

© Copyright 2015

Richard J. Callahan

**Growing Up in Taiwan: Counterfactual Models of Part-Time Work,  
Romantic Relationships, and Crime among Taiwanese Youth**

Richard J. Callahan

A dissertation

submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Dr. Ross L. Matsueda, Chair

Dr. William Lavelly

Dr. Min Li

Program Authorized to Offer Degree:

Sociology

University of Washington

**Abstract**

Growing Up in Taiwan: Counterfactual Models of Part-Time Work, Romantic Relationships, and Crime among Taiwanese Youth

Richard J. Callahan

Chair of the Supervisory Committee:  
Dr. Ross L. Matsueda  
Department of Sociology

Using marginal structural models for counterfactual inference and Bayesian Item Response Theory (IRT) models, this dissertation investigates the effect of two turning points in the life course, work and marriage, on deviance for Taiwanese youth between the ages of 15 and 26. The research draws on the Taiwan Youth Project (TYP) survey, a longitudinal study of youth in Taiwan that presents challenges for the analysis of crime because the survey items that measure deviance change with each wave. The dissertation overcomes this problem through a novel application of Bayesian IRT models to the measurement of crime. For the first time in criminology research, vertical scaling methods from the field of educational assessment are used to place latent IRT measures of deviance from different waves onto a common scale, allowing for the analysis of within-individual change in delinquency. A Bayesian approach is both

necessary given the large numbers of respondents each wave reporting zero deviant acts (a situation that leads to violations of frequentist IRT model assumptions in the case of the TYP data), and also makes possible a new approach for minimizing measurement error during vertical scaling using Dijkstra's algorithm. Substantively, work among Taiwanese youth of high school and college age causes an increase in deviance, while variability in work intensity and extensity does not explain variability in deviance. The findings also reinforce the importance of social control in Chinese culture: youth with low parental monitoring experience a strong, positive, causal effect of work on deviance, while for those with high parental monitoring, getting a part-time job does not cause any appreciable change in delinquency. Concerning romantic relationships, the intent to marry (either in general or one's dating partner) has a spurious association with deviance. Limitations of the available data restrict analyses of the impact of marriage on desistence to an analysis of smoking, and while Taiwanese who are married tend to not smoke, getting married does not actually cause people to quit. Overall, the findings suggest that for Taiwanese youth, school and family bonds affect both the timing of work and marriage and their subsequent effect on deviance.

## 摘要

本論文透過應用邊緣性結構模型之反事實分析法 (marginal structural model for counterfactual inference)和貝氏項目反應理論模型(Bayesian Item Response Theory [IRT] model)，研究 15 歲到 26 歲之台灣青少年的生命歷程中，婚姻和工作（包括打工）這兩個轉折點(turning point)如何導致其偏差行為的改變。

該研究的數據來自於《台灣青少年成長歷程研究》(TYP)項目。該項目對台灣青少年進行了追蹤調查。由於測量偏差行為的題目在每次調查年份都有些微改動，導致偏差行為之變項難以作貫時分析。為了解決該問題，本論文使用在犯罪學中從未應用過的貝氏項目反應理論模型和教育測量學(educational assessment)中的縱向量表化方法(vertical scaling methods)，以便把不同調查年份的偏差行為 IRT 潛變量分數放在同一標尺下，來分析個體偏差行為的變化。

本研究使用貝氏模型之必要性在於：1. 大部分被訪者自稱沒有不良行為，因此最大似然估計(maximum likelihood estimation) IRT 模型不能正確估算潛變量分數。2. 貝氏模型所支持的戴克斯特拉演算法 (Dijkstra's algorithm)，可將縱向量表化的測量誤差(measurement error)降到最低。

本論文之研究結論如下：1. 工作（包括打工、兼職工作和全職工作之變數）會導致 15 歲到 22 歲台灣人的偏差行為上升，但工作強度（工作時數）和長度（工作維持多久）不能解釋偏差行為的變化。

2. 社會控制理論(social control theory)對偏差行為在華人社會中扮演重要角色。受家長監管程度較低的少年，其工作對偏差行為的因果影響較大；但對受到家長監管程度較高的少年而言，從無工作的狀態到有工作的狀態之變化，對其偏差行為之因果影響幾乎等於零。

3. 在戀愛婚姻方面，有結婚意願（被訪者已確定結婚對象，或未確定結婚對象但準備將來結婚）和偏差行為無顯著關係。

另外，TYP 在調查婚姻對偏差行為的影響，僅限於抽煙行為。已婚的台灣人較少抽煙，但已婚狀態也並不會導致台灣人戒菸。

總之，本論文的研究成果表明台灣青少年跟家庭的關係緊密與否，影響到其工作和婚姻時間的早晚，而且他們跟家庭的關係並影響到其偏差行為。

## 摘要

本论文应用对反事实推断的边缘性结构模型(marginal structural model for counterfactual inference) 和贝氏项目反应理论模型(Bayesian Item Response Theory [IRT] model)，研究婚姻和工作这两个人生经历中的转折点(turning point)对15岁到26岁台湾青少年不良行为的影响。

该研究的数据来自于《台湾青少年成长历程研究》(TYP) 项目。该项目对台湾青少年进行了追踪调查。由於测量不良行为的题目在每次调查年份都有些微改动，导致难以分析不良行为的变化。为了解决该问题，本论文使用在犯罪学中从未应用过的贝氏项目反应理论模型和教育评估学(educational assessment)中的纵向量表化方法(vertical scaling methods)，以便把不同调查年份的不良行为 IRT 潜变量分数放在同一标尺下，来分析个体不良行为的变化。

本研究使用贝氏模型之必要性在於：1. 大部分被访者自称没有不良行为，因此最大似然估计(maximum likelihood estimation) IRT 模型不能正确估算潜变量分数。2. 贝氏模型所支持的戴克斯特拉演算法 (Dijkstra's algorithm)，可将纵向量表化的测量误差(measurement error)降到最低。

本论文的研究结论如下：1. 工作（包括打工、兼职工作和全职工作）会导致15岁到22岁台湾人的不良行为上升，但工作强度（工作时数）和长度（工作维持多久）不能解释不良行为的变化。

2. 社会控制理论(social control theory) 对不良行为的解释, 同样适用于中华文化环境。受家长监管程度较低的少年，其工作和不良行为的因果关系较大；但对受到家长监管程度较高的少年而言，从无工作的状态到有工作的状态之变化，对其不良行为之影响几乎等於零。

3. 在恋爱婚姻方面，有结婚意图（被访者已确定结婚对象，或未确定结婚对象但准备将来结婚）和不良行为无显著关系。

另外，TYP 在调查婚姻对不良行为的影响，仅限於抽烟行为。已婚的台湾人较少抽烟，但已婚状态也并不会导致台湾人戒烟。

总之，本论文的研究成果表明台湾青少年跟学校和家庭的关系紧密与否，影响到其工作和结婚时间的早晚，而且他们跟家庭的关係并影响到其不良行为的产生和发展。

## TABLE OF CONTENTS

List of Figures .....	ix
List of Tables .....	xi
Chapter 1. Introduction.....	1
Chapter 2. A Latent Variable Model for Modeling Deviance in Longitudinal Surveys with Changing Questions.....	11
Chapter 3. Literature Review of the Precursors of Delinquency for Taiwanese Youth.....	67
Chapter 4. Empirical Analysis of the Effect of Work on Delinquency for Youth of High School and College Age.....	113
Chapter 5. Modeling the Effect of Steady Romantic Relationships and Marriage on Deviance in Taiwan.....	201
Chapter 6. Conclusion.....	259
Bibliography.....	267
Appendix A. Detailed Information about Items Measuring Deviance in the Taiwan Youth Project Survey.....	282
Appendix B. Technical Details Concerning Implementation of the Bayesian Generalized Partial Credit Model for the TYP Data.....	299
Appendix C. Description of the Haebara and Stocking-Lord Methods for Vertical Scaling.....	321
Appendix D. Design of a Filter to Create Data for Work Intensity, Work Extensity, and Whether Respondents Worked During the Previous Year, Academic Year, and Holiday Breaks.....	324
Appendix E. Time-Varying Covariates Used to Model Selection into Work.....	329
Appendix F. Missing Value Imputation Procedure for Chapter 4.....	355
Appendix G. Covariates Used to Model Survey Attrition.....	359
Appendix H. Principal Components Analysis of Criteria for Marriage.....	361
Appendix I. Nonresponse among the J3 Variables Predicting Selection into Marriage.....	363

## LIST OF FIGURES

Figure 2.1. Item Characteristic Curve (ICC) associated with the dichotomous question about smoking in Wave 4 of the TYP Survey for the J1 cohort.....	25
Figure 2.2. Item Characteristic Curve (ICC) associated with the dichotomous question about fighting in Wave 4 of the TYP Survey for the J1 cohort.....	26
Figure 2.3. An example of a well-fitting test item: fraud in Wave 7.....	48
Figure 2.4. A poorly-fitting test item: drug use in Wave 4.....	48
Figure 2.5. Hypothetical example of vertical scaling for the delinquency items in the Taiwan Youth Project survey using common-item linking.....	55
Figure 2.6. Values of $\sigma_A + \sigma_B$ associated with the choices of characteristic curve method and common items for vertical scaling that minimize measurement error associated with the separate calibration of pairs of TYP waves.....	61
Figure 2.7. Choice of separate calibrations in Figure 2.6 and a “root” node for a scale that minimizes total measurement error for vertical scaling of the TYP delinquency items.....	62
Figure 4.1. A model for the causal effect of a time-varying treatment on an outcome at two time points.....	165
Figure 4.2. Nonparametric locally-weighted robust least squares regressions of deviance on work intensity and extensity among respondents to Wave 9 of the Taiwan Youth Project survey.....	193
Figure 4.3. Nonparametric locally-weighted robust least squares regressions of deviance on work intensity and extensity among respondents to Wave 8 of the Taiwan Youth Project survey.....	194
Figure B1. Derivation of informative priors to estimate latent constructs for criminality in the TYP data.....	311
Figure B2. Distribution of posterior deviance estimates for respondent with raw scores of 0 and 1 in Wave 1 (J3 cohort).....	312
Figure B3. Distribution of posterior deviance estimates for respondents with raw scores of 2 and 5 in Wave 1 (J3 cohort).....	312
Figure B4. Distribution of posterior deviance estimates for a respondent with a raw score of 11 in Wave 1 (J3 cohort).....	313
Figure B5. Distribution of EAP estimates for J3W1 survey respondents with raw scores of 0.....	315
Figure B6. Distribution of EAP estimates for J3W1 survey respondents with raw scores of 1.....	315
Figure B7. Distribution of EAP estimates for J3W1 survey respondents with raw scores of 2 and 3.....	316
Figure B8. Distribution of EAP estimates for J3W1 survey respondents with raw scores between 4 and 6.....	316
Figure B9. Distribution of EAP estimates for J3W1 survey respondents with raw scores of 7 or greater.....	317

Figure B10. Example of rescaling the mixture densities in Figures B7 and B9 to suit a hypothetical wave with a maximum observed raw score of 7.....	318
Figure B12. Example diagnostic plot showing autocorrelation among posterior draws for an item difficulty parameter in Wave 2 of the J3 survey.....	320
Figure B13. Example diagnostic plot showing convergence for posterior draws for an item difficulty parameter in Wave 2 of the J3 survey.....	320

## LIST OF TABLES

Table 1.1. Structure of the TYP questionnaires.....	6
Table 2.1. TYP delinquency items used in this dissertation.....	31
Table 2.2. Pearson product moment correlations of delinquency with depression.....	64
Table 3.1. TYP variables used to model selection into work that are based on a review of the criminology literature with respect to Taiwan.....	110
Table 4.1. TYP variables used to model selection into work that are based on a review of the causes and consequences of youth employment in Taiwan.....	138
Table 4.2. Distribution of work variables during J1 survey waves for high school and college.....	142
Table 4.3. Covariates used for the selection models into work.....	147
Table 4.4. Fixed effects negative binomial (“FENB”) regressions of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months.....	151
Table 4.5. Random effects negative binomial regressions of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months.....	151
Table 4.6. Item nonresponse patterns for variables affecting selection into work.....	154
Table 4.7. Fixed and random effects negative binomial regressions of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months (after multiple imputation).....	159
Table 4.8. Fixed and random effects negative binomial regressions of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months, controlling for time-varying covariates.....	159
Table 4.9. Auxiliary random effects negative binomial regression of delinquency on part-time work.....	161
Table 4.10. Random effects negative binomial regression of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months, including respondents only observed for one person-year and incorporating additional time-invariant covariates.....	162
Table 4.11. Logistic regressions predicting missingness in the Taiwan Youth Project survey (J1 cohort).....	174
Table 4.12. Logistic regressions predicting part-time work at any time during the year in the Taiwan Youth Project survey (J1 cohort).....	176
Table 4.13. Counterfactual inverse Gaussian and negative binomial regressions of delinquency on work during the year.....	181
Table 4.14. Counterfactual inverse Gaussian and negative binomial regressions of delinquency on work during the academic year.....	183
Table 4.15. Counterfactual inverse Gaussian and negative binomial regressions of delinquency on work during the summer or winter breaks.....	184
Table 4.16. Counterfactual Inverse-Gaussian and negative binomial regressions of delinquency on work during the academic year and on work during the summer or winter breaks.....	186

Table 4.17. Counterfactual inverse Gaussian and negative binomial regressions of delinquency on work during the year for respondents to Waves 5, 6, 8, and 9, with an interaction between parental monitoring and work.....	187
Table 4.18. Negative binomial regressions of delinquency on Mortimer’s work typologies during the year for Waves 8 and 9.....	195
Table 5.1. Random effects negative binomial regression of delinquency on dating.....	211
Table 5.2. Correlations between deviance and intent to marry for respondents to the Taiwan Youth Project survey who are currently dating someone.....	219
Table 5.3. Correlations between deviance and intent to never marry for respondents to the Taiwan Youth Project survey.....	220
Table 5.4. Negative binomial regressions of deviance on intent to never marry for respondents to the Taiwan Youth Project survey.....	221
Table 5.5. Negative binomial regressions of deviance on intent to never marry.....	222
Table 5.6. Number of married respondents for each cohort-wave in the TYP Survey.....	224
Table 5.7. Tetrachoric correlations for prospective and retrospective smoking questions for the TYP survey, J3 cohort.....	225
Table 5.8. Fixed and random effects logistic regression models for the effect of marriage on smoking for respondents to the Taiwan Youth Project survey who are at least 18 years old.....	228
Table 5.9. Fixed and random effects logistic regression models for the effect of marriage on smoking for respondents to the Taiwan Youth Project survey who are at least 12 years old.....	230
Table 5.10. Fixed and random effects multinomial logistic regressions for the effect of marriage on smoking for Waves 1-10 of the Taiwan Youth Project survey.....	232
Table 5.11. Multinomial logistic regressions for selection into intent to marry for Wave 6 of the TYP J3 cohort.....	238
Table 5.12. Logistic regressions predicting unit nonresponse in the Taiwan Youth Project survey (J3 cohort).....	245
Table 5.13. Covariates used for the selection models into marriage, for each survey wave used for the marriage analysis.....	248
Table 5.14. Logistic regressions predicting marriage in the Taiwan Youth Project survey (J3 cohort).....	251
Table 5.15. Counterfactual random effects logistic regressions for the effect of marriage on smoking for Waves 7, 8, and 2011 of the TYP survey, J3 cohort, using inverse probability weighting for unit nonresponse and selection into marriage.....	256
Table A1. Format of the delinquency question groups in the Taiwan Youth Project survey.....	282
Table A2. Response category wordings for the Taiwan Youth Project survey delinquency items.....	283
Table A3. Wording of the Taiwan Youth Project survey delinquency items.....	284
Table A4. Fraction of respondents responding to each category of the Taiwan Youth Project survey delinquency items with categorical responses on a frequency scale.....	287

Table A5. Fraction of respondents responding to each category of the Taiwan Youth Project survey delinquency items with dichotomous responses.....	290
Table A6. Results of one-dimensional confirmatory factor analyses for each cohort-wave in the Taiwan Youth Survey, along with Cronbach's alpha reliabilities.....	291
Table A7. Discrimination and item difficulty parameter estimates for Taiwan Youth Project survey delinquency items with dichotomous responses.....	295
Table A8. Discrimination and item difficulty parameter estimates for Taiwan Youth Project survey delinquency items with polytomous responses.....	296
Table A9. Number of independent draws from the posterior distribution of parameter estimates for each cohort-wave in the Taiwan Youth Survey.....	298
Table H1. Marital Qualification Factor Analysis Component Score Coefficient Matrix.....	362
Table I1. Item nonresponse patterns for variables affecting selection into marriage for the TYP survey.....	364

## ACKNOWLEDGEMENTS

Every dissertation must have a sole author, but no dissertation is ever the product of one person alone. I am grateful for the opportunity to thank the many people involved in these pages. First and foremost, I would like to thank my parents. This dissertation in particular is a tribute to rigor, and to staying committed to answering an important empirical question using the best available data despite any number of obstacles associated with the data or available methods. Such challenges often constitute opportunities for moving our field forward. My father, an inventor and a sonar engineer, demonstrated these qualities. My mother, who earned her Ph.D. in Education in the 1970s when few women in the United States were receiving doctoral degrees, and who later pioneered techniques in distance education at the University of Maryland when the Internet was still in its infancy, did so as well. I am blessed with two parents who have set a powerful example for me, both in their personal lives and in their careers. I was fortunate to grow up with an entire library in my basement, where I spent many afternoons reading such authors as the Hardy Boys, Ray Bradbury, and Victor Hugo. If anything in this dissertation is creative, it is due to my parents. The rest, due to simple persistence, is to their credit as well.

Next, I would like to thank all the people who have taught me Chinese over the years. Tony Chen, my first Chinese teacher, deserves special mention. A physicist from China who immigrated to the U.S. and worked his way to a job at NASA, he reinforced everything my parents taught me about hard work and then some. I was about 12 years old when we first met. He was a waiter in a Chinese restaurant and still brushing up on his technical English, so we started a weekly language exchange in which my mother helped him with his business English

for an hour and he taught me Mandarin. I still remember the first part of the paragraph he had me memorize when I was about 12: “Every Sunday I get up at 6:30. Then I exercise, and study...”

The afternoon that he asked me to repeat that passage after him in Mandarin and then translated to English what he had asked me to say, and my mother and I looked at each other across the dinner table and just laughed. But the message stuck: this is the level of commitment necessary to pull oneself up by the bootstraps, and I have always had tremendous respect for Chinese scholars. I would also like to thank the Hope Chinese School in College Park, Maryland, the Department of Asian and Middle Eastern Languages and Literature at Dartmouth College, my teachers at Beijing Normal University and Tunghai University in Taichung, Taiwan, and the staff at the Inter-University Program for Chinese Language Studies (IUP) at Qinghua University, Beijing, and the Inter-University Chinese Language Program (ICLP) at National Taiwan University, Taiwan. Preparing course material for just a few students in language programs with small class sizes can be exhausting and thankless work. Those teachers deserve a great deal of credit for this dissertation.

On the analytical side, I would like to start by thanking my good friend and mentor, Dr. Ross Matsueda. His intellectual imprint is on every page of this dissertation, from the substantive arguments to the models. Working with him has been a tremendous honor and privilege, and I highly recommend any graduate student or faculty member presented with the opportunity to work with Ross to strongly consider doing so. I’d also like to thank my committee members, in no particular order. Dr. Peter Hoff provided valuable input concerning important technical details concerning improving convergence for the Bayesian IRT model in Chapter 2, and his excellent graduate-level course on Bayesian statistics also was deeply informative. Dr. Thomas Richardson also provided extremely helpful feedback concerning the design of the marginal

structural models used in Chapters 4 and 5. Dr. Gary Hamilton and Dr. Bill Lavelly provided helpful feedback concerning the substantive portions of the dissertation. Dr. Min Li provided important advice concerning the IRT models, and one of her graduate students translated the abstract into simplified Chinese. Dr. Lin Wen-Hsu, a postdoctoral student at Academia Sinica, provided invaluable assistance with data extraction and cleaning for the counterfactual models in Chapter 5. And the Taiwan Youth Project survey team in the Institute of Sociology at Academia Sinica has produced a truly amazing data set, with very few missing cases given the number of respondents involved. Their efforts are truly impressive, and working with the fruits of their labor has been a privilege.

I would also like to thank the University of Washington Department of Sociology faculty and staff, and my fellow graduate students for their support. I would like to make special mention of the University of Washington's Center for Statistics and Social Sciences (CSSS), a unique organization that is not an academic department of itself but rather a center of funding through which professors who are experts in statistics impart their knowledge to graduate students from a wide variety of disciplines, thereby ensuring that the statistical methods that we learn in the social sciences are not watered down in any shape or form. What a rare opportunity it has been to learn from experts such as Dr. Hoff and Dr. Richardson. Other outstanding CSSS professors from whom I have taken classes include Dr. Adrian Dobra, Dr. Christopher Adolph, Dr. Jerald Herting, and Dr. Miruna Petrescu-Prahova. In the Department of Sociology, I would also like to thank Dr. Katherine Beckett, Dr. Kate Stovel, Dr. Edgar Kaiser, and Dr. Becky Pettit (who has since moved to the University of Texas-Austin). I would highly recommend the University of Washington to anyone interested in quantitative social science.

A few people from my undergraduate years at Dartmouth College deserve special mention here. Dr. Hua-Yuan Li Mowry, Dr. Susan Blader, and Phuoc Le helped to foster my love for the language, culture, and people of China. Dr. Robert Welsch taught me to write proposals, read critically, and conduct qualitative research. In the spirit of what I have learned from Dr. Welsch, I strongly recommend that future research identify some of the causal mechanisms underpinning the findings herein through qualitative interviews. On the quantitative side, I would also like to thank my undergraduate mentor at Dartmouth College, Dr. Joel Levine. As is the case for Dr. Matsueda, my main mentor at the University of Washington, I cannot say enough good things about Joel. Dr. Levine makes sure that all his students carefully consider the ethical implications of their work, a matter close to his heart as he is one of the pioneers of social network analysis, having worked with Harrison White's networks group at Harvard in the 1960s and early 1970s. Learning from Joel was also an honor and a privilege, and it is his willingness to stand up to anyone when defending his ethical principles that has stayed with me throughout my life and guided me long after graduation. Learning from Joel is also a rare opportunity, and I highly recommend Dartmouth's Mathematics & Social Science honors major, a hybrid program that allows students each year to design their own programs and conduct interdisciplinary research. The program is a wonderful model that I would argue should be used more widely in other colleges. Of course, whoever is reading this dissertation is most likely past the point of choosing an undergraduate college, but please pass the information along!

I would also like to thank the Santa Fe Institute (SFI), particularly Dr. David Feldman at the College of the Atlantic, for contributing to my interdisciplinary training through the Complex Systems Summer School, held in Santa Fe every year. I was fortunate to participate in the Beijing summer school from 2007-2009, once as a student and twice helping to handle logistics.

The SFI School had a strong influence on my thinking, and indeed my initial impetus for studying the causal effect of marriage on crime in a Chinese context stemmed from a presentation by Dr. Marcus Feldman at the Beijing summer school concerning China's gender imbalance. I am extremely grateful for the opportunity to have studied and worked there, and wish all of my classmates and colleagues during those years the very best. All of us at the school, students and staff, saw those grass-roots academic collaborations as part of the best way forward for international relations, and there was a high level of energy in those schools and workshops. I would strongly encourage the continuation of such wonderful programs as the SFI Complex Systems Summer School.

Part of the research for this dissertation was funded by a Fulbright grant from the Department of State. I would like to thank Mr. William Akeley at the Institute for International Education, the Taiwan Foundation for Scholarly Exchange, the Department of State, and the American Institute in Taiwan for their support while I conducted research at Academia Sinica. I especially would like to thank Drs. Wu Chyi-Yin and Yi Chin-Chun for their valuable guidance. I thoroughly enjoyed Academia Sinica's intellectual rigor and the wonderful environment for research in Nankang, Taipei. I hope that time will demonstrate that this project was an effective use of U.S. taxpayer money.

No project can be completed without the support of a strong network of friends. My girlfriend, Buffy Wang, deserves special recognition for supporting me in ways both large and small as I have invested so much time and energy in this project. I would also like to thank her family. I already mentioned Dr. Joel Levine concerning ethics, but my old scoutmaster, Kurtis McDowell, and the entire staff of Camp Spencer at Broad Creek Memorial Scout Reservation

from 1996-1998 also deserve special mention. All my teachers at Arundel Senior High School, Gambrills, MD, also should share in the credit.

What can I say about my family, except that they are deserving of endless gratitude and praise. My two brothers David and Bill, my two sisters-in-law Lisa and Lisa, my nephew Ray, and my four wonderful nieces Lily, Brianna, Emily, and Maisy have put up with me while I have spent countless hours and even some Christmases on the other side of the country and world. And then there are other members of my extended family – my uncle and cousins – who I have had only rare opportunities to see during these years in graduate school. To all of you, I wish I could say that I will somehow make up for lost time, but I know that while I will try all my life, the reality is that I can never repay that debt. As Marcus Aurelius writes in his *Meditations*, we only live in the present, and our actions on a day-to-day basis are all that matter; everything else is but wind. So all I can really say to my family is thank you, and God bless you all.

I would also like to mention some remarkable people who I had the honor of meeting while working for Linder & Associates as a junior-level management consultant for the Los Angeles Police Department from 2003-2005. John Linder is devoted to the reduction of crime through cultural transformations of police organizations, and I learned a great deal from him and from my other two colleagues, Richard Sawyer and Dr. Kevin Unter. Credit for any part of this dissertation that is well-written is due in large part to their wonderful guidance during those years. I also had the pleasure of working with a number of inspiring people at the Los Angeles Police Department, who are simply too numerous to mention by name. Chief among them (no pun intended) was Chief William J. Bratton.

I learned two important lessons from Chief Bratton, a truly remarkable man. The first involves setting “stretch goals.” While he was Chief of Police, every year the Los Angeles Police

Department set for itself a jaw-dropping goal of 20% reduction in crime, with a focus on Part I crimes (such as murder, rape, robbery, burglary, and auto theft). For any individual police officer, attempting to achieve that goal involves increasing the number of contacts with members of the public, potentially leading to a greater number of uses of force. It also involves a temptation to misclassify serious crimes as less-serious crimes. An amazing trait of Chief Bratton is that he is able to walk the fine line between supporting his officers in the field, defending them from accusations of malpractice by the public when they are frivolous, and yet promoting only officers of the finest ethical caliber that hold themselves and their colleagues accountable for any policing practices that are not considerate or constitutional. In doing so, he presses his officers to do their utmost to achieve these goals for crime reduction, holding each other accountable and coming up with new strategies and tactics, and in the process earning the respect and admiration of those who they lead. Because of Chief Bratton's leadership style, those under him when at their best are willing to put their careers on the line each day, thinking of their roles in a greater goal for their police department while trusting their Chief to support them all the same if their best is short of the stretch goal. In the end, crime in Los Angeles dropped by 49.8% from 2003-2012, and the police and public alike loved him for it. The lesson, which Chief Bratton projected to his officers, was simple: believe in something, be willing to fail, do your utmost, and be smart about it. Then, even if you come up short, you will have been moving in the right direction and your colleagues and the people you serve will support you the entire way.

The second lesson I learned from Chief Bratton was really one from Jack Maple, a legend in the New York Police Department and an old colleague of Chief Bratton and John Linder. Mr. Maple devised the four principles of Compstat, one of the country's most effective and popular crime reduction programs, which are outlined further in his book (Maple 2000): accurate and

timely intelligence, effective tactics, rapid deployment, and relentless follow-up and assessment. It is by adhering to these four principles that the police can lower crime, and my time in Los Angeles showed me that the most challenging obstacles to implementing Compstat are not technological but rather cultural. I am convinced that sociologists have a role to play in lowering crime, and that while we must maintain our objectivity to be dispassionate observers of the nature of crime there is no reason why we cannot research the questions that policymakers, kids, and parents would like to have answered and couch these questions in counterfactual logic, regardless of the actual research methods used. In this way, our results are not dry formulas but rather immediately useful to the public. Indeed, I believe that the text of this dissertation contributes to the first two principles of Compstat: accurate and timely intelligence (about the causal effects of work and marriage on general deviance and smoking, respectively, in Taiwan), and effective tactics (the use of Bayesian IRT models for the measurement of crime).

In that spirit, I would like to add an extra-special “thank you” to you, the reader of this dissertation. Whether you are a researcher, practitioner, or policy maker, this text was written for you, and you will know best how to use it and how to evaluate the findings in light of future work – some of which may be yours, whether it be on the research side or the implementation side. These responsibilities map to the last two principles of Compstat: rapid deployment and relentless follow-up and assessment. That is, the research tools described herein should be put into practice, and both their effectiveness in measuring crime and this dissertation’s substantive findings should be rigorously assessed. Those tasks are left as an exercise to the reader. Thank you in advance for your help.

This dissertation is an endeavor that is part of something larger: an effort to determine how marriage and work lower crime in a Chinese cultural context, just in case China’s gender

imbalance could affect the crime rate. A question so complex, with so many political, cultural, and technological impediments to its resolution, cannot be solved by one person. This dissertation only scratches the surface – mostly on the technological side, with some useful results concerning work and marriage – and even as a modest contribution it is clearly the work of many hands. I have endeavored here to thank the people involved, but I am sure I have missed many and ask the forgiveness of those who I have not mentioned. Any errors or omissions in the analyses and writing are my own.

## **DEDICATION**

In loving memory of my mother, Barbara J. Callahan, Ph.D., who wrote her dissertation on a typewriter.

I can't imagine how you did it.

## Chapter 1: Introduction

This dissertation investigates the causes of deviant behavior for Taiwanese youth from their high school years through age 26. One of the most important reasons that sociologists systematically investigate human behavior is to grow a body of knowledge concerning the causes of social problems in order for readers of their work to then mitigate the negative impacts of these social problems on society. Life course research, which follows people over time and examines how changes in their development and social environment influence their behavior, is a common theme for research oriented around social problems, because carefully designed studies of the circumstances in which people change their behavior allow for the testing of hypotheses about how outside interventions could lead to desirable social outcomes. “Turning points” in the life course – such as graduating from school, getting a job, marriage, and becoming a parent – are compelling areas of focus for life course research, as they constitute powerful, replicable ways to measure how a person’s changing bonds to society affect behavior.

This dissertation, with its focus on how work and marriage affect deviant behavior for Taiwanese youth as they enter adulthood, falls squarely within this tradition of research. Turning points in the life course impact crime in numerous, dynamic ways, and the present work seeks to apply some lines of inquiry from Western studies to a Chinese context. Studying youth is important because of the “age-crime curve,” a characteristic of societies that spans cultures and historical time periods in which youth commit crime at higher rates than those in other age groups, then eventually desist as they transition into adulthood (e.g. Gottfredson and Hirschi 1990). Understanding the mechanisms through which desistence occurs for people of these ages can therefore help decision-makers – from policy-makers to parents – to promote it.

One life course theory, Sampson and Laub's age-graded theory of informal social control (Laub and Sampson 2003; Sampson and Laub 1993), predicts that desistance occurs during turning points in the life course because social bonds associated with the adult roles into which people transition reduce an individual's motivation to offend and alter his or her social networks as well. Entering the labor force constitutes one such turning point, especially if the job is a stable one (Mortimer 2003; Sampson and Laub 1993). Steady romantic relationships and marriage constitute a second. Among various life-course transitions, those from school to work and from singleness to dating and then marriage are particularly important because they are so common. The purpose of the present research is to extend our knowledge of the effects on deviance of work, dating with an intent to marry, and marriage to Taiwanese youth of high school and college age, using counterfactual models designed to explicitly model causal relationships.

This dissertation topic and the choice of a Taiwanese data set are compelling because of its potential utility to Western criminologists as well as to Taiwanese and Chinese sociologists and policy-makers.<sup>1</sup> Driven by the importance of the topic, this investigation has been faithful to its initial conception and substantive hypotheses despite a number of technical challenges, one of which is so complex as to have required an improvement on existing methodologies for the measurement of crime using latent variable methods, addressed in the second chapter of this dissertation. Knowing why the present topic under investigation is so compelling as to motivate such an investment of time in developing new methodologies requires an understanding of the unique role and importance of social control in Taiwan. Bonds to social institutions such as

---

<sup>1</sup> Following a civil war in 1949, the governments of the People's Republic of China (PRC) and Republic of China (ROC) both respectively claim to be "China" with either Beijing or Taipei, respectively, as the capital of China. In referring to the PRC as "China" and the ROC as "Taiwan," and to its residents as "Chinese" and "Taiwanese" respectively, this dissertation follows the convention in most contemporary scholarly publications written in English, and does not constitute any statement by the author concerning the political status of these two parts of the world.

school and the family are thought to be particularly strong in collectivist cultures such as those in China and Taiwan, and according to control theory (Hirschi 1969), such bonds deter would-be offenders from crime. Work and romantic relationships are turning points most directly linked to social control in the criminological literature, and so an investigation of the effects of these turning points on deviance in a culture with high social control helps to extend our understanding of how control theory matters with respect to life course transitions in general. Additionally, Chinese culture places a high importance on conformity to conventional activities, such that most parents encourage their children to prioritize homework and test preparation over part-time work and to delay dating until college. In the United States, work (e.g. Mortimer 2003) and marriage (e.g. Sampson, Laub, and Wimer 2006) have been associated with positive outcomes for youth, including desistance from crime. But many of the people who benefit from these life-course changes do so because they are moving from a relatively unstructured environment, into a more structured one, with colleagues or loved ones who could suffer materially and psychologically as a result of one's getting arrested. Because of the relative importance of school and the family in Taiwan, however, many young people who transition from a heavy investment in school and the family into work and marriage are moving from an environment of high social control to another environment of high social control. Therefore, it is possible that work and marriage may not be as important to desistance from crime as they are in the West, or perhaps would even lead to *increased* crime if they happen so early in life that they are counter to social norms about the importance of getting a good education first, and thus constitute a type of deviant behavior in and of themselves. Yet even if deviant people tend to enter these turning points earlier in life, they may still deter people from crime. Controlling for the plethora of background variables in order to reach causal inferences net of these sorts of selection effects therefore constitutes a central

focus of the literature review in Chapter 3 and the empirical analyses in Chapters 4 and 5, which respectively cover work and romantic relationships.

In an increasingly competitive world in which young people everywhere are delaying marriage and investing more time and money in higher education, a critical question will be whether delays in these life course transitions in the general population would lead to increasing crime rates. Taiwan is at the vanguard of this shift, with a recent acceleration in the rate of increase of age at first marriage (Chen and Chen 2014) accompanied by an explosion in the number of students in higher education. Research that explicitly uses counterfactual models for causal inference is therefore quite important to Taiwanese policy makers: were marriage merely associated with lower crime but not causal, then delayed marriage due to additional schooling would not cause an increase in crime – but if the relationship were in fact causal, then the delay of this turning point within the general population would result in rising crime rates.

Outside of Taiwan, while the results of this research are not directly useful for policy makers because the sample used is only representative of Taiwan (and northern Taiwan at that), they can nonetheless help to motivate and guide designs for further empirical research. The methodologies used for the first time in Chapter 2 to model deviance in a survey with changing questions are more generally usable, and the findings concerning a society with high social control can yield some insight into how research might be conducted in other societies with high social control. Mainland China is also experiencing a similar shift towards higher education and later marriage. More seriously, since the 1980s the country has experienced a growing imbalance in the sex ratio at birth, and demographers have projected that as many as 40 million young men will be unable to find wives by 2020 (South China Morning Post Staff Reporter 2008), with as many as 9.3 “missing girls” identified in the 2000 census alone (Ebenstein 2010), predicting a

greater marriage squeeze ahead. Any hypothetical Chinese policy to adjust future marriage rates by bringing the sex ratio at birth closer to its natural value would need to be informed by studies in China that involve exactly the sort of counterfactual questions concerning the causal effect of marriage on crime that are addressed in Chapter 5. Moreover, any policy to correct the gender imbalance would not take effect for years, but policies connected to youth employment and crime such as that addressed in Chapter 4 could take effect much more quickly.

This dissertation, while modest in what it can achieve with the available data, is therefore part of a much greater set of endeavors. And given the importance of research conducted in a Chinese context, Taiwan is a natural choice as the sort of high-quality longitudinal surveys of delinquency necessary for deriving causal inferences about the effect of work and marriage on deviance are still not available in China (Liu 2008), while the ongoing Taiwan Youth Project (TYP) survey at Taiwan's Academia Sinica has painstakingly gathered and cleaned a rich source of new data source for such research (Yi 2012). The discussion now turns to the TYP data set.

### *The Taiwan Youth Project Survey*

This dissertation draws on the Taiwan Youth Project (TYP) data, a ten-wave, longitudinal survey of 5,586 Taiwanese youth between the ages of 13 and 26 (Yi 2012). The survey was originally designed to measure the effect of an educational reform in junior high school, which runs from 7<sup>th</sup> to 9<sup>th</sup> grade. It contains two birth cohorts, one with respondents who were in 7<sup>th</sup> grade during the first wave of the study (the "J1" sample), and one with respondents in 9<sup>th</sup> grade (the "J3" sample). Respondents were sampled from 40 schools in Taipei city, Taipei County outside the city, and Taoyuan County (a rural sample). The stratified cluster design involved clustering on junior high schools in order to identify school-level effects on the respondents, but did not include social network questions in a manner similar to the National Longitudinal Survey

of Adolescent Health. Parents and teachers were also interviewed as part of the survey. Table 2 shows the survey design (Academia Sinica Institute of Sociology 2011). In Table 2, the years that respondents were in junior high school are prefixed with a “J”, those corresponding to senior high school (10<sup>th</sup> through 12<sup>th</sup> grades) are prefixed with an “S”, and college years and beyond are prefixed with a “C”. The table displays the format of the survey during each wave (a questionnaire administered in class, a telephone survey, or an in-person interview at home), the year and month of the survey (which differ slightly within waves for the J1 and J3 samples), and whether parents or teachers were also interviewed during each wave.

**Table 1.1.** Structure of the TYP questionnaires (Academia Sinica Institute of Sociology 2011).

Sam- ple	Resp.	W1	W2	W3	W4	W5	W6	W7	W8	W9
J3	Time	3/00	10/00	10/01	3/03	2/04	10/04	12/06	11/08	
	Y*	J3 Quest.	S1 Phone	S2 Phone	S3 Interview	C1 Phone	C2 Interview	C4 Interview	C5 Phone	
	P*	Quest.			Interview		Interview	Interview		
	T*	Quest.								
J1	Time	3/00	3/01	3/02	10/02	10/03	3/05	3/06	6/07	6/09
	Y*	J1 Quest.	J2 Quest.	J3 Quest.	S1 Phone	S2 Phone	S3 Interview	C1 Phone	C2 Interview	C4 Interview
	P*	Quest.		Interview			Interview		Interview	Interview
	T*	Quest.	Quest.	Quest.						

\*Y = youth survey; P = parent survey; T = teacher survey

While the Taiwan Youth Project survey has the distinct advantage of being useful for deriving causal inferences, like any longitudinal survey it also has limitations. First, survey attrition could be connected to the variables of interest, biasing results. This dissertation uses conventional statistical methods for correcting for bias due to both item non-response (cases in which a respondent completes a questionnaire but does not complete all the questions on the survey) and unit non-response (cases in which a respondent does not complete a survey at all

during a particular year). Unit non-response is perhaps the more important for this research because some of the analysis in Chapter 4 demonstrates that a respondent's deviance at the beginning of the TYP survey predicts his or her loss to follow-up, with more deviant respondents less likely to be interviewed later. The causal models in this dissertation therefore all correct for unit non-response, and some of them also address item non-response.

Second, the TYP survey questions used to measure the same concept change in wording from one survey year to the next, rendering invalid a direct comparison of the responses across survey years and constituting a serious obstacle for measuring deviance in a manner that allows an analysis of the causes of within-individual change in criminality, the central focus of this dissertation. Problems like this one are nonetheless common in longitudinal surveys because budgetary considerations, changing research priorities, the need to revise survey questions to more reliably measure the intended variable of interest, and personnel changes can all lead to changes in the wording of the surveys over time. All of Chapter 2 is dedicated to addressing this problem, using models originally from educational assessment to overcome obstacles in measuring crime in longitudinal surveys with changing delinquency questions. In particular, building on the first application of Item Response Theory (IRT) models to criminology by Osgood et al. (2002), Chapter 2 applies a set of methods known as vertical scaling from the literature on educational assessment and psychometrics to criminological research for the first time, and proposes a new approach for minimizing measurement error when placing latent variables from multiple waves of a longitudinal survey onto a common scale.

The chapter demonstrates that crime can be reliably measured even in surveys where the questions change, but that doing so may require researchers to use Bayesian models.<sup>2</sup>

---

<sup>2</sup> While the source code from this research is not currently in the form of a user-friendly software package, I am happy to make all the code available to other researchers to use (the Bayesian IRT models are written in R).

Longitudinal surveys have an important major strength, however, that justifies rigorously addressing all their limitations: they are extremely useful for helping researchers to answer causal questions outside of clinical settings. In an example of experimental treatment design – the gold standard of causal inference – to calculate the causal effect of a treatment on the odds of a patient recovering, medical researchers randomly assign patients to either receive the drug or receive a placebo. In the present case, however, forcing Taiwanese youth to either work part-time or not, or to either enter into a romantic relationship or not, just to observe the effects of these changes in their lives on their subsequent delinquency would be both unfeasible and unethical. Absent the availability of this proven research methodology for deriving causal inferences, the current research adopts the next best approach: statistical models for causal inference from observational data.

#### *Marginal Structural Models for Causal Inference from Observational Data*

A brief description of the methodology underpinning causal inference is in order, as it motivates and structures the rest of the dissertation. In a similar interdisciplinary spirit to using psychometric models to model deviance in a survey with changing delinquency items, to derive causal inferences from observational data this dissertation borrows a methodology from the epidemiological literature: a class of counterfactual models that use inverse probability of treatment weighting, also known as IPTW (Hernán, Brumback, and Robins 2000; Robins 2000). IPTW models compensate for non-random assignment to treatment groups by first modeling selection into the treatment using observed variables, and then weighting observations based on the inverse of the probability of receiving the treatment that a respondent actually receives. If the selection model accounts for all the background variables that influence both the outcome and

the odds of receiving the treatment (known as *confounders*), then the differences between the groups receiving and not receiving the treatment become statistically random in the reweighted sample. Such models avoid biases currently found in structural equation models for time-varying treatments (Hernán, Brumback, and Robins 2000; Robins, Hernán, and Brumback 2000).

Due to its focus on causal inference, one goal of the present study is to identify as many variables as possible that could confound the relationship between employment and deviance and between marriage and deviance in Taiwan, and then to control for them when fitting a causal model (e.g. Sampson, Laub, and Wimer 2006). As the literature associated specifically with selection into delinquency and work in Asia is sparse, I review the literature for any background variables that have been shown in Taiwan to affect either work, juvenile delinquency, or both, and then include both sets of covariates in the models for selection into work in Chapter 4. The strategy in Chapter 5 for identifying variables affecting both romantic relationships and deviance in Taiwan is similar. As both chapters include covariates predicting deviance in the selection models, Chapter 3 reviews the causes of delinquency in Taiwan. This broad approach to identifying variables for the IPTW selection models is justified because including a potential confounder in a selection model that in fact is orthogonal to either work or crime has no adverse effect on the reweighting process, but excluding a confounder biases causal estimates.<sup>3</sup> Moreover, as both part-time work and dating while still in high school and college are sometimes themselves considered deviant in Taiwan, however, a significant overlap is expected between the predictors of deviance and the predictors of marriage and part-time work.

The rest of the dissertation proceeds as follows: Chapter 2 addresses the problem of measuring deviance in the Taiwan Youth Project survey using a latent variable Item Response

---

<sup>3</sup> Occasionally conditioning on a variable that is not a confounder can introduce bias in a causal model due to collider stratification, but the risk of doing so is generally less than the risk of improperly omitting a confounder (Greenland 2003).

Theory model. The chapter demonstrates the feasibility of modeling deviance even when the survey items change, as long as pairs of survey waves retain common survey items that make possible the rescaling of latent variables for criminality from one wave to the next, in order to place respondents' scores for all survey waves onto the same scale. Chapter 3 then presents a literature review of the etiology of juvenile delinquency in Taiwan. Chapter 4 completes the review of selection into part-time work, and then presents empirical analyses of the causal effect of part-time work on deviance among Taiwanese youth of high school and college age. In a finding demonstrating the importance of social control in Chinese culture, these analyses determine that part-time work does indeed cause a mild increase in deviance, but only for youth who are not well-monitored by their parents. Chapter 5 reviews selection into dating and marriage, and then investigates the relationship between dating and deviance and intent to marry and deviance. Limitations of the data prohibit a model of marriage and general deviance, but the chapter does include counterfactual models for the causal effect of marriage on smoking. Fixed effects models for marriage and smoking find that marriage is associated with lower smoking in much the same way that it is associated with lower deviance in the West (Sampson, Laub, and Wimer 2006), while the counterfactual models find that the relationship is in the general direction of marriage lowering smoking but is not statistically significant, either due to lack of statistical power or simply because the relationship is not causal in Taiwan.

For all these models to be feasible, an assessment of the reliability of self-reported deviance is necessary. After all, to what extent can we trust someone who has broken the law to admit that in an interview? The next chapter begins with that discussion prior to demonstrating how deviance can be measured in a survey in which the items measuring deviance constantly change.

## **Chapter 2: A Latent Variable Model for Modeling Deviance in Longitudinal Surveys with Changing Questions**

The present research approaches the study of crime from a practitioner's approach: crime must be properly explained so that its prevalence in society can be reduced through effective policies. In order for crime to be explained, however, it first must be measured. That is, criminologists must first define what crime is, then develop instruments that capture that concept in a manner that is reliable, valid, and reproducible. Only then can they strive to describe its prevalence, develop and test theories that explain its causes, and ultimately implement policies that lower it.

This chapter addresses an important practical problem for criminologists: researchers who seek to compare responses associated with the same person or that reflect a particular school or community over time, especially those conducting secondary data analysis, often find a need to calibrate their instruments used to measure crime change from year to year or differ by location. Perhaps a local police department may change how it defines a robbery, affecting official records<sup>4</sup>; or surveys involving self-reported crime may introduce new types of crimes, such as illegal downloading of music or "jailbreaking" a smartphone, that reflect broader cultural changes. These sorts of comparisons are almost always essential for research that involves causal inferences. A problem arises, however, when attempting to compare measurements taken with slightly differing instruments on a scale that makes sense.

---

<sup>4</sup> Official classifications of crimes are often less straightforward than they seem. The present researcher recalls one question posed by an official at a major police department in the United States during a public meeting to discuss crime prevention strategies: "If a prostitute is paid for her services and the check bounces, is that a rape or a robbery?"

The present research seeks to partially address this problem by focusing on a particular survey involving self-reported delinquency in which the survey questions change over time, and provides four contributions. Substantively, it quantifies the relative seriousness of types of delinquency as asked of Taiwanese youth. Additionally, it extends Osgood, McMorris, and Potenza's initial application of Item Response Theory (IRT), a methodology from educational assessment, to the study of deviance (2002), showing how vertical scaling, a technique usually used to compare test scores, can allow criminologists to model changes in individual deviance in longitudinal surveys in which the survey items change. Third, building on a model by Richard Patz and Brian Junker (1999b), it presents the use of Bayesian IRT models for criminological research for the first time, in a setting in which only the Bayesian model is valid, and provides an extension of Patz and Junker's open-source software for researchers with programming expertise to use. Finally, the present research improves upon the current methodology for vertical scaling by proposing an algorithm for improving the precision of vertical scaling techniques in Bayesian research settings. The chapter begins with a selective review of the literature on measurement of crime, focusing on studies of the reliability and validity of self-reported crime, and concluding with a discussion of Item Response Theory models as used in criminological research. It then proceeds with a discussion of the reliability and validity of the survey items used to measure deviance in the Taiwan Youth Project (TYP) survey, and presents details of the current model and methodological contribution.

### Literature Review: Measurement of Crime

Short and Nye (1957) addressed a major concern that contemporary criminologists had about the reliability of officially reported delinquency by simply asking people how much crime

they had committed. Placing “trap” questions in the survey designed to identify people seeking to paint an overly-rosy picture of themselves and assuming that respondents self-reporting an excessive number of crimes would not be at large, they found that only a small percentage of respondents completed surveys in a manner that suggested that they were understating or overstating their deviance. The results opened the doors for the study of crime using self-reported delinquency (SRD).

Two decades later, Hindelang, Hirschi, and Weis (1981) published *Measuring Delinquency*, a tour de force in which they carry out a careful analysis of the reliability and validity of crime as measured through official and self-reported data, in an attempt to evaluate the use of SRD for measuring crime. They note that more than 20 years after Short and Nye’s paper, no researchers had yet conducted a rigorous investigation of the reliability and validity of SRD measures. The following section reviews the literature covering the reliability and validity of self-reported delinquency as a research methodology, using *Measuring Delinquency* as a point of departure.

### *Reliability of Self-Report Responses*

The reliability of a survey instrument denotes its precision as a measurement tool (Huizinga and Elliott 1986:295). A common method for calculating reliability is Cronbach’s alpha reliability score (Cronbach 1951), which ranges from 0 to 1 and is used to determine the extent to which two or more survey variables measure the same underlying concept. A reliability of 0 entails that a set of survey items are completely unrelated, whereas a reliability of 1 indicates that the concept in question is perfectly measured by every survey item, such that a single survey item could replace the set of questions without loss of precision.

Hindelang, Hirschi, and Weis (1981) note that several factors matter in assessing the reliability of self-reported delinquency items. First of all, researchers must know whether the concept they seek to measure is unidimensional or multidimensional. Secondly, a respondent must give the same answer when asked the same question more than once (“test-retest reliability”). Third, the type of survey (questionnaire vs. face-to-face or telephone interviews) could affect the reliability of responses. Fourth, the degree of anonymity of the respondent may affect the veracity of their responses. Fifth, the type of scale used to measure delinquency (such as count data or broader categories such as “zero”, “once or twice”, “more than five times” and so on) could impact reliability. Finally, the reliability of a survey instrument may systematically vary based on respondents’ background characteristics, including level of education, race, gender, and social class.

Researchers constructing reliability survey instruments must know something about the dimensionality of crime. If criminals tend to specialize, focusing on some types of crimes and not others, then the alpha reliability scores for scales theorized to capture a general measure of delinquency would be lower than those associated with a particular domain. Researchers defining subscales for crime would therefore find their survey instruments to be more precise. Hindelang and his colleagues (1981) conduct an empirical analysis of “homogeneity among self-report items” (p. 46). Through their literature review and also carrying out factor and cluster analyses of data from a Seattle survey containing subscales for crime (official contact, serious crime, delinquency, drug use, and deviance related to the school and family), the authors find a high reliability within subscales, as well as a moderate correlation between scales. They conclude that criminals specialize, but that all forms of deviance are to some degree explained by a common thread. Following up on this line of investigation with an investigation of the first five

waves of the National Youth Survey, Huizinga and Elliott (1986) note that general scales for delinquency are still useful because they capture the breadth of an individual's behavior (and therefore may constitute more valid tests of a particular criminological theory), but because individuals to a degree specialize, reliability measures based on inter-item correlations within scales may be inappropriate for assessing the reliability of such scales calculated from self-report data, and test-retest measures may therefore be better assessments of SRD reliability.

Test-retest reliability measures for self-reported delinquency have been shown to be high. Huizinga and Elliott (1986) re-interview a subset of the National Youth Survey respondents to determine whether their responses change. They find high reliabilities for both individual self-report items and for scales comprised of these items, noting that test-retest reliabilities are particularly high for scales measuring serious crime, but still strong for general delinquency scales. Hindelang, Hirschi, and Weis (1981) also calculate measures of test-retest reliabilities for respondents self-reporting delinquency in Seattle, as part of a test of whether the reliability of self-reported delinquency responses varies by the type of survey. They randomly assign respondents to each of four survey types: anonymous face-to-face interviews (in which respondents sit on the other side of a screen and their identities are not known to the interviewers), non-anonymous interviews, anonymous questionnaires, and non-anonymous questionnaires. The authors find a high test-retest reliability among responses regardless of whether the survey was administered via interview or on paper, and regardless of whether respondents were anonymous. In so doing, they put to rest concerns that anonymity or survey methodology (interviews vs. questionnaires) affect the reliability of SRD instruments.

The scoring method of an SRD scale may also impact its reliability. Reviewing the literature, Hindelang, Hirschi, and Weis (1981), along with Huizinga and Elliott (1986:301) in a

literature review, note that scales comprised of variety scores (a measure of the number of different types of crime that an individual has committed) have higher test-retest reliabilities than scales based on taking questions in which a respondent reports the total number of times they have committed a particular type of crime, then summing the results. In practice, however, reliabilities for both types of scales are reasonably high; using alpha reliability measures for scales, Hindelang, Hirschi, and Weis (1981: 80-81) report reliabilities of about .9 for variety scales and about .8 for frequency scales, and these numbers persist after they stratify their sample by race, gender, and socioeconomic status.

Huizinga and Elliott (1986) caution readers that the skewness of self-reported delinquency creates a problem for assessing SRD reliability, because measurement error is higher for respondents reporting higher delinquency: those who commit many crimes over the course of a year may forget about some of them, while those who have committed zero can be certain of the exact value of their response. As a result, simple reliability scores based on correlations may be overestimates of true reliability, and this fact is particularly true for scales and especially for scales based on frequency scores rather than variety scores. As a result, the authors recommend that researchers assess test-retest reliabilities for individual items as well as for scales.

### *Validity of Self-Report Responses*

The validity of a measuring instrument denotes the extent to which its measurements are consistent with common sense and other data known to properly reflect the concept of interest. Face validity refers to the former: does a question about deviance appear to actually ask about a behavior that is deviant? Huizinga, Elliott, and Weis (1981) refer to three long-held approaches that researchers when defining the boundaries of what they consider delinquency when

constructing self-report measures of it. First, researchers could seek to reproduce official delinquency statistics in order to measure violations of the law among both those who have been arrested and those who have not yet been arrested. Second, their goal may be to test theories of deviance, in which case the measures must accurately map to specific theories. Third, they may wish to explain behavior perceived as socially harmful. In the second and third cases, acts that are not necessarily illegal may still be defined as deviant. Measures of deviance in the TYP survey, referred to as “problem behavior” in Mandarin by the research team, were most likely constructed following the third approach. Hindelang, Hirschi, and Weis (1981) note that face validity for the items in such surveys can only be ascertained on a case by case basis. Huizinga and Elliott (1986: 310) further caution readers that scales containing items with low face validity can easily lose face validity themselves, even though they may contain many items with high face validity. The reason is that often the behaviors with low face validity are not illegal and occur frequently, thereby dominating the resulting scale and masking the importance of more serious crimes which may have higher face validity.

A second type of validity, criterion validity, involves comparing self-report responses with other types of delinquency (such as official delinquency or that reported by teachers and parents) for the same individuals. Criterion validity can be either *concurrent* (in which the external data reflects the same time period as the self-report data) or *predictive* (in which the self-report data reflect a time period prior to that of the external data). Two methods for assessing criterion validity – correlational analyses and record checks – are prevalent in the literature. Correlational analyses involve comparing the correlation between officially reported delinquency and self-reported delinquency, and tend to yield somewhat low numbers due to the fact that many self-reported crimes are either never officially recorded or else are recorded outside the jurisdiction in

which a researcher is checking official records. Record checks, in contrast, involve verifying that every crime committed by a survey respondent that is officially recorded by the police is also self-reported.

Concerning correlational validity checks, Hindelang, Hirschi, and Weis (1981) find that self-reported delinquency is relatively highly correlated with measures of official delinquency such as court referrals – on the order of about 0.5, although the results vary by study. The lowest correlation between self-reported and official delinquency in the Seattle data occurs for black males at 0.3, although self-reported official contacts and official reports of delinquency are quite highly correlated (on the order of .7 to .8 in Seattle) regardless of race or gender. They also find a moderately high correlation between self-reported and official deviance regardless of survey type (anonymous interview, nonanonymous interview, anonymous phone, and nonanonymous phone) – although black males tend to give responses unrelated to official statistics in anonymous interviews and nonanonymous questionnaires. One implication for research using the Taiwan Youth Project surveys, which are administered in different years using different methods (face-to-face interviews, telephone interviews, and questionnaires), is that all of these methods are appropriate for measuring deviance. Huizinga and Elliott (1986), in a review, note that self-reported delinquency is not very highly correlated with official arrest data, with correlations generally around 0.3-0.5, but as official data concerning arrests is not necessarily measured well, perhaps these somewhat low positive correlations are as high as can be expected.

Record checks yield higher validity. In a review, Huizinga and Elliott (1986) find that more than 80% of official offenses are self-reported, although Hindelang, Hirschi and Weis (1981) find that this number varies substantially by race and seriousness of the crime, and to a lesser extent by gender. In particular, they find that white males do not report 20% of serious offenses

and 5% of nonserious offenses, black males do not report 57% of serious offenses and 18% of nonserious offenses, white females do not report 16% of serious offenses and 13% of nonserious offenses, and black females do not report 22% of serious offenses and 21% of nonserious offenses.<sup>5</sup> Digging further, they find that for whatever reason, black males who have low grades are more likely to underreport their crime. Concerning gender, they find that the ratio of the rate of male to female reported crimes in self-report data for individual items map closely to their equivalent ratios in official data, although the numbers become a bit more distorted when the individual items are aggregated into scales. Their conclusion is that self-reported delinquency is in general a valid research approach, although the method is questionable for black males, and in particular the approach might not constitute an effective way to compare the prevalence of delinquency across groups. Following up with a record check using the National Youth Survey, Huizinga and Elliott (1986:318) find that respondents overreport delinquency about as often as they underreport it, but black males tend to underreport official delinquency substantially more than other groups. More recently, using a longitudinal survey of serious offenders, Piquero et al. (2014) found a high concurrent validity of self-reported deviance that is stable over time and more or less invariant across race and gender. Using odds ratios, they show that respondents who appear in official records are many times more likely to self-report having committed the delinquent act. The authors also review the literature and note that 15 of 17 studies have found moderate to strong criterion validity for self-reported delinquency.

Elliott and Huizinga (1983) note the discrepancy between class differences in the prevalence of delinquency in official reports and the absence thereof that Hindelang, Hirschi, and Weis (1981) find in self-report data. They conclude that the self-report method is therefore still

---

<sup>5</sup> The authors examine a total of 1,877 offenses, with at least 100 offenses in each of the above categories except for serious crimes for females (16 for white females and 22 for black females).

questionable and merits further research. Using the first five waves of the National Youth Survey, a nationally-representative survey of delinquency, the authors show that the discrepancy can likely be explained by the choice of questions and scale in self-report surveys. They note that the use of scales in the measurement of delinquency often leads to situations in which one item with high incidence, usually a type of crime that is not serious, dominates all the others. Additionally, some crimes are more common among the working classes while others are more common among middle-class respondents, leading to situations in which suppressor effects may dilute the apparent importance of class in a delinquency scale. Constructing a number of subscales for crimes such as theft (serious and minor), assault (serious and minor), robbery, and status offenses, they find class differences in several of the subscales, with stronger differences for more serious crimes, but not in their general delinquency indices. These differences between the indices persist regardless of race or gender. They offer a cautious conclusion that measures of serious delinquency may be more valid than general delinquency scales, at least to the extent that researchers believe that the official collection of and reporting of nonserious crimes is both unbiased with respect to class and contains low measurement error.

Few studies have been conducted assessing concurrent validity of self-report measures for Asian respondents in particular. Gould (1969) is one exception, in which the author compares the correlation between official reports and self-reported deviance for 104 Caucasian, 221 African-American, and 38 Asian-American 7<sup>th</sup>-grade students in two Seattle middle schools. Using Goodman and Kruskal's gamma (Goodman and Kruskal 1954), a measure of association between ordinal variables scored on a scale similar to Pearson's R, he finds gamma correlations between self-reported and official delinquency of 0.34, 0.09, and 0.07, respectively, for the three groups. Given the low correlational validity for Asian-American students and the low prevalence

of official and self-reported delinquency for this group, Gould concludes that cultural expectations within the Asian American community that emphasize that Asian children should not and do not commit crimes may lead youth to self-identify as not deviant and thus literally forget minor commissions of deviant acts. His methodology, however, suffers from both a small sample size and the use of correlational validity instead of a records check. Using records checks, Huizinga and Elliott (1986) find that respondents belonging to a broad ethnic category of “other” that includes Asian Americans report delinquency known to the police as often as Caucasians do. Using a longitudinal survey from Seattle, Jolliffe et al. (2003) are able to assess both concurrent and predictive validity of self-report data (their measure of predictive validity involves determining whether respondents with court referrals reported committing the delinquent act during the previous year, presumably before they had been officially charged). They find strong validity for most respondents – with black males’ self reports some of the most valid – but the lowest validity occurs for Asian females, while Asian males have reports with high validity. The authors test whether respondents who state they committed a crime in the past year (so-called “prospective validity”) also acknowledge having committed the crime years later (“retrospective validity”). This percentage is greater than 66% for white males and females, black males and females, and Asian males, but only 39.1% for Asian females (a highly statistically significant difference). The sample of Asian (American) respondents contains mostly those of Chinese and Filipino heritage. These two studies provide some initial evidence that perhaps self-reported measures of delinquency might be less valid in the case of Taiwan than for North America and Europe, but nonetheless Gould and Jolliffe et al. use Asian Americans as their unit of analysis, and most of those of Chinese descent are likely to have come from the People’s Republic of China, which is culturally distinct from Taiwan.

An indirect way to assess the validity of self-reported deviance, absent better measures such as record checks or concurrent validity studies, is to compare self-reported delinquency to self-reported victimization. One comparison of 862 Chinese youth in Hangzhou and 2,041 American youth in five U.S. cities (Webb, Ren, Zhao, He, and Marshall 2011) finds that Chinese youth report less victimization than their American counterparts, but far less delinquency; 1% of Chinese respondents who were not in street gangs reported being assaulted during the past year compared to 4% of U.S. non-gang respondents, while 0.12% of the same Chinese sample compared to 1% of the U.S. sample reported assaulting someone during the past year. For the ratio of the rate of offense to the rate of victimization to be half in China of what it is in the U.S. suggests either underreporting of crimes in Chinese self-report surveys or that more crimes in China are committed by just a few individuals compared to the United States, certainly a possibility given the recent success of the COMPSTAT policing model for identifying and arresting offenders responsible for large numbers of crimes (Bratton and Knobler 1998; Maple 2000; Zimring 2007).

Despite lower absolute numbers of reported crimes in China, self-report surveys have nonetheless been commonly used to test major theories of the etiology of crime in a Chinese context (Bao, Haas, Chen, and Pi 2014; Bao, Haas, and Pi 2007; Cheung and Cheung 2010; Jessor, Turbin, Costa, Dong, Zhang, and Wang 2003; Lin, Dembo, Sellers, Cochran, and Mieczkowski 2014; Zhang and Messner 1995; Zhang and Messner 1996; Zhang and Messner 1999; Zhang, Messner, and Liu 2007). Absent more rigorous reliability or validity checks in China, the assumption underlying all of these studies is that even if Chinese respondents do underreport their involvement in crime, as long as the mechanism causing the underreporting is orthogonal to the mechanisms causing the actual crime, regression parameters are still unbiased

and criminological theories can be tested. The present research uses concurrent validity checks, however, as a way to demonstrate the suitability for research of the IRT scores for deviance that follow.

### *The Use of Item Response Theory in Criminological Research*

The debate concerning how to appropriately devise scales to measure crime is still inconclusive. In their analysis of the effect of class on delinquency, for example, Elliott and Huizinga (1983) favor scales based on specific counts of deviance over items with broader categories such as “seldom,” “sometimes,” and “often,” claiming that count data provides more information. Just two years later, however, using the same data set they use categorical responses for theory testing “because they have better distributional characteristics and are less skewed” (Elliott, Huizinga, and Ageton 1985:97). Hindelang, Hirschi, and Weis (1981:80) directly compare the reliability of “ever-variety” scales (scales comprised of variables with just two categories, “ever” and “never”) with scales based on counts, and find the variety scales to be more reliable (on the order of .9) but the frequency scales to still be reliable (on the order of .8). With either type of scale acceptably reliable to most researchers, the remaining issue is one of interpretability of the resulting scale: how does one actually interpret a general delinquency measure associated with three thefts, one speeding ticket and ten days of playing hookie with another measure associated with two thefts, one day of truancy, and one burglary? A variety scale would score these two individuals as equally delinquent; a frequency scale might score the first respondent as more delinquent, unless the crimes were weighted by severity – yet criminologists have not reached an agreement on how to conduct or interpret such weighting (Osgood, McMorris, and Potenza 2002). Similarly, suppose that the two respondents were

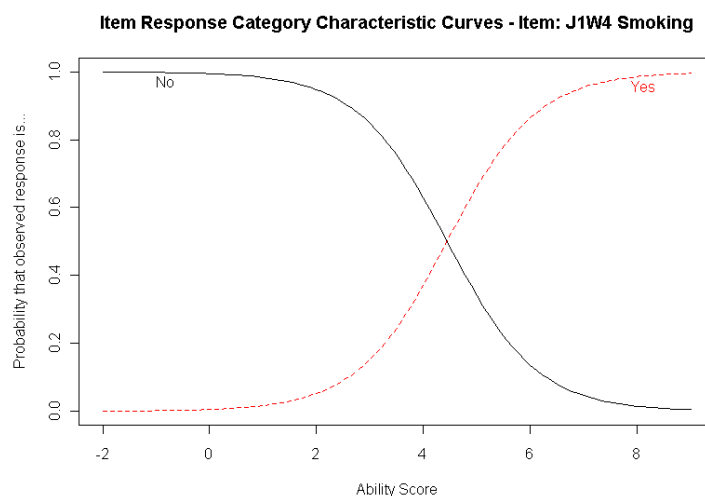
interviewed during different decades, such that the first was asked if he was “truant” and the second was asked if she “played hookie.” The literature on measuring criminology has nothing to say about comparing such responses, perhaps because the problem of simply constructing scales is hard enough.

Osgood et al. (2002) address these issues by applying Item Response Theory (IRT), an approach to constructing measures using nonlinear latent variable models (Lord 1952; Rasch 1960; Samejima 1969), to criminology for the first time. IRT models, which have been most widely used in educational assessment and psychometrics, assume that survey or examination questions have an inherent difficulty, and individuals responding to the questions have an ability to respond to them. The probability that a respondent correctly answers a test question is then a function of the item difficulty and the individual’s ability. Equation 2.1 gives a simple example of an IRT model, the 2-parameter logistic model, used to model responses to test questions, such as true-false questions, that have exactly two categories (Baker 2001:21):

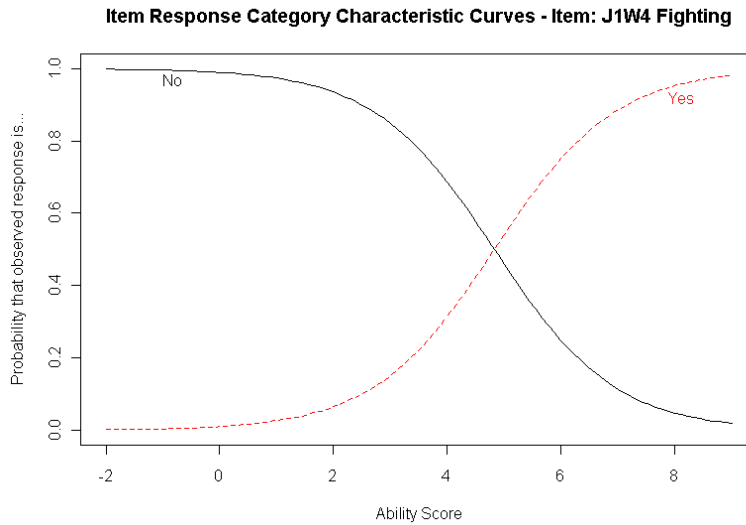
$$P(X = 1) = \frac{1}{1 + e^{-a(\theta - b)}}. \quad (\text{Equation 2.1})$$

In Equation 2.1,  $P(X=1)$  denotes the probability that a respondent correctly answers the question,  $\theta$  denotes the respondent’s ability level, and  $b$  denotes the difficulty of the test question. The greater a respondent’s ability to answer the question, the greater the value of  $P(X=1)$  in Equation 2.1, and the greater the difficulty, the lesser the value of  $P(X=1)$ . The parameter  $a$  (always positive) is known as the *discrimination parameter*, and is so named because high discrimination parameters have a higher rate at which small changes in individual ability or item difficulty affect the value of  $P(X=1)$ , and hence they are better able to “discriminate” between individuals with high and low ability.

Equation 2.1 can then be complicated in various ways, such as by allowing  $\theta$  to be multidimensional instead of simply unidimensional, as might be the case for a test question requiring a respondent to demonstrate both reading comprehension and critical thinking. Another common variation in the case of social science research involves respondents answering questions on Likert scales, such that the greater a respondent's score on a latent construct (such as depression), the more strongly he or she would agree with a question measuring that construct (such as "I often feel depressed.") The distinction between item difficulty and discrimination parameters is often presented graphically with a question's Item Characteristic Curve (Figure 2.1). ICC curves are obtainable for a wide class of IRT models, and allow a researcher to quickly understand the relationship between item discrimination parameters, item difficulty parameters, and the probability that a respondent will answer a test or survey question in a particular way. These plots are not associated with individual respondents, but rather with the behavior of a particular survey item for the entire sample.



**Figure 2.1.** Item Characteristic Curve (ICC) associated with the dichotomous question about smoking (with an item discrimination parameter of 1.19) in Wave 4 of the Taiwan Youth Project Survey for the J1 cohort.



**Figure 2.2.** Item Characteristic Curve (ICC) associated with the dichotomous question about fighting (with an item discrimination of 0.94) in Wave 4 of the Taiwan Youth Project Survey for the J1 cohort.

Figures 2.1 and 2.2 show two Item Characteristic Curves computed using data from the Taiwan Youth Survey and Muraki’s Generalized Partial Credit Model (Muraki 1992), an approach well-suited for Likert-style questions that, for dichotomous questions such as those in the figure, closely matches Equation 2.1 in its mathematical formulation. The y axis denotes the probability that a respondent answers a survey question in a particular manner (“yes” for the dashed line and “no” for the solid line), and the scale for the x axis is in standard deviations of respondents’ ability scores, with positive numbers denoting the number of standard deviations above the mean and negative numbers denoting standard deviations below the mean. Thus, the average Taiwan Youth Project survey respondent is quite unlikely to report having smoked or gotten in a fight in this particular wave (associated with 10<sup>th</sup> grade). The item difficulty parameter is associated with the point on the x axis at which a respondent becomes more likely to answer “yes” to the survey question than “no”; in Figures 2.1 and 2.2, this point is slightly further to the right, indicating a question that is more difficult to answer. The discrimination

parameter affects the steepness of the two lines, with the dashed line taking a shape similar to a logistic curve. The greater the discrimination parameter, the steeper the ICC curve is, and the better the precision of the question in determining an individual's exact ability. (One possible explanation for the relative lack of precision for the fighting variable might be that some respondents who said they were in fights were not the aggressors.) Osgood et al. (2002) note that item discrimination parameters are comparable to factor loadings in a confirmatory factor analysis, except for the fact that factor loadings represent the level of precision of a survey item in capturing a latent factor regardless of respondents' characteristics, while an IRT question may provide a great deal of information about respondents with an ability corresponding to the steepest gradient of the ICC curve, but little information about respondents with abilities corresponding to sections of the ICC curve that are nearly flat. Thus, as might be expected, fighting in this example is more able to distinguish among respondents of high and low criminality, as might be expected of a delinquent behavior that at face value appears to be more serious.

Osgood and his colleagues (2002) were the first to notice that these psychometric models can directly be applied to measure deviance, such that ability scores map to a latent measure of an individual's criminality, and item difficulties constitute straightforward, data-driven measurements for the severity of a crime. Using a sample of 9,596 person-years from the Monitoring the Future survey, the authors implement Samejima's Graded Response Model (Samejima 1969), demonstrating that item discrimination parameters map to criminologists' general beliefs about the severity of crimes, with offenses such as robbery and arson associated with higher discrimination parameters than offenses such as trespassing and theft of \$50 or less. Additionally, their measures are categorical, and the Graded Response Model allows them to

associate item difficulties with specific categories. They demonstrate that, as expected, the more serious a crime, the higher the initial difficulty associated with reporting at least one offense. They also demonstrate a smaller difference in thresholds associated with going from reporting three or more crimes to five or more crimes of a particular type than from going from one to two. Additionally, they can demonstrate that as the difficulty associated with reporting five or more crimes usually implies that respondents doing so must have criminality scores that are more than three standard deviations above average, attempts to gain additional precision at the high end of these scales are likely to be unsuccessful, thus validating the truncation of reported crime at five counts for their particular data set. All of these results are consistent with the literature. Finally, they point out (p. 287) that the Graded Response Model accurately distinguishes among responses for the most delinquent 7% of respondents, assigning to them ability scores of between 1.5 and 3.5 on a standardized scale – a much-needed level of precision to replace highly skewed responses on deviance scales that more often than not lead to influential outliers in regression analyses.

In short, the SRD method has been shown to be a reliable and valid approach to the study of delinquency, with some caveats: first, researchers must examine delinquency items for their face validity, and second, additional research needs to be conducted to assess the validity of self-reported delinquency in Taiwan. Further, Item Response Theory constitutes a valid, reliable approach to measuring crime that offers a solution to previous problems with interpreting delinquency scales and incorporating differential seriousness of crimes into measures of individual deviance.

The present chapter makes three contributions to research on the measurement of crime, and one contribution to Item Response Theory methods in general. Concerning the measurement of

crime, it extends the application of Item Response Theory to cover vertical scaling, an approach commonly used in educational assessment to allow practitioners to directly compare ability scores for the same person taking different tests and to directly compare the difficulty of questions on different tests when the same people take both tests (Kolen and Brennan 2004). Osgood et al. (2002: 277) note that vertical scaling methods “have little bearing on our aims, however, which concern the use of measures in social research rather than for individual testing or diagnosis.” In fact, vertical scaling techniques allow criminologists to compare measurements of individual criminality across waves of longitudinal surveys in which the questions change – an approach not possible with other methods. Second, this chapter presents a situation in which the normal assumptions of frequentist IRT models are violated, and only a Bayesian approach is valid. Accordingly, it serves as a caution to researchers using this approach to carefully test model assumptions before fitting IRT models. Third, item discrimination and difficulty parameters are provided for measures of crime in a Taiwanese sample. Finally, of interest to IRT methodologists in general, this chapter presents an algorithm for improving the precision of vertical scaling techniques in Bayesian research settings. This chapter continues with a brief introduction to vertical scaling in order to motivate the use of IRT models in the present research, and then discusses the reliability and validity of the Taiwan Youth Project delinquency items in light of the previous literature review before presenting the specific IRT model used here.

#### *Motivation for Using Item Response Theory and Vertical Scaling in the Present Research*

Table 2.1 shows the wording of the delinquency items in the Taiwan Youth Project used for the analyses in the present research. The precise wording of these items along with their

corresponding categories in Chinese, with the English translation, is also detailed in Appendix A in Tables A-1, A-2, and A-3. The fact that no single item in Table 2.1 remains on the survey for all waves constitutes a significant challenge for analysis. A model of the contemporaneous effect of work on crime for the J3 cohort, for example, would use crime in Waves 3, 6, and 7 and could not be fit if the same items were used to measure crime in all three years. A model of the lagged effect of work on crime would use crime in Waves 4, 6, and 7 and may be possible, but would restrict the variables used to measure crime to theft and drug use. Very few people report theft and drug use compared to the other items on this list; out of more than 1800 respondents, the number of people reporting theft, for example, did not exceed two dozen in any of these years. The result is very little variance in delinquency to explain.

The good news, however, is that Item Response Theory (IRT) methods not only can allow criminologists to obtain more precise estimates of self-reported criminality and the relative seriousness of self-reported crimes (Osgood, McMorris, and Potenza 2002), but through a process known as “vertical scaling” (e.g. Kolen and Brennan 2004), ability scores for the same individual in different survey years can be placed onto the same scale to investigate the etiology of change in criminality even if the survey items themselves change. Such rescaling is possible as long as waves of the survey that are linked together using this procedure always have at least one common item. This approach also has the benefit of allowing all survey items in a wave to be used for analysis, even those items that appear in only a single survey year (they would be used to measure crime but not for vertical scaling). The procedure requires making an assumption of invariance of the discrimination and difficulty parameters for the same survey item appearing in multiple surveys. In the field of educational assessment, vertical scaling is often used for purposes such as test security, for two students seated next to each other could

**Table 2.1.** TYP delinquency items used in this dissertation. Wave abbreviations include the cohort followed by the wave, such that “J3W1” refers to Wave 1 of the J3 Cohort, and so on.

Item	Scale Type	J3W 1	J1W 4	J3W 2	J1W 5	J3W 3	J1W 6	J3W 4	J1W 8	J3W 6	J1W 9	J3W 7
Battery	Frequency	X		X								
Non-prescription drug use	Dichotomous			X								
Smoking	Dichotomous		X	X	X	X						
Drinking	Dichotomous		X	X	X	X						
Smoking	Frequency						X					
Drinking	Frequency						X					
Smoking or drinking	Frequency	X						X				
Causing trouble at school	Dichotomous		X	X	X	X						
Running away from home	Dichotomous			X		X		X				
Running away from home	Frequency	X					X					
Skipping class	Dichotomous		X	X		X						
Skipping class	Frequency	X										
Theft	Frequency	X		X			X	X	X	X	X	X
Drug use	Frequency	X		X			X	X	X*	X	X*	X
Drug use	Dichotomous					X						
Fighting	Dichotomous		X		X	X	X	X				
Vandalism	Frequency	X							X	X	X	X
Fraud / cheating	Frequency								X	X	X	X
Playing Mahjong (gambling)	Frequency								X	X	X	X
Drag racing	Frequency								X	X	X	X
Pornography	Frequency								X	X	X	X
Chewing betelnut	Frequency	X							X		X	
Illegal downloading	Frequency										X	

\* The question was worded using a five-item frequency scale, but recoded to dichotomous because too few respondents stated that they had been involved in the behavior with high frequency.

receive copies of standardized tests containing entirely different questions, and their scores could be directly compared so long as a subset of questions were the same. In the present case, the situation is the same as the test security example, except that instead of comparing the scores for two respondents given different surveys at the same time, the focus is on comparing scores for the same respondent given different surveys at different times. As this technique is unique to IRT models, it not only presents a useful and informative way to model delinquency, but is absolutely essential. Of course, for the TYP survey items to measure a latent deviance construct they must exhibit face validity and reliability, so the present chapter proceeds with a discussion of the reliability and validity of these items.

#### *Face Validity of the TYP Survey Items*

Although the items in Table 2.1 are all mild forms of deviance, a delinquency item's low severity does not imply that it lacks face validity as a measure of delinquency. Indeed, Gottfredson and Hirschi (1990) make a strong case that general criminality predicts a wide variety of different crimes, both mild and serious, as well as a variety of status offenses and behaviors, such as drug, alcohol, and tobacco use, that are not necessarily illegal but are nonetheless more likely to be observed among those with high criminality. Furthermore, Tables A4 and A5 in Appendix A demonstrate that relatively few Taiwan Youth Project survey respondents self-report engaging in even mild forms of delinquency (such as drinking or smoking) – and some survey items, such as drug use and running away from home in Wave 4 for the J1 cohort, could not be included in this research because not enough respondents admitted to the behavior. Accordingly, questions about mild forms of deviance make sense for the Taiwan

Youth Project survey given the characteristics of the sample, as these items are useful for measuring criminality among youth with only a modest tendency towards deviance.

Many of the deviant behaviors in Table 2.1 have face validity because they involve breaking the law or school rules – including theft, drug use, vandalism, getting in fights, and illegally downloading music. In Taiwan, the “fraud” question involves cheating on tests or other forms of dishonesty, with a typical example of the latter often involving a respondent lying to one’s parents about where he or she has gone or how he or she has spent lunch money that day. An answer of “yes” to this question usually involves deceiving someone in a position of authority, a behavior that is deviant anywhere and particularly in Chinese culture, where respect of one’s parents and teachers is emphasized. Skipping class, running away from home, and “causing trouble” in school, for similar reasons, are considered deviant, and Hindelang, Hirschi, and Weis (1981:63) include these questions in an index of school and family offenses. Additionally, the legal age for smoking, drinking, and viewing pornography in Taiwan is 18 years old, so these activities clearly have face validity as delinquent acts for minors. Moreover, according to Gottfredson and Hirschi (1990), these activities retain their face validity as measures of criminality among those old enough to legally engage in these activities, as they reflect the absence of self-control, a predictor of crime, and are correlated with other forms of deviance.

Three types of deviance – playing Mahjong, chewing betel nut, and motorcycle racing – deserve special discussion because they are somewhat culturally unique to Taiwan. The Chinese term for motorcycle racing (飆車) does not usually involve racing against a friend, but rather roaring out of an intersection after waiting for a red light, or speeding in order to ride faster than another stranger just for the thrill of it. As motorcycle racing involves risk-taking, a sense of

excitement, and is a public action, Gottfredson and Hirschi (1990: Chapter 5) would most certainly describe it as deviant.

Mahjong is a type of game with tiles that involves four players and almost always involves money. Gambling is technically illegal in Taiwan, but in practice many Taiwanese play Mahjong with family members on special occasions, such as Chinese New Year. Accordingly, depending on the respondent's family traditions, playing Mahjong just once or twice may not be particularly deviant at all, although playing frequently would be more suggestive of a low-level degree of impulsivity and risk-taking regardless of holiday customs. Perhaps a better wording of the question might involve "playing Mahjong for money" or "playing Mahjong (other than during holidays)", but given the question is always asked with a frequency scale containing "sometimes," "often" and "all the time" as possible responses, the survey item as a whole retains face validity as an instrument capturing mild deviance. The validity of the item might additionally be more of a concern if the majority of respondents to the TYP survey reported playing Mahjong at least once or twice, but the distribution of responses to the questions about Mahjong in Table A4 maps that of the other items in the survey, suggesting that it does represent a measure of low-level deviance.

Betel nut chewing is not illegal in Taiwan for minors, but is widely seen as deviant. Betel nut is a stimulant analogous to chewing tobacco with mildly addictive properties, and is often associated in Taiwan with blue-collar workers such as truck drivers and construction workers, who chew it to stay alert. Socially, it is seen as undesirable because, just as with chewing tobacco, using it involves chewing the nut with large motions of the jaw and spitting out the unused part of the nut on the ground. Youth may chew betel nut in order to feel like adults, but due to its association with blue-collar workers, the action may be seen as a rejection of the

modern work ethic of social advancement through hard work in school. Additionally, girls may be particularly unlikely to chew betel nut because long-time users often have mouths that are permanently stained from the juice. Betel nut use has additionally been linked to oral cancer, so its use is discouraged for public health reasons. For all these reasons, betel nut use is considered deviant in Taiwan, so much so that in fact so few people reported chewing betel nut during some survey years that the survey responses could not be used to fit an IRT model.<sup>6</sup>

In summary, all the delinquency items in the Taiwan Youth Project survey, except perhaps for a single category of playing Mahjong, have high face validity. The present chapter continues with a discussion of survey item reliability, using Confirmatory Factor Analyses (CFAs).

#### *Reliability of the TYP Delinquency Items*

Table A6 shows individual factor loadings associated with confirmatory factor analyses for each cohort-wave of the TYP survey, along with alpha reliabilities reflecting the internal consistency of the survey items were they to be directly used to create scales. The standardized factor loadings in Table A6 suggest that most of the survey items are only modest measures of the underlying latent construct for deviance, with factor loadings above 0.8 in only a few cases. Similarly, the error variances in Table A6 denote the percentage of variance in the latent variable for criminality left unexplained by an individual survey item; the fact that few individual items explain more than half of this variance suggest that the latent construct is a composite of multiple measures of deviance. Hindelang, Hirschi, and Weis (1981: Chapter 2) demonstrate that even as some specialization occurs among criminals, most crimes are modestly predictive of other crimes, suggesting that even as criminologists studying high-risk populations can devise

---

<sup>6</sup> One other reason betel nut chewing may not be prevalent in the TYP survey is that betel nut use is (anecdotally) more prevalent in southern Taiwan, and the TYP sample is limited to northern Taiwan.

subscales to measure serious crime with high reliability, they can also capture a single underlying dimension for deviance in the general population with scales for general delinquency comprised of a wide variety of items. The factor loadings and error variances in Table A6 are therefore consistent with a successful implementation of the latter approach by the TYP research team, with a diverse set of survey items capturing a latent dimension for criminality.

The alpha reliabilities of raw scales derived from the measures for each wave also suggest that the individual survey items are not particularly strong measures of general deviance, as they usually are between .55 and .65, and Waves 2 and 4 of the J3 cohort have reliabilities between .3 and .4. While Hindelang, Hirschi, and Weis (1981: 59-68) generally report alpha reliabilities of between .7 and .9 for serious crime and drug use, they do note a lower reliability for an index of school and family status offenses that includes skipping classes, running away from home, and causing trouble at school. For this index they observe a reliability of .62.<sup>7</sup> Consistent with these findings presented in *Measuring Delinquency*, many of the TYP survey items with low factor loadings in Table A6 are school and family offenses, including skipping classes, causing trouble at school, staying away from home without permission, and cheating on tests or otherwise using deception (presumably with respect to one's parents or teachers). In other words, even though the internal reliability measures in Table A6 may be somewhat disappointing given the much higher internal reliabilities of SRD scales of serious crime and drug offenses in Western studies, they are nonetheless entirely consistent with Western research.

In the present case, however, alpha reliabilities may provide a poor estimate for the extent to which the delinquency items measure what they are intended to measure. Cronbach's alpha

---

<sup>7</sup> The authors observe a reliability of .62 for males, which comprise about 75% of their sample, and a higher reliability of .71 for females – but the latter reliability is calculated for a subset of the test questions after dropping from the scale some items exhibiting low communality with the other items in the index, and the authors do not report initial reliability measures based on the raw questions for females.

provides a lower bound on reliability, a bound which can be a substantial underestimate of reliability when the number of items is small and the items are heterogeneous (Lord and Novick 1968:91), both of which are true for the TYP data. In addition, alpha is generally less than or equal to the reliability of a measure, but this bound becomes an equality when the *true scores* associated with each item, or “expected observed score with respect to the propensity distribution of a given person on a given measurement” (Lord and Novick 1968:30), are all linearly related (Lord and Novick 1968:90). This linear relationship makes sense if a set of measurements are Tau-equivalent (i.e. they have exactly the same true score), as in the case of test-retest reliability. Indeed, Hindelang, Hirschi, and Weis (1981) note that given the breadth of different forms of deviance that are often incorporated into deviance scales, test-retest reliabilities are a more salient measure of SRD item reliability. Unfortunately, such information is unavailable for the TYP survey. Particularly for waves such as Wave 2 for the J3 cohort that contain both dichotomous and polytomous items, there is no reason to believe that the expected values of the responses associated with widely heterogeneous offenses – some of which very few respondents report committing and some of which are quite prevalent – would at all be linearly related. Hence, Cronbach’s alpha provides little information about the actual reliability of a test measure.

One additional detail about the factor analyses is noteworthy: mahjong exhibits one of the lowest loadings, good news given the previous discussion about its face validity. Were mahjong to have a high factor loading, then there might be greater concern that the potential invalidity of the “once or twice” category for that particular measure could be unduly impacting the validity of the overall scale. The low factor loading also suggests that the variable could be omitted from the present analysis completely, but I choose to leave it in as gambling is not captured by any other survey questions and preserving the item itself leaves open more possibilities for placing

ability scores in different survey waves onto a common scale using the vertical scaling methodology presented later in this chapter.

In conclusion, the Taiwan Youth Project survey team has placed particular emphasis on low-level deviance, especially school and family offenses, due to previous results showing low rates of reporting more serious types of crime among Asian respondents and the particular salience of school and the family in explaining why Taiwanese youth do not offend. In doing so, they have implemented adequate measures of juvenile delinquency based both on local cultural practices and the literature on measurement, but the literature has also shown that the particular signal being measured in this case is a weak one. Accordingly, much of the methodological rigor in the rest of this dissertation is geared towards preserving as much of this signal as possible.

One question related to reliability left unaddressed thus far is whether the TYP questions are capturing a measure of criminality that is merely unidimensional or possibly multidimensional. Most classes of IRT models additionally can be specified as having unidimensional or multidimensional latent ability scores, but whatever number of dimensions is specified, that specification then becomes one of the model assumptions. Accordingly, this chapter next addresses the question of dimensionality explicitly through confirmatory factor analyses.

#### *Verifying Unidimensionality of the Latent Variable for Criminality*

The dimensionality of the latent deviance construct for each cohort-wave can be shown using confirmatory factor analyses. For the present research, the LISREL 8.54 software package and its accompanying pre-processing software PRELIS were used to first compute polychoric correlations among the observed crime variables, then use that correlation matrix to run two confirmatory factor analyses using Weighted Least Squares (WLS) (Byrne 1998): one assuming

that all crime variables measure a single latent variable, and a second allowing for two latent variables. The procedure involves assigning specific delinquency items to specific dimensions based on a theory; in the present case, as subscales for substance use are often used in criminological research and smoking, drinking, and drug use are fairly ubiquitous items in the TYP survey, I designated one dimension for drug use and a second for other measures of deviance.<sup>8</sup> Technically, a formal test of dimensionality in LISREL would involve setting the scale for each of the two latent factors to be measured in terms of standard deviations by fixing its variance to 1.0 (such that  $\phi_{11} = \phi_{22} = 1.0$ ), then estimating all the unstandardized factor loadings in the second column of Table A6 instead of estimating one of them. The resulting models would produce the standardized factor loading estimates in the first column of Table A6, and the LISREL parameter  $\phi_{12}$  would yield the correlation between the two latent factors. An additional model fixing  $\phi_{12}$  to 1.0 would effectively estimate a one-dimensional model, and a test of the difference in the chi-square fits for the two models on one degree of freedom would constitute a formal test of whether a second dimension explains any additional variability in the survey items. That said, such a test is quite strong, and highly correlated factors that explain slightly more variance in the survey items than a single factor might not really represent two substantively distinct dimensions of criminality. Additionally, with the test repeated ten times, one or two tests may lead to rejecting the null hypothesis that the two latent factors have a correlation of 1.0 simply due to random chance alone. Four factors were therefore considered in assessing the dimensionality of the latent construct: a comparison of the Root Mean Squared Error of Approximation (RMSEA) for overall model fit for both the single-factor and two-factor models for each wave; a comparison of the chi-squared fits of the two models; the correlation between the two factors in the two-dimensional model; and the value of the Parsimony Normed

---

<sup>8</sup> Other specifications of dimensions for the J3W6 cohort led to the same conclusions presented here.

Fit Index (PNFI), which penalizes complex models. For all ten cohort-waves in the study, the model comparisons pointed strongly to the fact that the two-dimensional model was not explaining more about delinquency than the one-dimensional model. The RMSEA for the model with a single latent variable was less than 0.05, indicating a good fit of the data to the model with just a single dimension (as well as the presence of at least one dimension), and the RMSEA for the model with two dimensions was only marginally lower. Additionally, the decrease in the chi-square value for the model fit associated with relaxing the unidimensionality constraint was always less than 3.84, and the two factors were 90% correlated or higher. Finally, the value of the Parsimony Normed Fit Index (PNFI) was always less for the model with two latent variables compared to an equivalent model with just one, suggesting that the simpler, one-dimensional model is a more parsimonious fit. Accordingly, the following research is based on the conclusion that for all waves of the survey involving delinquency, a latent variable for criminality is unidimensional.<sup>9</sup>

#### *Selection of the IRT Model to Use*

The present research makes use of a Bayesian Generalized Partial Credit Model (GPCM) to model the TYP delinquency items. Frequentist IRT models have previously been used to model delinquency (Osgood, McMorris, and Potenza 2002; Osgood and Schreck 2007; Pyrooz and Decker 2013), but in the present case a Bayesian approach is the only viable one, and also constitutes a first usage of such models in criminology. This section discusses the functional

---

<sup>9</sup> The factor analytic approach used here uses only the correlations among the items to assess unidimensionality, but see also Swygert, McLeod, and Thissen (2001) for a presentation of using all information associated with the frequencies of response category patterns to conduct a factor analysis directly based on IRT using the software application POLYFACT. Hambleton, Swaminathan, and Rogers (1991:56) also provide a list of other approaches, including eigenvalue plots, for assessing unidimensionality. POLYFACT (as reviewed in 2001) uses exploratory factor analysis, while for theoretical reasons I prefer to use confirmatory factor analysis, dividing the questions about crime into dimensions reflecting substance use and all other crime.

form for this model and the justification for its use. Model selection involves the choice of a frequentist or Bayesian approach, a fixed or variable discrimination parameter, and the specific functional form to use given the categorical nature of the responses.

The choice of a Bayesian model is necessitated by characteristics of the data. In the field of educational assessment, maximum likelihood estimation (MLE) of a respondent's precise latent ability is impossible in the case of perfect test scores – either answering all the questions correctly or getting them all wrong – because the person's ability is not located within the range of ability scores captured by the test items. Some software packages for IRT estimation provide a partial solution to this problem by leaving individuals with all questions answered incorrectly out of the analysis when estimating the model parameters, then assigning to those individuals an ability slightly lower than that associated with only answering only the easiest test question correctly (Scientific Software International 2003). In the present case, most of the respondents in any given wave – 61.7% of 1,841 respondents for Wave 6 of the J3 survey, for example – reported committing no deviant acts at all, and removing all the respondents committing zero crimes from the study distorts the latent construct among the remaining cases to the extent that it becomes multidimensional instead of unidimensional. Accordingly, using a frequentist approach to fit an IRT model in this situation is simply infeasible.<sup>10</sup>

Bayesian estimation, however, allows for inclusion of perfect zero scores in the model, as an informative prior distribution provides enough information for these scores to be fit. Note that the prior *must* be informative in this case: a model with a completely flat prior that provides no

---

<sup>10</sup> A high percentage of respondents reporting no delinquency is not a problem unique to the current study. While IRT analyses of self-report data from Chinese samples are still uncommon, Pyrooz and Decker (2013) report that 44% of a sample of 2,245 students in a city in Northern China also reported that they committed no deviant acts. And in a survey of 3,500 U.S. high school students, Osgood and his colleagues (Osgood, McMorris, and Potenza 2002) find that 37% of respondents report zero delinquency. These authors may have retained a unidimensional latent construct even after eliminating the respondents reporting zero deviance from their data sets, but unfortunately they do not state whether this was in fact the case in describing their methodology. The present research brings attention to the importance of doing so.

information about the distribution of abilities on a latent scale is as unable to identify abilities for respondents reporting all zeroes as its MLE counterpart. The challenge of using Bayesian IRT models with user-defined prior distributions, however, is that researchers must currently write their own software routines rather than work with out-of-the-box statistical software, and the open-source R software is therefore popular among Bayesian researchers (Fox 2010).

IRT models can be broadly grouped into those belonging to the Rasch family and all others, and the selection of a model family has to do with the aims of the researcher and the characteristics of available data. Rasch models do not incorporate discrimination parameters into their functional form: the two-parameter logistic model in Equation 2.1 would belong to the Rasch family, for example, if  $a$  were fixed to the same constant for every test item. Rasch models have advantageous mathematical properties in that the item difficulties and individual responses are independent (Rasch 1960), a concept referred to as *specific objectivity* (Wright and Masters 1982). A major benefit of specific objectivity is that estimated item difficulties are invariant across data sets, so they could be applied to different samples in the same general time and cultural setting.<sup>11</sup> In fact, many researchers in psychometrics who recommend the use of the Rasch model take the somewhat unusual approach in research of suggesting that practitioners designing test questions exclude from their surveys items that do not fit the Rasch model, while retaining those that do (Bond and Fox 2007). Their general point is that Rasch models reflect a tool for measurement in much the same way that a ruler measures length, and one does not reject the ruler as an instrument simply because a particular object is difficult to measure with it. This

---

<sup>11</sup> As an example of why results from Rasch models cannot be applied to different cultural settings, consider two tests of English reading comprehension, one given to American students and one given to Chinese students, that refer to a passage about a politician who uses the rules of American football to describe a political situation. Because fewer Chinese students than American students know the rules of football, the question would be harder for the Chinese students to answer regardless of their English ability. Thus, a researcher would be mistaken to administer the same reading test in China and the United States, observe that the data fit a Rasch model in both samples, then assume that the item difficulties are the same and directly compare the ability scores for each sample. The same logic applies to comparing the results of tests administered decades apart.

approach, however, presupposes that a researcher has full control over the collection and analysis of data, along with a significant pool of test questions that all measure an underlying theoretical construct. Most researchers using longitudinal data, however, do not collect the data themselves, nor do they have the luxury of discarding questions simply because they do not fit a Rasch model. Instead, minimizing measurement error is a high priority, particularly in the case of deviance which is difficult to measure with precision. Accordingly, the present research relaxes the assumption that the discrimination parameters for every test item are equal in value. This approach is validated by some initial analyses with the data that compared fits of Muraki's Generalized Partial Credit Model (GPCM) and the Partial Credit Model (PCM) due to Masters (1982), which belongs to the Rasch family. The GPCM achieved a better fit to the data, supporting previous research showing that the GPCM generally achieves better fits to empirical data than its Rasch family equivalents (Thissen, Nelson, Rosa, and McLeod 2001:170-172).<sup>12</sup> Tables A7 and A8 in Appendix A also demonstrate that the resulting estimated discrimination parameters differ substantially in magnitude among survey items in the same wave and cohort, and Figures 2.1 and 2.2 demonstrate these differences graphically with two Item Characteristic Curves.<sup>13</sup>

The final selection of a Generalized Partial Credit Model (GPCM) is also based on characteristics of the data, as well as available software. Because many of the TYP delinquency questions are polytomous (using an ordinal scale based on the frequency with which respondents

---

<sup>12</sup> During initial analyses, I fit the GPCM and PCM to wave 2 of the J3 TYP data using the eRm ("Extended Rasch Modeling") package in R (Mair and Hatzinger 2007). As the eRm package uses a frequentist approach, however, the analysis required first eliminating from the data set 60% of all responses, associated with respondents reporting no delinquency at all for that year. A confirmatory factor analysis in LISREL further showed that the latent construct for the remaining 40% of cases was not necessarily unidimensional, violating the model assumptions for both a one-dimensional GPCM and PCM. A true comparison of these models would therefore require implementation of a Bayesian PCM, outside the scope of the present research.

<sup>13</sup> The difference in ICC slopes for polytomous questions is visually more dramatic, but the presence of multiple item location parameters per question (see Equation 2.2) means that discrimination parameters cannot be directly associated with the slope of the ICC curve for each category.

committed a particular type of delinquent act in the previous year), strong candidates include Muraki's Generalized Partial Credit Model, or GPCM (Muraki 1992), and Samejima's Graded Response Model (Samejima 1969), both of which model ordered categorical variables and do not constrain all item discrimination parameters to be equal. Osgood et al. (2002) implement Samejima's model, for which the item difficulty parameters are a little easier to interpret than those of the GPCM, but the present work uses the GPCM due to the availability of open-source software implementing the GPCM in a Bayesian framework (Patz and Junker 1999a; Patz and Junker 1999b). As part of his doctoral dissertation, Patz (1996) implemented Markov Chain Monte Carlo estimation of the GPCM for items with the same number of categories. The present implementation extends his source code to allow for items with different numbers of categories, to include an informative prior for values of  $\theta$ , to incorporate tests for item goodness of fit, and to speed convergence of the Markov chain using methods suggested by Patz (Patz and Junker 1999a:355).<sup>14</sup>

In the Generalized Partial Credit Model, the probability that the  $i^{\text{th}}$  respondent will check the box associated with the  $k^{\text{th}}$  category of item  $j$  is

---

<sup>14</sup> The improvements involve jointly drawing the difficulty parameters and log of the discrimination parameter from a multivariate normal distribution, and are suggested by Patz although he did not need to implement them for his own work. I am grateful to Richard Patz for making his code publically available, as the present work would not be possible otherwise. When updating the code, I identified a small coding error in Patz's `draw.pc()` function, which generates draws from the joint posterior distribution of the item parameters for the generalized partial credit model. The line

```
irt0 <- apply(log(tmp), 2, sum) +
  apply(matrix(log(dnorm(lalpha0, sd=sqrt(nu2))), d, J) - 2*lalpha0, 1, sum) +
  apply(matrix(log(dnorm(eta0, sd=sqrt(tau2))), ncol=K), 1, sum)
```

should read

```
irt0 <- apply(log(tmp), 2, sum) +
  apply(matrix(log(dnorm(lalpha0, sd=sqrt(nu2))), d, J) - 2*lalpha0, 2, sum) +
  apply(matrix(log(dnorm(eta0, sd=sqrt(tau2))), ncol=K), 1, sum),
```

and a similar adjustment should be made for the variable `irt1`. While I have checked my own work several times, as the occasional mistake in a complex endeavor is perhaps inevitable I would also appreciate readers of the present work sharing any errors in my source code that they find.

$$p(X_{ij} = k | \theta_i, a_j, \beta_{.j}) = \frac{\exp[\sum_{v=1}^k a_j(\theta_i - \beta_{vj})]}{\sum_{h=1}^K \exp[\sum_{v=1}^h a_j(\theta_i - \beta_{vj})]}, \quad (\text{Equation 2.2})$$

where for survey item  $j$  with  $K$  categories,  $\theta_i$  denotes the  $i^{\text{th}}$  person's criminality,  $a_j$  is the positive discrimination parameter that indicates how well item  $j$  distinguishes criminality, and  $\beta_{vj}$  denotes the difficulty of responding to the  $l^{\text{th}}$  category of item  $j$  and is referred to as the "item step parameter" in Masters' original Partial Credit Model (Masters 1982), where  $\beta_{1j}$  is arbitrarily set to 0 for all  $j$  for identifiability purposes.<sup>15</sup> Model assumptions for the GPCM include unidimensionality of the latent construct (previously verified), as well as assumptions that item discrimination parameters and difficulties are independent of each other, and that individual abilities are independent. The notion of "partial credit" stems from the fact that the categories are ordered by increasing level of difficulty: imagine a math problem with several steps, such that the completion of each successive step requires successful completion of the previous step. As an individual's ability increases, he or she will be more likely to successfully complete more and more steps, receiving greater "partial credit" until finally receiving full credit for completing the entire problem. The  $\beta_{vj}$  values thus denote a series of thresholds, such that if  $\theta > \beta_{vj}$ , then  $p(X_{ij} = v) > p(X_{ij} = v - 1)$  and if  $\theta < \beta_{vj}$ , then  $p(X_{ij} = v) < p(X_{ij} = v - 1)$  (Muraki 1992:162).

Patz's implementation of a Bayesian GPCM model (Patz 1996; Patz and Junker 1999a; Patz and Junker 1999b) involves MCMC draws from the posterior distribution of the individual ability and item discrimination and difficulty parameters using a Metropolis-Hastings-within-Gibbs algorithm. A particularly elegant feature of this approach is the fact that it allows for the estimation of missing values during an additional step in the algorithm, as long as a respondent

---

<sup>15</sup> Muraki (1992:161-162) shows that  $\beta_{vj}$  could be set to any number, as the term is cancelled from the numerator and denominator in Equation 2.1, and 0 is simply chosen for convenience and to follow convention.

has answered at least one survey question about delinquency during a given wave. The details of these draws are covered in Appendix B, along with the procedure for estimating missing values and extensions of Patz's code to speed convergence of the Markov chain.

### Determining Prior Distributions for Ability and Item Parameters

Given the need for an informative prior for the individual abilities  $\theta$ , it is preferable to draw one from a Chinese cultural context, and so the present research does so with a wave of the TYP survey not otherwise used for analysis. Table 2.1 contains delinquency items for Wave 1 of the J3 cohort, items not used to analyze the effect of work or romantic relationships on delinquency (as respondents are too young in that wave). Fitting the GPCM to these items thus yields a posterior distribution of individual abilities that then can be used as a prior for the other waves. Estimating a Bayesian GPCM for Wave 1 in turn needs a prior distribution, of course, and that one is based on knowledge of the distribution of criminality in the West. For the details of this procedure, please refer to Appendix B.

Prior distributions for item parameters are much more straightforward to compute. Following Patz and Junker (1999a:350), the prior distribution for item difficulty parameters is a normal distribution with mean 0 and variance  $\sigma_{\beta_j}^2$ , set to 2 for the present research; the prior distribution for the discrimination parameter is a lognormal distribution with variance  $\sigma_{\beta_2}^2$ , set to 0.7, and with mean  $\mu_{\beta_2}$ , set to 1 such that the corresponding normal distribution has mean 0.

With prior distributions specified, obtaining independent draws of the parameter estimates was straightforward. Table A9 shows the number of independent draws from the posterior distribution of the ability and item parameters used for analysis, and Appendix B also contains additional details concerning burn-in time and diagnostic plots.

### Assessing Model Fit

Determining whether the GPCM adequately fits the data for each wave makes use of a test statistic for item fit reported by Wilson (2005). The procedure requires first calculating the expected value of responses to each item  $j$ :

$$E_{ij} = \sum_{k=1}^{K_j} k \cdot p(x = k | \theta, a, \boldsymbol{\beta}) \quad (\text{Equation 2.3}).$$

Next, the squares of model-based residuals,  $W_{ij}$ , are calculated for each respondent:

$$W_{ij} = \sum_{k=1}^{K_j} (k - E_{ij})^2 \cdot p(x = k | \theta, a, \boldsymbol{\beta}) \quad (\text{Equation 2.4}).$$

The squares of observed residuals are then calculated:

$$Y_{ij}^2 = \sum_{k=1}^{K_j} (X_{obsij} - E_{ij})^2 \quad (\text{Equation 2.5}).$$

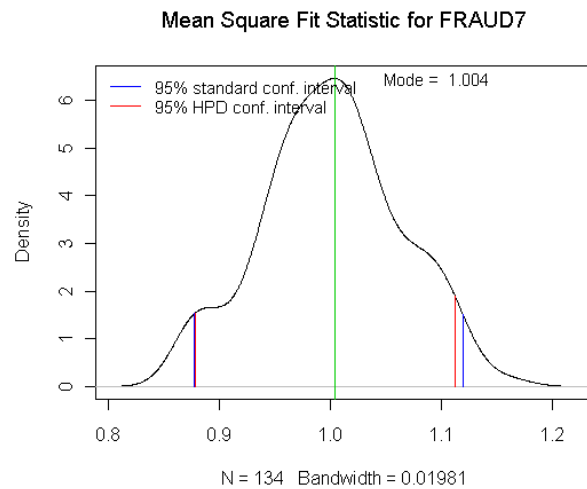
Finally, a mean squared error for the fit for item  $j$  is obtained as follows:

$$MSE_j = \frac{\sum_{i=1}^N Y_{ij}^2}{\sum_{i=1}^N W_{ij}^2}$$

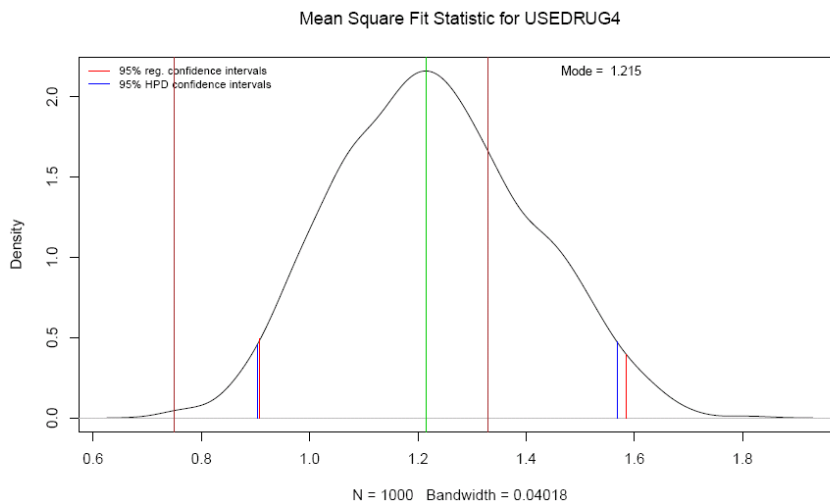
(Equation 2.6).

According to Wilson (2005), a mean squared error between  $\frac{3}{4}$  and  $\frac{4}{3}$  constitutes an acceptable item fit. The strategy for assessing item fit therefore involves computing a mean squared error for each item for each draw from the posterior distribution of  $(\theta, a, \boldsymbol{\beta})$ . If this error lies between  $\frac{3}{4}$  and  $\frac{4}{3}$  95% of the time, then the item is left in the model; otherwise, the variable

is transformed by collapsing categories with few responses together, or else removed from further analysis entirely. After survey items were recoded or removed from analysis, additional confirmatory factor analyses reconfirmed the unidimensionality of the new latent construct. Figure 2.3 shows an example of a good-fitting item, fraud as measured in Wave 7. The entire density of fitted values lies within 0.75 and 1.33.



**Figure 2.3.** An example of a well-fitting test item: fraud in Wave 7.



**Figure 2.4.** A poorly-fitting test item: drug use in Wave 4.

The only serious obstacle encountered during the model fitting process lies in the case of drug use in Wave 4 of the J3 cohort (Figure 2.4). Drug use and theft are the only two common items available to link respondents in their senior year of high school (J3 Wave 4 / J1 Wave 6) and their second year of college (J3 Wave 6 / J1 Wave 8), but theft fits even more poorly than drug use. The workaround involved linking J3W4 (seniors in high school) to J1W9 (seniors in college) using different variables. A model in which every item fits well is in turn a good-fitting model (Wilson 2005), so this procedure confirms that the data adequately fit the model. These findings suggest that the survey items are more reliable than their factor loadings or the overall alpha reliabilities for each latent factor might imply (see Table A6). The primary reason is that alpha provides a lower bound on reliability, a bound which can be a substantial underestimate of reliability when the number of items is small and the items are heterogeneous (Lord and Novick 1968:91), both of which are the case here. In addition, alpha is generally less than or equal to the reliability of a measure, but this bound becomes an equality when the *true scores* associated with each item, or “expected observed score with respect to the propensity distribution of a given person on a given measurement” (Lord and Novick 1968:30), are all linearly related (Lord and Novick 1968:90). This linear relationship makes sense if a set of measurements are Tau-equivalent (i.e. they have exactly the same true score), as in the case of test-retest reliability, and explains why test-retest reliabilities for delinquency items have alpha reliabilities that are so much higher than that for delinquency scales (Hindelang, Hirschi, and Weis 1981). In the present case, especially in waves such as Wave 2 for the J3 cohort that contain both dichotomous and polytomous items, there is no reason to believe that the scores associated with widely heterogeneous items, some of which contain very few responses, would at all be linearly related,

and hence Cronbach's alpha provides little information about the actual reliability of a test measure.

### Discussion of Parameter Estimates

Tables A7 and A8 contain the GPCM item parameter estimates for dichotomous and polytomous items, respectively, for each wave and cohort in the TYP study. Note that for polytomous items, the magnitude of an item difficulty parameter indicates how much more difficult a particular category is to answer than the previous category. Comparing "smoking" to "drinking" in J1 Wave 6, for example, the lower value of  $b_{\text{once or twice}}$  indicates for drinking implies that, a respondent is more likely to report drinking at least once or twice than to report smoking once or twice. The lower value of  $b_{\text{sometimes}}$  compared to  $b_{\text{once or twice}}$  for drinking does not imply that a respondent is more likely to report sometimes drinking as opposed to drinking once or twice, but rather that for drinking in Wave 6 for the J1 cohort, the relative increase in difficulty associated with going from drinking "once or twice" to drinking "sometimes" is less than the relative increase in difficulty associated with going from drinking "never" to drinking "once or twice". Particularly noteworthy is the fact that for almost every type of delinquency in every wave containing polytomous survey items, the difficulty parameter for "sometimes" is lower than that for "once or twice," suggesting that for many of the deviant behaviors listed here, committing a crime the first time is associated with a high threshold, and having crossed that threshold moving to occasional commission of the crime is easier. This pattern particularly makes sense for survey items that might involve some level of addiction or compulsive behavior – such as smoking, drinking, viewing pornography, playing mahjong, chewing betelnut, and illegally downloading music – but it also holds for drag racing, staying away from home, and

cheating on tests / deception. In fact, the only wave in which the transition from committing a delinquent action “once or twice” “sometimes” is more difficult than the transition from “not at all” to “once or twice” is Wave 1 for the J1 cohort, a wave which has the highest alpha reliability. This fact is not a coincidence, as alpha reliabilities assume a linear relationship between the latent variable and the survey items, and the decrease in the magnitude of the item difficulty parameters for the “sometimes” category compared to both “once or twice” and “often” is indicative of nonlinearity in the survey items.

Directly comparing item difficulty parameters across waves is problematic because the parameters are on different scales, but doing so is possible within waves. Comparing the values of  $b_{\text{once or twice}}$ , drinking is an easier crime to commit the first time than smoking, and perhaps for this reason the discrimination parameter is also lower, implying that the smoking item is better able to distinguish between respondents of high and low criminality. Skipping class is also easy, while theft, fighting, motorcycle racing, and betel nut use are much more difficult. Illegal downloading (in Wave 9 for the J1 cohort) appears to be one of the easiest crimes, even to the point that the item difficulty parameter for “sometimes” is negative – that is, given that a respondent has illegally downloaded software or music at least once or twice, he or she is actually *more* likely to report having done so at least “sometimes” rather than to only report having done so once or twice. The consistently low discrimination parameter for Mahjong compared to other items in the waves in which the item appears also hints at some of the concerns about the validity of the “once or twice” category mentioned previously. In short, the findings are intuitive, and provide another confirmation of the applicability of Item Response Theory to creating rational, substantive rankings of different types of deviance based on their seriousness while also shedding light on the relative seriousness of different types of delinquency

in Taiwan. One drawback to the use of the Generalized Partial Credit Model is that item difficulty parameters for categories above “once or twice” cannot be directly compared across questions, as the interpretation of their magnitudes in terms of their effect on the probability of offending depends on the magnitudes of difficulty parameters for less serious categories. In this sense the Graded Response Model is more straightforward to interpret, but given the lack of available software, implementation of a Bayesian Graded Response Model is beyond the scope of the present research.

### *Overview of Vertical Scaling Using Common-Item Linking*

With independent draws from the posterior distributions of the IRT parameters for each survey wave in hand, the next step in this analysis involves placing the ability scores onto common scales that enable comparisons of respondents’ propensity for deviance from one wave to the next. For this step in the analysis, I used vertical scaling using common-item linking (Kolen and Brennan 2004), the same procedure used to measure responses to the SAT on the same scale from one year to the next, or to provide for test security by enabling the comparison of test scores among students in the same class who take different examinations with only a few common questions. The present section reviews techniques for vertical scaling, in order to explain the particular methodology used for the present research and present a new algorithm for maximizing the precision of sequential common item linking when Bayesian IRT models are linked together.

Vertical scaling assumes that the discrimination and difficulty of the common items used for scaling are invariant from one administration of a test to the next. In the present research, this assumption means that over time a person’s criminality may change, but the inherent difficulty

of responding “yes” to having committed a crime such as theft does not. Accordingly, for the  $j^{\text{th}}$  question in Equation 2.1, the discrimination parameter  $a_j$  and difficulty parameters  $\beta_{lj}$  for each of the  $l$  categories in item  $j$  are assumed to be invariant across waves. Although they may be on different scales, they are linear transformations of each other (Lord 1980). In particular, for the Generalized Partial Credit Model in Equation 2.1 with discrimination parameters  $a_j$  and item difficulty parameters  $\beta_{lj}$  for item  $j$  with  $K_j$  categories, given  $A$  and  $B$  such that  $a_{jO} = a_{jN} / A$  and  $\beta_{kjo} = A\beta_{kjN} + B$  for the  $k^{\text{th}}$  category of item  $j$ , where the subscript  $N$  denotes the “new” scale and  $O$  denotes the “old” scale, it can be shown that  $\theta_{iO} = A\theta_{iN} + B$ , for all respondents  $i$  in the survey (Kim and Kolen 2005). Once the constants  $A$  and  $B$  that enable this transformation are calculated, they can also be used to place the ability parameters,  $\theta_i$ , onto the same scale for comparison across pairs of waves.

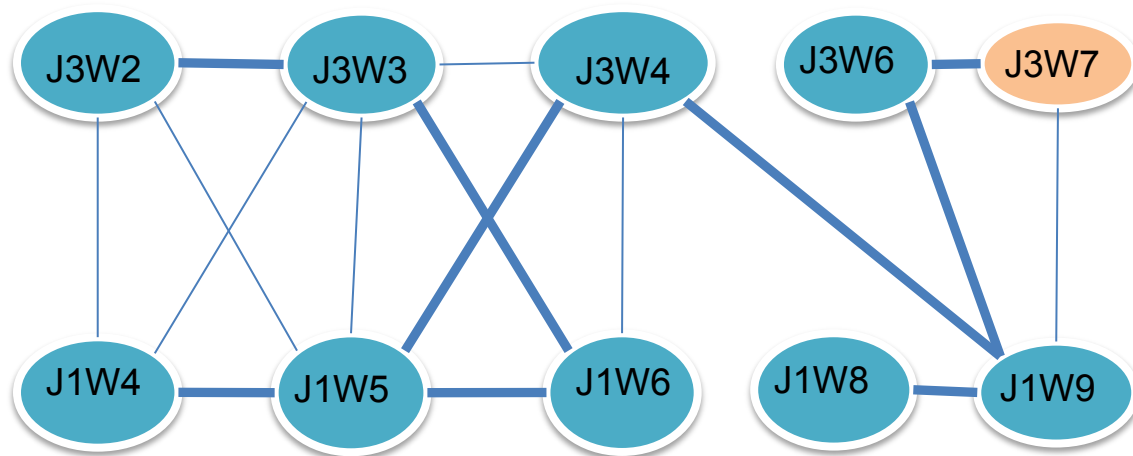
Of course, due to measurement error, estimates of the discrimination parameters and difficulty parameters for common items in consecutive waves are never exact linear transformations of each other. The challenge then becomes to find the  $\hat{A}$  and  $\hat{B}$  that minimizes the total transformation error over all pairs of common items. The two most common and tested types of estimations in the literature are based on preserving the moments of the item difficulty parameters, and on preserving the shape of the item characteristic curve (ICC) or the test characteristic curve (TCC).<sup>16</sup> Because the number of ability parameters is greater than the number of difficulty parameters, a procedure that minimizes the error in transforming the  $\theta$  parameters is generally a more accurate approach than one that only focuses on the item discrimination and difficulty parameters. Characteristic curve methods take advantage of this feature while moment methods do not. Additionally, ensuring that the error in rescaling is

---

<sup>16</sup> If a plot of  $P(\theta|a_j, \beta_{lj})$  for  $l=1, \dots, K_j$  vs.  $\theta$  gives the item characteristic curve for item  $j$ , then a plot of  $\sum_{j=1}^J P(\theta|a_j, \beta_{lj})$  vs.  $\theta$  gives the test characteristic curve for a test containing  $J$  items.

minimized for deviant respondents who have particularly high criminality is particularly important for this research, and so maintaining the entire shape of the item characteristic curve matters. Accordingly, this research uses characteristic curve methods for vertical scaling.

Another practical matter in vertical scaling involves the choice of *separate* vs. *concurrent* calibration methods. With separate calibration, if ability scores associated with more than two tests are to be placed on the same scale (as in the case of the present research), then pairs of tests are linked together (“calibrated”) separately, and the entire data set is then linked together as a chain (Kolen and Brennan 2004). Imagine the set of tests or surveys to be linked together as nodes in a graph, with the common items linking them as undirected edges. Separate calibration requires a researcher to select a set of edges that comprise a spanning tree, as in Figure 2.5. The approach requires selection of a “base” scale to use as a metric for all tests or survey waves (i.e. a root for the tree), and often the choice of edges in this graph to use for calibration may appear arbitrary. The hypothetical spanning tree in Figure 2.5, for example, is intuitively suboptimal, as expressing delinquency in Wave 2 for the J3 cohort using the scale in Wave 7 for the J3 cohort would involve no fewer than seven linear transformations. Selection of a more central node to be the root and spanning trees with a shorter average path length would clearly appear to reduce measurement error, but still the question of which spanning tree would reduce total measurement error remains difficult, and methodologists often view the vertical scaling process to be just as much of an art form as it is a science (Harris 2007; Petersen 2007).



**Figure 2.5.** Hypothetical example of vertical scaling for the delinquency items in the Taiwan Youth Project survey using common-item linking. Nodes in the graph represent survey waves, edges between nodes denote the presence of at least one set of common delinquency items to calibrate pairs of waves, and thick lines denote a hypothetical configuration for placing all ability scores for the survey on the same scale using separate calibration. J3W7 is a hypothetical root for the spanning tree.

Concurrent calibration, in contrast, only involves estimation of a single model, with all the parameters for common-item linking determined at the same time. In one sense it is more versatile than separate calibration in that a researcher could choose an approach for calibration corresponding to a set of edges in Figure 2.5 that covers all vertices and also creates cycles. In that case, optimum values for linear transformations associated with all edges in the graph could be obtained through an iterative procedure similar to that used to fit non-recursive structural equation models in statistical packages such as LISREL. In practice, however, separate estimation has been argued to be more robust to mild violations of IRT model assumptions including unidimensionality (Kolen and Brennan 2004) and concurrent estimation currently requires specialized proprietary software, so for both pecuniary and methodological reasons the present research uses separate calibration.

### Characteristic Curve Methods for Common-Item Linking

The two characteristic curve methods for common-item linking most commonly used in the literature on vertical scaling are the Haebara method (Haebara 1980) and Stocking-Lord method (Stocking and Lord 1983). Appendix C describes the two methodologies in greater detail. For both procedures, the error associated with transforming ability scores from one scale to another is calculated for one individual at a time and the error differences are summed over all individuals, but the Haebara method is based on the item characteristic curve (ICC), while the Stocking-Lord method is based on the test characteristic curve (TCC). The literature on measurement in criminology provides no theoretical reason to choose either of these two methods over the other, and I could find no recommendations in the literature on vertical scaling regarding which method to use; indeed, Kolen and Brennan (2004) call for more empirical research comparing the efficacy of the two methods.

The Bayesian methods used in the present study do allow for the drawing of empirical inferences concerning which method is more accurate, in addition to the selection of an optimal set of common items and pairs of waves for placing measures of criminality in a longitudinal survey onto a common scale using sequential common-item linking. This is possible because instead of a single number estimated for each respondent's proficiency score, several thousand draws from a posterior distribution are available. Accordingly, one way to decide whether to use the Haebara or Stocking-Lord method for any given common-item linking is to obtain thousands of values of  $A$  and  $B$  through a Monte Carlo procedure, then select the approach which creates estimates of  $A$  and  $B$  with the smallest standard deviation. Similarly, for each pair of waves with items in common, the same approach enables the selection of an optimum set of common items that minimizes the summed variance for estimates of  $A$  and  $B$  for pairs of survey waves

corresponding to the edges in Figure 2.5.<sup>17</sup> Moreover, given a selection of characteristic curve method and a choice of common items associated with the most precise scaling of ability scores in adjacent waves, these precisions can be translated into weights for the edges in Figure 2.5, enabling an objective choice of root node and common-item linkages that minimize the total measurement error associated with vertical scaling, with the most central node in the graph the best candidate for the root. The next section formalizes this reasoning into an algorithm for determining an optimal choice of wave and common-item linkages in order to minimize measurement error during the vertical scaling process.

*An Algorithm for Optimizing Precision During Vertical Scaling Using Bayesian Estimation*

Let  $G = (V, E)$  be a graph containing a set of vertices  $V[G]$  that represent a set of tests or surveys to be linked together using vertical scaling, as well as a set of weighted, undirected edges  $E[G]$  that join all pairs of vertices for which at least one common item exists. Let the weight for each edge  $e$  in  $E[G]$  correspond to the error associated with transforming the standardized ability scores for one of the endpoints of  $e$  to the scale for the other endpoint (also standardized), using a set of common items and a characteristic curve linkage method that minimizes measurement error for that edge. Further, let  $E_{BEST}[G]$  be the set of edges in  $G$  associated with the rescaling that minimizes total measurement error for the vertical scaling process, and let  $V_{ROOT}$  represent the test or wave to which all ability scores will be rescaled for this optimum scaling. Let  $\delta(V_{ROOT}, v)$  be the path in  $G$  corresponding to a transformation of the scale for  $v$  to that for the root node that has the lowest possible measurement error compared to all other paths from  $V_{ROOT}$  to  $v$ , and accordingly is a shortest path. As all edge weights are positive,  $\delta(V_{ROOT}, v)$  must also

---

<sup>17</sup> The present choice of the sum of the variance of A and the variance of B as a criterion for maximizing precision is a heuristic, and future research could likely improve upon this technique.

be acyclic. (Should  $\delta(V_{ROOT}, v)$  contain a cycle, then a shorter path than  $\delta(V_{ROOT}, v)$  could be produced by removing the cycle from the path, contradicting the assumption that  $\delta(V_{ROOT}, v)$  is a shortest path.) Total measurement error is computed as the sum over all vertices of the edge weights in each path  $\delta(V_{ROOT}, v)$ . Thus, assuming for the moment that  $V_{ROOT}$  is known in advance, the problem reduces to a single-source shortest paths problem for a weighted graph with non-negative edge weights, and Dijkstra's algorithm can solve this problem in  $O((V+E)\lg V)$  time (Cormen, Leiserson, Rivest, and Stein 2001). Every vertex must be reachable from  $V_{ROOT}$  in order for scaling to be feasible, and in this situation Dijkstra's algorithm produces a shortest-path spanning tree. With  $V_{ROOT}$  not known in advance, the problem simply reduces to running Dijkstra's algorithm  $V$  times, allowing each vertex in  $V$  to be the root, and finally selecting for  $V_{ROOT}$  and  $E_{BEST}[G]$  the choice of root and spanning tree that minimizes total measurement error. The running time for this algorithm is  $O(V\lg V(V + E))$ .

### Performing Vertical Scaling Using POLYST

The present research uses the free POLYST software package (Kim and Kolen 2003). POLYST is perhaps older than other software applications for vertical scaling, but is useful in that it is open source and can also be executed from a DOS command line, with the user specifying an input file containing the parameters to be estimated and a syntax file controlling the analysis itself. The estimated values of  $A$  and  $B$ , using both the Haebara and Stocking-Lord method, are then printed to an output file. POLYST was designed for a maximum likelihood framework, but as a DOS shell script could call the program an arbitrary number of times from the command line, the software could in fact generate a distribution of parameter values  $A$  and  $B$  given draws from the posterior distribution as input, using a brute force approach. Using the full

distribution of  $A$  and  $B$  through Monte Carlo techniques also carries with it the benefit of propagating measurement error in delinquency to the next stage of analysis if desired.

Performing the vertical scaling for each wave requires first translating Patz and Junker's reparameterization of the item difficulty parameters (Equation B1) back into Muraki's notation (Equation 2.2), computing  $\beta_{lj} = \frac{\beta_{\delta lj}}{\beta_{2j}}$ . Second, because the ability scores in different waves can and do have different scales, they must be standardized to be comparable (Kolen and Brennan 2004:161). The process of standardization is itself a linear transformation:

$$\frac{\theta - \mu_{\theta}}{\sigma_{\theta}} = A\theta + B, \quad (\text{Equation 2.7a})$$

where

$$A = \frac{1}{\sigma_{\theta}}, B = -\frac{\mu_{\theta}}{\sigma_{\theta}}. \quad (\text{Equation 2.7b})$$

The item difficulty parameters can be similarly transformed using Equations C2 and C3. Treating the standardized ability scores as the "old" scale and the draws of ability scores from their posterior distribution as the "new" scale (such that  $\theta_{01J3W1} = A\theta_{J3W1} + B$  for the J3 Wave 1 ability scores for example), substituting values for  $A$  and  $B$  in Equation 2.7b then yields

$$a_{jO} = a_{jN}\sigma_{\theta_N} \quad (\text{Equation 2.8})$$

and

$$\beta_{jvO} = \frac{\beta_{jvN} - \mu_{\theta_N}}{\sigma_{\theta_N}} \quad (\text{Equation 2.9})$$

for each discrimination parameter and item difficulty parameter. The first item difficulty parameter,  $\beta_{j10}$ , may however retain a value of 0 after standardization, because the parameter drops out of Equation 2.2 and can therefore have any value.

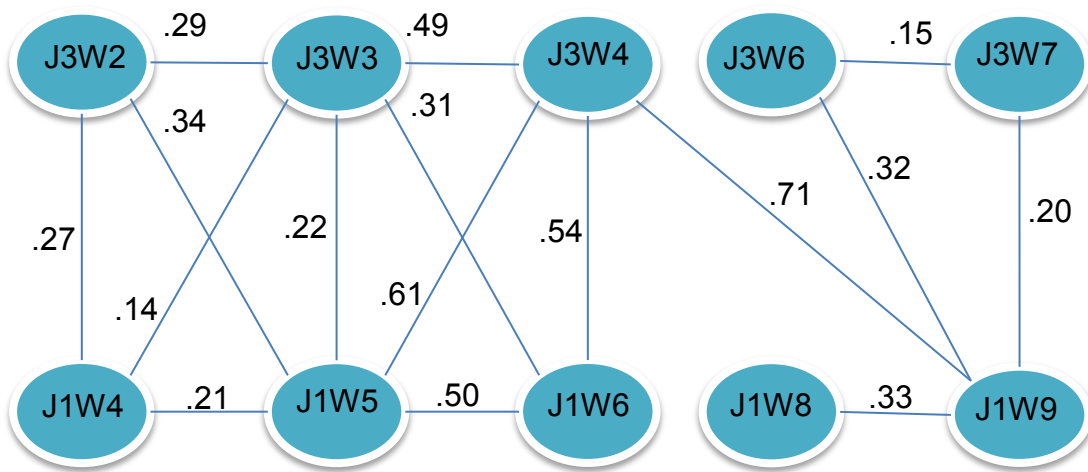
Next, for every pair of waves containing at least one common item<sup>18</sup> (corresponding to an edge in Figure 2.5), defining one wave as the “old” wave and one as the “new” wave, 4,025 random pairings of draws from the joint posterior distribution of the item parameters for each wave are generated.<sup>19</sup> Given  $n$  common items linking each wave,  $\sum_{i=1}^n \binom{n}{i}$  possible subsets of common items exist that could be used for the vertical scaling. For each of these subsets, the present approach generates a POLYST input file that passes in as input the number of items and number of categories for each item, the item discrimination and difficulty parameters for each item, and a command to complete the vertical scaling based on 100 “proficiency” points spaced at equal intervals between the smallest and largest observed standardized ability scores for this particular wave (Equations C1 and C4). These files also specify that the scaling transformations should be symmetric (see the discussion of the Haebara and Stocking-Lord methods in Appendix C), a property that Lord describes as necessary because “*A basic requirement of equating is that the result should be the same no matter which test is called [old] and which is called [new]*” (Lord 1980:198, emphasis his own). DOS batch files are then generated and run that call POLYST 4,025 times for each set of common items, then automatically read the generated output files with another program, extracting the parameters  $A$  and  $B$  as calculated using the Haebara and Stocking-Lord method. Finally, the characteristic curve method and set of common items are selected that minimize  $\sigma_A + \sigma_B$  for each pair of waves containing common items

---

<sup>18</sup> Two items in separate waves are common items if they have the same wording and the same number of categories (for some survey items, some categories had to be collapsed together in order for the item to fit the GPCM).

<sup>19</sup> The number 4,025 corresponds to the number of posterior draws for the wave with the greatest number of posterior draws, and has no particular significance here.

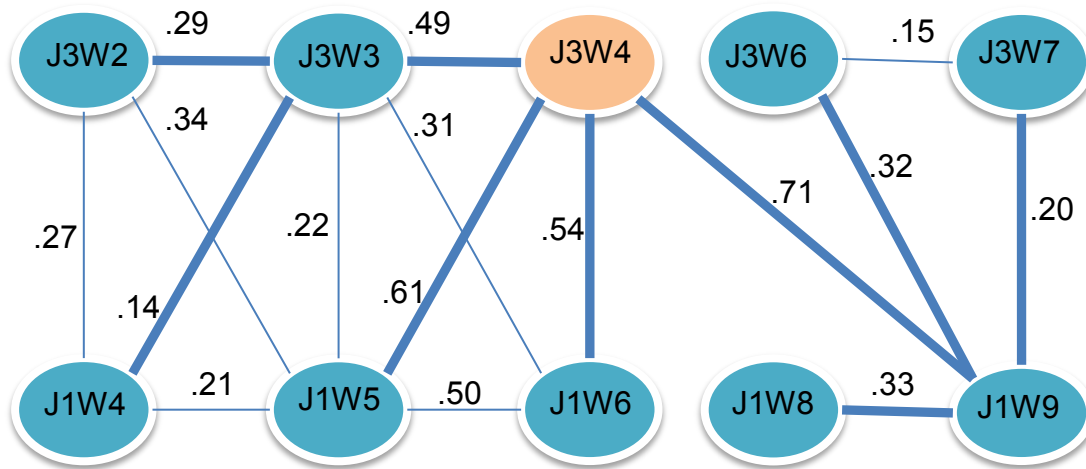
(Figure 2.6).<sup>20</sup> For 9 of the 16 edges in Figure 2.6, the Stocking-Lord method yielded the minimum value of  $\sigma_A + \sigma_B$ ; for two edges, the Haebara method yielded the minimum; and for five edges, the two methods yielded the same results because only one common item exists for that pair of waves. Considering the 11 edges for which more than one common item could be used for the linking, then, the average percent improvement in precision for using the Stocking-Lord method compared to the Haebara method was 10.6%.



**Figure 2.6.** Values of  $\sigma_A + \sigma_B$  associated with the choices of characteristic curve method and common items for vertical scaling that minimize measurement error associated with the separate calibration of pairs of TYP waves.

Using the `igraph` package in R (Csardi and Nepusz 2006) to apply Dijkstra’s algorithm to the graph in Figure 2.6 as described above produces a root node and spanning tree, as shown in Figure 2.7.

<sup>20</sup> In some cases, POLYST returns an invalid termination code indicating a “lower point cannot be found,” and in spite of the software’s statement that the convergence of the algorithm is “probably successful” in such cases, in a few cases with a single dichotomous variable and one common item for which Equation C1 is equal to Equation C4, the Haebara and Stocking-Lord values for A and B differed, often considerably. As this difference caused  $\sigma_A + \sigma_B$  when using the Haebara and Stocking-Lord methods to differ by as much as 3% when they should have been exactly the same, to be safe I removed from analysis all cases yielding improper termination codes. For the vast majority of combinations of common item linkings, POLYST converged in more than 99.5% of all runs of the program; in the case of only one single common-item linking for a single pair of waves was the rate of non-convergence as high as 5.8%, and the set of common items associated with that linkage was not ultimately selected as the preferred set for that edge. Accordingly, I do not believe that in rejecting some POLYST runs I have significantly biased the values of the parameter draws used for the linking and thus the values for A and B ultimately selected for the scalings.



**Figure 2.7.** Choice of separate calibrations in Figure 2.6 and a “root” node for a scale that minimizes total measurement error for vertical scaling of the TYP delinquency items.

Additionally, some analyses in Chapter 5 only require the use of Waves 6, 7, 8, and 9, so Dijkstra’s algorithm was run a second time on a subgraph containing only J1W6, J3W6, J3W7, J1W8, and J1W9 (as the graph does not contain J3W4, I simply add an edge from J1W6 to J1W9 to this graph, with the weight equal to the sum of the weights for the edge from J1W6 to J3W4 and from J3W4 to J1W9). In this new graph, the algorithm identifies J1W9 as the root node for the scaling process.

### Concurrent Validity Checks

Wu (2003) has shown that depression and delinquency are correlated at approximately 0.24 in Taiwan. Accordingly, after converting all ability scores associated with a latent propensity to offend to the J3W4 scale (for a rescaling of the full data set) and to the J1W9 scale (for analyses in Chapter 4), the rescaled Expected A Posteriori (EAP) values for delinquency were correlated with depression in the TYP survey, a measure for which the questions never change.

Additionally, all the rescaled posterior draws for latent delinquency for each wave were also

correlated with delinquency to generate 95% confidence intervals for these correlations. Table 2.2 shows the correlations for this study, with confidence intervals. In Table 2.2, correlations of delinquency with depression are attenuated, likely due to measurement error. The correlations are higher for the EAP estimates, suggesting that these estimates more closely approximate the respondents' true delinquency compared to a typical data set comprised of one random draw from the posterior distribution of the ability score for each survey respondent (far right column). Accordingly, the rest of this dissertation uses the EAP draws to conduct analyses.<sup>21</sup>

---

<sup>21</sup> Note that the magnitude of the correlations in Table 2.2 are unaffected by measurement error introduced as part of the vertical scaling process, because for each draw from the posterior distribution of the ability scores, one posterior draw for each ability score is multiplied by the same values of  $A$  and  $B$  to perform the vertical scaling. As the vertical scaling involves a linear transformation only, each set of rescaled ability scores will have the same correlation with the depression scale as the original set of ability scores.

**Table 2.2.** Pearson product moment correlations of delinquency with depression in the ten cohort-waves used for the analyses in this dissertation. 95% confidence intervals for these correlations are shown in parentheses; they are a frequentist interval as returned by the `cor.test` function in R's `stats` package for the Expected A Posteriori (EAP) delinquency estimates, and Bayesian confidence intervals for thousands of correlation coefficients calculated from posterior draws of delinquency in the far right column.

Wave	Alpha Reliability of Depression Scale	Correlation of Depression with EAP Delinquency Estimates	Average Correlation of Depression with Delinquency Estimates
J1W4	.743	.171 (.132,.210)	.125 (.102,.149)
J1W5	.705	.103 (.060,.147)	.077 (.051,.103)
J1W6	.885	.172 (.127,.216)	.137 (.114,.161)
J1W8	.897	.113 (.066,.159)	.080 (.053,.107)
J1W9	.887	.169 (.124,.212)	.131 (.106,.157)
J3W2	.669	.132 (.093,.170)	.083 (.058,.108)
J3W3	.670	.124 (.085,.163)	.098 (.075,.119)
J3W4	.885	.172 (.130,.214)	.117 (.088,.145)
J3W6	.891	.122 (.077,.167)	.095 (.071,.119)
J3W7	.900	.166 (.121,.209)	.129 (.105,.153)

*Summary*

This chapter extends Osgood, McMorris, and Potenza's initial application of Item Response Theory (IRT) to the study of deviance (2002), demonstrating how vertical scaling can allow criminologists to model changes in individual deviance in longitudinal surveys in which the survey items change. The applications of the nonlinear Generalized Partial Credit Model presented here also quantify the relative seriousness of types of delinquency as asked of

Taiwanese youth. Characteristics of the data in Taiwan, in which the majority of respondents report having committed no delinquent acts in each survey wave, both justify the inclusion of minor forms of deviance in the survey and make necessary Bayesian methods for fitting IRT models to the data. The Bayesian approach in turn provides an opportunity for further refinement of the vertical scaling process, using posterior draws of the IRT parameters and Dijkstra's algorithm to generate optimal combinations of characteristic curve methods, sets of common items, and pairs of waves to link together with sequential common item linking such that total measurement error due to the vertical scaling process is minimized. Bayesian IRT modeling of delinquency also allows for graceful handling of missing values when estimating the latent criminality scores, and for improving the quality of evaluating item fits.<sup>22</sup> Some evidence suggests that the use of IRT calibration methods based on the test characteristic curve (the Stocking-Lord model) may allow for more precise calibrations than those based on the item characteristic curve (the Haebara model).

The software used for the analysis here can be further developed in several ways. The analysis presented here has the shortcoming of not testing that the validity of the EAP estimates is robust to alternate specifications of the prior distribution. The use of a wave of the TYP survey to generate an informative prior distribution increases the credibility of the particular prior used here, but an important goal for future research would be to adapt the current implementation of the Bayesian IRT model into a user-friendly R package that allows for quick and easy specification of alternate prior distributions for the model parameters. Additionally, the Bayesian algorithms presented here are somewhat slow to converge, but improvements to the software could speed them by at least an order of magnitude. An implementation of the MCMC algorithm

---

<sup>22</sup> In the present case, the Taiwan Youth Project survey team was so careful in their data cleaning that only a handful of cases (fewer than a dozen) were missing for each delinquency variable for each wave, so Bayesian methods for missing value imputation would potentially be much more important in other research settings.

in C instead of R (while preserving the front-end of the package in R) would likely make the code run two to three times faster, as C is one level of abstraction closer to machine language than R (which is implemented in C) and is almost always faster. Additionally, the Metropolis Hastings within Gibbs procedure to obtain posterior draws of the ability and item parameters, because of the conditional independence of the steps in the chain, is an excellent candidate for the use of concurrency. Multiple threads would generally accelerate the computation by a factor approaching the number of a user's processors, and using distributed computing techniques to run the program on a cluster of computers would speed computation still further. Should the approach still be slow, using Empirical Bayesian methods could produce still more improvements in speed. The vertical scaling algorithm similarly is easily broken down into independent pieces, and is a prime candidate for multithreading. The use of batch files to repeatedly run POLYST also adds a great deal of administrative overhead to the software, and including POLYST as a library in another program instead would also speed computation.

The selection of  $\sigma_A + \sigma_B$  as a heuristic to capture precision in the vertical scaling is admittedly ad hoc, and further work is needed to identify a measure of calibration error that combines information about  $A$  and  $B$  in a manner that has desirable statistical properties. In future work, it would also make sense to repeat this analysis with a data set for which we know respondents' "true" delinquency because the survey items do not change over time. With the methodological challenge of modeling deviance in the TYP survey now addressed, the focus of the dissertation now returns to an investigation of the causal effects of part-time work and steady romantic relationships on delinquency, and of marriage on smoking, for Taiwanese youth. In order to set up the causal models, Chapter 3 proceeds to review the literature on the causes of delinquency in Taiwan.

## **Chapter 3: Literature Review of the Precursors of Delinquency for Taiwanese Youth**

This chapter reviews the literature about the causes of delinquency in Taiwan, in order to set up selection models for work and marriage in the next two chapters. Most of the research on the etiology of crime in Taiwan involves tests of Western theories, as the majority of Taiwanese sociologists received their formal education in the West before finding academic positions in Taiwan. Even so, many Asian criminologists, through empirical research, have identified aspects of their own cultures that the Western theories do not address. Their empirical findings based on adaptations of the Western theories to Asia are discussed in this chapter, as they have implications for the variables included in selection models in this dissertation, as well as for the social context in which work and romantic relationships may affect delinquency. Although these findings are usually too recent for the ideas they raise to have been directly incorporated into the Taiwan Youth Project survey in the form of new survey items, they still inform the selection of other variables that are not theorized to be as proximate a cause of delinquency but still are useful controls. For similar reasons, studies involving Mainland China, Hong Kong, and Chinese Americans in the U.S. are referenced in this chapter when findings from Taiwan are not available for some variables.

This review is organized mainly by criminological theory given the focus of Taiwanese criminologists on testing them, with corresponding evaluations of variables available in the TYP survey associated with these theories and their likely explanatory power with respect to deviance in Taiwan. A drawback of much of the available research covering Mainland China is the fact that very few longitudinal surveys of crime are available in China (Liu 2008), although the case

is a little better for Taiwan, and several high-quality findings are available in the form of recent master's theses published in Chinese. The review of the etiology of crime in a Chinese cultural context covers social control theory, Gottfredson and Hirschi's general theory of crime, differential association / social learning theory, strain theory, and social disorganization theory.<sup>23</sup> In each case the review focuses on variables associated with these theories that explain variation in work or deviance. In assessing the relative strengths of these variables, the review describes the extent to which different studies consistently report significant relationships between these variables and delinquency, the methodological rigor of the studies, and the types of causal mechanisms supported by the studies.<sup>24</sup> The chapter also pays particular attention to studies that use the Taiwan Youth Project survey data, and highlights whenever they do. Many of the potential confounders discussed here have been identified via research based on theories that have different and sometimes conflicting assumptions,<sup>25</sup> but the fact that these assumptions conflict does not hinder for the analyses in this dissertation, which seek to explain variability in part-time work and marriage and then test the relationship between these variables and either general deviance (for part-time work and intent to marry) or smoking (for marriage). These competing theories are competing precisely because they both explain the behavior of real people, and are thus both associated with confounders that should be incorporated into these causal models.

---

<sup>23</sup> Rational choice theory matters as well, but is more linked to the direct relationship between employment and delinquency, and I could not find any studies in Taiwan examining this relationship using a rational choice perspective.

<sup>24</sup> The review does not directly compare effect sizes of these variables in explaining delinquency, because crime is often measured quite differently in these various studies.

<sup>25</sup> For example, self-control theory assumes that everyone has a similar understanding of what is moral and what is not, and all that distinguishes a criminal from a non-criminal is his or her level of self-control, while social learning theory assumes that different people have different definitions of morality that are more or less favorable to crime.

### Social Control Theory

A body of research finds support for social control in lowering delinquency. Travis Hirschi argues that bonds to social entities such as the family, school, community organizations, and peers can lower crime by essentially giving a person something to lose, and “delinquent acts result when an individual’s bond to society is weak or broken” (1969:16). These bonds can be measured through attachment to those entities, commitment to the conventional goals of these entities, involvement in conventional activities, and beliefs unfavorable to delinquency. Empirical tests of social control measure bonds to specific social institutions, so this section assesses available research findings regarding attachment, commitment, and involvement, and beliefs (when studied) with respect to family, school, peers, and the workplace. It then presents some caveats associated with empirical tests of social control theory in Taiwan: in particular, Asia and the West may differ in terms of the overall strength of social bonds and the nature of moral beliefs.

### Variables Associated with Family Social Control in Taiwan

This section addresses the role of parental monitoring and attachment, commitment to family-related goals, and involvement in family activities in affecting deviance in Taiwan. As parental monitoring and parental attachment are considered the same by Hirschi (1969:88-90) this section addresses both concepts, but following convention this review and subsequent analysis treats the two variables as separate.<sup>26</sup> A number of cross-sectional studies of middle and

---

<sup>26</sup> For Hirschi, parental monitoring (e.g. “Does your mother (father) know whom you are with when you are away from home?”) is a component of parental attachment, a concept operationalized in most studies with questions such as “How close do you feel to your mother?” Parental monitoring could be viewed as a property of the bond that is objective in that it could be measured through direct observation, and which is determined more by the parent than

high school students in Taiwan and China consistently find that parental monitoring (Deng and Roosa 2007; Li, Martin, Armstrong, and Walker 2011; Sheu 1988), sometimes generalized as “family supervision” (Wang and Jensen 2003), predicts lower delinquency. Using structural equation models with a sample of junior high school students in China, Deng and Roosa (2007) show that parental monitoring predicts lower delinquency by decreasing deviant beliefs and increasing parental attachment for boys and girls. For girls in particular, they also find that parental monitoring leads to greater rewards for conventional behavior and lower resulting deviance, suggesting that they are expected to conform to conventions in Chinese society more than boys.<sup>27</sup> Wang and Jensen (2003) also find that having grandparents living at home in Taiwan also is associated with lower deviance, but only for children from broken families (with one or more parents not living at home)<sup>28</sup>. Whether the salient factor lowering deviance is associated with monitoring or some other aspect of the social bond, however, is unclear. Because of the multiple mechanisms through which parental monitoring has been linked to deviance and its consistent appearance as a relevant explanatory variable for deviance, and because youth who are carefully monitored by their parents may be less likely to work part-time and less likely to date – an activity of which many Taiwanese parents disapprove – it is clearly an important potential confounder to be controlled for in selection models for work and marriage in this dissertation.

In line with Hirschi’s theory, attachment to the family, commitment to family-related goals, involvement in family activities, and moral beliefs as influenced by the family have all been

---

the child. Attachment is usually considered more subjective and defined more by the child than the parent, leading to the two being treated as separate concepts in most studies.

<sup>27</sup> These gender-specific findings are consistent with those of Huang (2006a), who finds that junior high school girls in Miaoli County who misbehave in school are more likely than boys to upset their teachers, creating conflict with them.

<sup>28</sup> These results are consistent with Carol Stack’s findings for African-American families living in America’s inner cities (e.g. Stack and Burton 1994).

linked to lower deviance in Chinese samples (Chang 2010; Deng and Roosa 2007; Goo 2007; Juang and Nguyen 2009; Lee 2011; Li, Martin, Armstrong, and Walker 2011; Zhang and Messner 1995). Concerning parental attachment, Hirschi notes that it is important because it captures “the essence of internalization of norms, conscience, or super-ego” (1969: 18). The literature covering the effect of parental attachment on deviance in a Chinese cultural context is largely limited to cross-sectional data, but contains consistent findings demonstrating that parental attachment is associated with lower crime and appears to act through several separate mechanisms.

A U.S. study of Chinese-American adolescents (Juang and Nguyen 2009) found that a sense of obligation to honor the family is associated with lower juvenile delinquency, but only for boys. Gender aside, in a set of in-depth qualitative interviews of respondents to the Taiwan Youth Project survey, Pan and Yi (2011) show that survey participants from disadvantaged families who reached high levels of education did so because their parents consistently expected them to do well in school, and because they did not wish to let their parents down. These results suggest a strong relationship between family attachment and educational attainment in the same data set used for the current study, and also imply that respondents with high family attachment may be involved in deliberate, long-term planning related to family goals that could affect selection into work, romantic relationships, and deviance.

Deng and Roosa (2007), using structural equation models, further delve into the mechanisms through which parental attachment and a child’s deviant beliefs affect juvenile delinquency in a cross-sectional sample of junior high school students in Beijing. They find that parental rewards for conventional behavior affect juvenile delinquency only indirectly by increasing parental attachment, which in turn lowers delinquency. They also observe a gender difference in the

mechanism through which parental attachment is associated with lower deviance: for girls, attachment to parents reduces their deviant beliefs, in turn lowering deviance.<sup>29</sup> In contrast, for boys parental attachment has no effect on deviant beliefs, but both parental attachment and deviant beliefs have independent, direct effects on deviant behavior. Given that deviant beliefs are not consistently measured in the TYP while parental attachment is, the implication of these findings for the present research is that including parental attachment in selection models for work and marriage while omitting deviant beliefs is potentially less problematic for girls than for boys.

Chang (2010), also using structural equation models and analyzing the TYP data with structural equation modeling, finds that parental attachment is associated with lower deviance, both directly and indirectly by lessening the difference between the moral beliefs of the parent and child. He uses only a single wave of deviance due to the changing nature of the delinquency items addressed in the previous chapter. Chang also finds that boys have moral beliefs that are more in keeping with those of their parents than girls.<sup>30</sup> Finally, in a cross-sectional study of Taiwanese students aged 11-15, Liu (2003) predicts deviance while controlling for both attachment to each parent and the extent to which the child feels socially distant from his or her

---

<sup>29</sup> For measures of deviant beliefs, the authors use survey items that ask respondents “how wrong” they consider a dozen types of delinquent activities. They then refer to these beliefs in terms of definitions favorable or unfavorable to crime, an orientation more in keeping with differential association theory than control theory. Their theoretical model, however, is an application of the Social Development Model (Hawkins and Weis 1985) in a Chinese cultural setting for the first time, and this model is a hybrid of social learning, control, and differential association theories. Despite the assumptions of their theoretical model deviating somewhat from those of control theory (which assumes that the relative weakness of moral beliefs, not the presence of beliefs favorable to delinquency, is what leads to crime), their model is still relevant here in explaining one mechanism through which parental attachment affects delinquency.

<sup>30</sup> Chang’s measurement methodology is worth noting here. Both the TYP youth and parent survey respondents are asked to indicate how important they consider up to 20 moral principles; Chang determines the absolute value of the difference between the parent’s and child’s responses for the remaining questions, then conducts a principal axis rotation to identify two latent factors onto which the remaining eight questions load. One relates to personal responsibility related to school achievement and hygiene, and the second relates to empathy and modesty. His final measure of consistency between the moral beliefs of the parent and child is a latent variable measured by these two dimensions.

parents (anomie); in keeping with Hirschi's theory that deviance occurs when bonds are weak, anomie has strong effects on light, moderate, and serious deviance while net of anomie, attachment to either parent is not associated with deviance. The literature thus provides strong evidence that attachment to parents predicts lower delinquency, and weak evidence that feeling socially distant from parents in particular may lead to higher delinquency. Especially as youth who have close relationships with their parents would likely be more likely to comply with their parents' views concerning whether they should work or date while in school, parental attachment is therefore an obvious confounder to include in selection models for both work and dating.

The case for family involvement is fairly weak, as few studies examine the relationship between involvement and deviance in China or Taiwan. Deng and Roosa (2007) show that boys are more rewarded for involvement in conventional family activities than girls and such rewards are associated with greater parental attachment, providing empirical support for a cultural preference for sons in China – but they find no direct effect of family involvement on deviance. Hirschi (1969) also finds little effect of involvement on deviance, and notes that he may have been wrong in hypothesizing that any involvement in other activities is associated with lower crime, given that crime requires relatively little time to commit. The present research includes family involvement as it may be a useful proxy for attachment to family members other than the parents, but this variable is not expected to be a strong predictor of delinquency, and the fact that it is not consistently measured across survey waves is likely no great loss.

#### *Variables Associated with School Social Control in Taiwan*

Findings concerning the role of school social control in a Chinese context are as strong as they are for families – and possibly even stronger. This section covers school attachment,

commitment, and involvement. As with family social control, most available studies are cross-sectional in nature, but results from one or two high-quality longitudinal studies that do not use deviance as the dependent variable but involve other consequences of school social control are also available. School attachment, along with commitment as measured by grades and class rank, predicts lower delinquency, and while involvement in school conventional activities predicts higher delinquency the findings are not as conclusive.

### School Attachment

The literature in the PRC and Taiwan indicates particularly robust protective effects of school attachment on delinquency. While cross-sectional studies do not provide as convincing a set of results as longitudinal studies in general, some such studies in China and Taiwan have shed light on potential causal mechanisms. Studying 150 students in Tianjin between the ages of 15 and 18, Zhang and Messner (1996) find that family attachment is negatively associated with juvenile delinquency net of controls including family deviance, but these effects then become null when school attachment is included in the model with strong negative effects on delinquency – suggesting that the school has more direct effects on adolescent behavior in China than the family. A study of 636 11<sup>th</sup>-grade students in four high schools in Jiangsu Province, China (Li, Martin, Armstrong, and Walker 2011) also indicates that lower problem behavior is associated with attachment to school and expectations about behavior set by the high school.<sup>31</sup> Using a sample representative for all of Taiwan, Goo (2007) finds that both attachment to teachers and attachment to school in general is associated with lower delinquency for Taiwanese

---

<sup>31</sup> The authors refer to their measure for school attachment as “school commitment,” but an example question the authors present for their scale, worded “How interesting are most school subjects to me?” more closely resembles Mortimer’s (2003) measures for intrinsic motivation towards school, a measure of school attachment. For the present study, the boundary between school attachment and commitment is not all that critical to establish, but it is worth noting that various authors in the field mean different things when they use the word “commitment.”

junior high school students – and attachment to teachers has the strongest effect on delinquency of any variable in the model, triple the effect size of general school attachment.<sup>32</sup> Looking at Taiwanese students in 5<sup>th</sup>-9<sup>th</sup> grade in Nantou County, Liu (2003) also finds that the more students perceive their teachers as fair, the less likely they are to commit moderate forms of deviance (such as theft, gambling, and fighting), while this impression about the teacher has no effect on minor deviance (such as smoking and drinking) or serious deviance (such as drug use, motorcycle vandalism, or extortion). The implication of this particular study for the TYP deviance data, which consists of a general scale derived from a combination of measures that Liu defines as mostly minor and moderate deviance, is ambiguous, but overall attachment to teachers seems associated with lower deviance in Taiwan in cross-sectional studies.

Longitudinal studies also find positive outcomes associated with school attachment in Taiwan. In a high-quality analysis of four waves of the Taiwan Education Panel Survey (the most comprehensive longitudinal survey related to education in Taiwan ever conducted, with an initial sample of 20,004 students in 333 schools around Taiwan and four waves of data covering six years from junior high school through the end of high school), Lee (2011) uses latent growth curve modeling to find that school attachment explains both lower initial delinquency and desistance from crime, net of peer delinquency and family attachment; this effect is particularly strong for girls, at least double the effect size of nearly every other variable in the study. Notably, this effect for attachment<sup>33</sup> is about five times the magnitude of the effect of commitment.<sup>34</sup>

---

<sup>32</sup> In particular, Goo (2007) finds that a standard deviation increase in attachment to teachers is associated with a decrease in general deviance of 0.145 standard deviations.

<sup>33</sup> School attachment is a scale comprised of three Likert-scale questions concerning the extent to which respondents like their school, believe that they receive fair treatment and examinations, and believe that their school cares about its students.

<sup>34</sup> Lee conducts separate models for each gender as well as a joint model, and the result described here is for the joint model. In explaining how commitment is measured in the TEPS study, Lee (2011:44) explains that “success in school in the TEPS dataset is not estimated using classical measurement theory, but rather is a score estimated with the three-parameter model belonging to modern measurement theory, which is then placed onto a common scale for

Using the Taiwan Youth Project data, Chang (2010) finds that lagged school attachment predicts lower delinquency for junior high school students, also net of parental attachment, peer attachment, and educational aspirations but not peer delinquency. These two studies might have been stronger had they used fixed or random effects, but in establishing a clear temporal order they still provide much stronger findings than the cross-sectional studies, and Lee's analysis is particularly important in capturing change in deviance over time.

Finally, in a high-quality longitudinal design modeling depression among TYP survey respondents using a latent growth curve model, Yi et al. (2009) also demonstrate the pressure that Taiwanese students experience in high school and its relationship to depression. They show that family factors such as cohesion and parents' strategies for educating their children affect the children's initial levels of depression in junior high school, but subsequent change in depression through the end of high school is determined by school factors, not family factors. In what the authors posit to be evidence of just how stressful Taiwan's secondary education can be, classroom cohesion (the extent to which classmates support each other's academic endeavors) predicts lower initial depression, but a positive linear slope in depression. These findings suggest that as examinations approach, students feel an even keener sense of competition with classmates with whom they share a strong social bond, thus increasing their depression. This result is in keeping with in-depth interviews of TYP respondents and their families that find that students at the head of their classes are often unpopular (Pan and Yi 2011:377). Due to these findings, the selection model for work during high school in Chapter 4 includes classroom cohesion. The

---

all four survey waves using equating methods [等化方式], to enable comparisons across waves." The model sounds like a three-parameter logistic IRT model for which multiple survey waves were placed onto a common scale using vertical scaling techniques; respondents to the TEPS survey must have been administered a general knowledge test as part of the survey. Lee's thesis tests Hirschi's control theory among other things, although in contrast with Hirschi (1969), she refers to academic success as capturing involvement in school activities and educational aspirations as capturing beliefs; Hirschi describes both as measures of school commitment.

overall implication of these studies is that school attachment should be one of the strongest variables available in the TYP survey in explaining deviance, and cannot be omitted from a selection model without a high probability of producing biased results.

### School Commitment

Both educational performance and educational aspirations operationalize Hirschi's theory that commitment to conventional activities predicts lower crime (Hirschi 1969), and several studies use one or both of these measures to confirm the theory. High educational achievement, as measured either by class rank or by grades, consistently predicts lower delinquency in Taiwan (Huang and Lin 2009; Lee 2011; Sheu 1988; Wang and Jensen 2003). One of these studies uses the TYP survey and a fixed effects framework (Huang and Lin 2009), providing fairly convincing evidence that higher grades are associated with lower deviance by controlling for unobserved time-invariant within-individual heterogeneity.<sup>35</sup> Lee (2011) in a longitudinal study with data representative of all of Taiwan, demonstrates that an increase in performance on a standardized test over time is associated with a corresponding reduction in deviance. Especially as grades affect work patterns as well, with those with high grades tending to work fewer hours per week in the West (Mortimer 2003), educational commitment as operationalized by either grades or class ranking constitutes an important confounding variable for work as well as for deviance.<sup>36</sup>

Looking to mental distress as a dependent variable, high grades appear related to lower distress but those who care about grades may also be distressed by approaching entrance

---

<sup>35</sup> To compare deviance across waves, given the changing wording of the TYP delinquency survey items the authors restrict their analysis to drug use, drinking, and smoking during three survey waves.

<sup>36</sup> For some waves in the present research, class ranking is not available, so investment in homework is used as an alternate measure for educational commitment.

examinations. In a random effects model that accounts for between-individual and within-individual heterogeneity in strain, Huang and Chien (2013) show that students of higher academic ranking in high school have less trouble sleeping. Yi et al.'s high-quality longitudinal study of depression using the TYP data (2009) found, however, that students with higher grades were more likely to suffer from *increasing* levels of depression over time, suggesting a connection to the entrance examinations prior to graduation from high school. School commitment, in general, is associated with lower deviance, but may have time-specific effects on strain.

Hirshi (1969) theorizes that educational aspirations also measure commitment to school. Several studies (Chang 2010; Goo 2007; Sheu 1988), two of which are longitudinal studies of relatively high quality (Chang 2010; Goo 2007), find that high educational aspirations predict lower deviance, while Lin and Li (2013) show null effects, using longitudinal data and a latent growth curve model to estimate the lagged effect of educational aspirations on deviance. Using multivariate logistic regression methods, Goo (Goo 2007) finds that attachment to teachers explains the most about deviance, followed by educational aspirations; that said, the model only explains about 4.2% of the variance in delinquency.<sup>37</sup> Lee (2011) also finds that educational aspirations do not predict initial deviance in Taiwan during the first year of junior high school net of a raft of other social control variables and peer deviance. For males only, however, the more positive the rate of change in educational aspirations as adolescents age through junior high and high school, the more negative the rate of change in deviance. These findings are consistent with a “zero-sum” hypothesis which holds that children who lose interest in school are more

---

<sup>37</sup> The TEPS delinquency items consist of a set of questions about minor deviance similar to the TYP survey. In particular, TEPS respondents are asked if they cheated on a test, skipped class, fought or otherwise caused trouble in school, viewed pornographic magazines or CDs, stayed away from home, committed theft or vandalism, or “smoked, drank alcohol, or chewed betel nut.”

likely to become deviant – and one way they could lose interest, of course, is by intensively working outside school with little remaining time for homework (Greenberger and Steinberg 1986; Steinberg and Dornbusch 1991; Steinberg, Greenberger, Garduque, and McAuliffe 1982), or similarly by spending too much time dating.

High educational aspirations could lead to conventional behavior and away from crime, but they could also serve as a cause of strain and thus crime (Agnew 2006b) – a classical theoretical problem that Hirschi also recognized in *Causes of Delinquency* (1969-171). Yi et al.’s findings concerning grades and depression (2009) allude to this. Chang (2011), using a longitudinal design interviewing 415 7<sup>th</sup>-grade students three times over nine months in eight schools in southwestern Taiwan, examines the effect of educational aspirations on the onset of delinquency while controlling for strain. Students with high educational aspirations have a lower probability of initial delinquency net of strain, with a large effect size robust to model specification.<sup>38</sup> Chang’s findings are compelling, but also restricted to seventh-graders. In contrast to the importance of commitment, Chang finds no effect of self-efficacy on the odds of initial deviance; in other words, for his sample of 7<sup>th</sup> graders, their personal belief in their ability to succeed in school does not keep them from becoming deviant so much as the extent to which they desire and consciously work towards success.<sup>39</sup> In other words, grades and time spent on homework might predict lower deviance far better than aspirations or self-efficacy. When the above articles are taken together, the findings in Taiwan with respect to the effect of educational aspirations on deviance are inconclusive, but the highest-quality findings suggest that high educational

---

<sup>38</sup> An increase in educational aspirations of one unit above the mean is associated with a reduction in the odds of initial deviance within the next three months of 42.3%. These findings are robust to whether the effect of educational aspirations on the change in the odds of deviance between the first and second wave and between the second and third wave is assumed to be constant or allowed to vary.

<sup>39</sup> These findings remind me of a quote by the coach of the Los Angeles Bruins in explaining why his highly-acclaimed players barely beat a losing team: “I think the three worst words used are 'I got it,' or 'We got them.’” (Foster 2014)

aspirations are associated with lower deviance. With respect to work, using U.S. data Mortimer (Mortimer 2003) shows that respondents with high educational aspirations are less likely to report that their extracurricular workload is stressful and that they experience a trade-off between work and grades, thus linking educational aspirations to a so-called “agentic” model in which those who are forward-thinking in their decision-making derive more positive benefits from work. Educational aspirations, as a measure of commitment to school, are therefore a key control variable to include in a model for selection into work in Taiwan – but grades and time invested in school are probably stronger predictors of lower deviance.<sup>40</sup>

### School Involvement

Research to date in Taiwan concerning the effects of school involvement on deviance has yielded conflicting findings. In a cross-sectional analysis using one wave of the TEPS data, Goo (2007) finds that while attachment to teachers and to school in general is associated with lower delinquency, involvement in school activities is associated with *positive* delinquency. As the study did not control for peer delinquency, however, this unusual finding for school involvement may be upwardly biased due to unobserved peer deviance. Hirschi (1969) notes that time spent on homework is a traditional indicator of school involvement, and Sheu (1988) and Chang (2010) find that Taiwanese students who spend more time doing homework commit less crime. Chang’s findings are in a longitudinal framework and are more convincing. Studying Jiangsu, China, Li, Martin, Armstrong, and Walker (2011) find the opposite – involvement in and reward for pro-social high school activities is associated with low problem behavior – but the authors did not control for delinquent peers either. Additionally, their methodology, a cross-sectional

---

<sup>40</sup> Educational aspirations are not available for respondents observed in their college years and later, so the fact that grades may matter more than aspirations for predicting deviance is therefore reassuring.

analysis of about 600 students from four high schools in one province, is neither particularly rigorous nor necessarily representative of a broader population. Simply attending school also matters: in a fixed effects framework, Huang and Lin (2009) also find that net of peer deviance and lagged deviance, TYP respondents between ages 16 and 18 who are not in school are more likely to smoke, drink, and use drugs. The methodology is strong, but going to school as a binary indicator variable is at best a very coarse measure of involvement, a concept usually captured by measures such as engaging in extracurricular activities and possibly holding leadership positions. Finally, Chang Fung-Mei's longitudinal study of 7<sup>th</sup> graders in southwestern Taiwan (2011) uses a latent measure for involvement in school activities and finds that net of strain and commitment to school, a one-unit increase in involvement above the mean is associated with a reduction in the odds of initial deviance of 29.5%. These findings are promising, but restricted to junior high school. Overall, the findings for the effects of school involvement on deviance in Taiwan are weaker and less conclusive than those for school attachment and commitment, a fact that is not surprising given that Hirschi (1969) also found time spent on homework to be negatively related to delinquency after controlling for school commitment, but in general found involvement to be one of the weakest predictors of deviance. The selection models for work and deviance in the present chapter include involvement out of an abundance of caution.<sup>41</sup>

### *Variables Associated with Peer Social Control in Taiwan*

Relatively little is known about the effect of peer attachment on deviance in Taiwan because the available studies are either cross-sectional or use a single wave of longitudinal survey data, and do not control for peer deviance, and what research is available suggests that peer bonds are

---

<sup>41</sup> Additionally, for some years in college, the survey includes questions about time spent on homework but not about grades.

not as protective as bonds to school or the family. In a cross-sectional study that does not control for peer deviance, Goo (2007) finds that bonds to peers predict lower deviance, but the variable for peer attachment is measured using questions that could more adequately be described as “peer commitment to school.”<sup>42</sup> Chang (2010) and Liu (2003) both find no effect of peer attachment on deviance and use better measures of peer attachment that reflect an emotional connection to peers and have high reliability and validity. And comparing Japanese college students to U.S. college students, Fukushima et al. (2009) also note that peer attachment is positively associated with deviance in Japan and negatively associated with deviance in the U.S., suggesting the potential for a social learning mechanism to be at play in Japan (they do not control for peer deviance). All four of these studies are severely limited by the fact that they do not control for peer deviance, even as social learning theory accounts for a great deal of the variance of delinquency in the West – as much as a third (Akers and Jensen 2006). Peer attachment may also explain selection into part-time work to the extent that some youth may choose such work as a way to spend time with their friends, but I could find no studies from Taiwan to support or refute this hypothesis. Still, spending time with friends has been associated in the West with longer work hours for high school youth and both higher earnings and work-derived status (Mortimer 2003), and the fact that time spent with friends should be positively correlated with dating is also intuitive. Peer attachment is therefore included as a potential confounder in the present research because it is theorized by Hirschi (1969) to predict lower

---

<sup>42</sup> The questions related to peer bonds in the first wave of the Taiwan Education Panel Survey (TEPS) include “My classmates often discuss homework and reading,” “My classmates often discuss matters related to educational advancement,” and “I often go with friends to attend exhibitions and listen to presentations outside of school.” The fact that the author uses questions specifically referring to classmates as part of a variable measuring peers would likely be considered a methodological flaw in a US study, but in Taiwan reflects general knowledge that students are grouped into tight-knit clusters of classmates who take all the same classes and progress through grades together, such that a student’s group of classmates dominates his or her peer network during junior high and high school.

delinquency, and out of caution because relatively little is known about the role of this variable in explaining delinquency in Taiwan.<sup>43</sup>

*Attachment to the Workplace in Chinese Culture*

The role of the workplace as an institution affecting crime through social control is important for this research, but also not well-understood in a Chinese context, and the present review could only identify a single relevant study based in Mainland China, not Taiwan. Looking at members of state-run and collective work units in Tianjin, Zhang and Messner (1999) find the strength of the social bond to one's work unit (*danwei*) to be negatively associated with official delinquency after controlling for peer deviance, deviance in the neighborhood, family attachment, and other demographic background variables. That said, the magnitude of this protective effect is less than that of the family bond and 1/3 that of the deleterious effect of delinquent peers, the findings are limited by the cross-sectional survey design, and the fact that the *danwei* in China in the 1990s were mostly semi-government institutions sharply limits the extent to which inferences can then be drawn about bonds to work in Taiwan's free market system. The authors further point out that their findings related to school, family, and *danwei* social controls (Zhang and Messner 1995; Zhang and Messner 1996; Zhang and Messner 1999) could potentially be spurious, because the fact that they use official delinquency as their dependent variable implies that youth performing poorly in school, with parents who are deviant, or working in a disreputable *danwei* could be more likely to be picked up by the police due to labeling effects. The present study leaves out this potential confounder for practical purposes – consistent data on attachment to the workplace are not consistently available, and when the

---

<sup>43</sup> Concerning commitment and involvement, Hirschi emphasizes commitment and involvement in conventional activities and has nothing to say with respect to peers, so researchers testing social control theory typically do not investigate "peer commitment" or "peer involvement."

questions are asked in the TYP survey, they are not asked of all youth who had worked during the previous year.

### *Differences in the Nature of Moral Beliefs in Taiwan and the West*

Hirschi (1969) theorizes that attachment, commitment, and involvement in social institutions all act on deviance by strengthening a person's moral beliefs, so when measured, the strength of moral beliefs should constitute an important proximate cause of delinquency according to control theory. The present section covers a recent research finding in Taiwan that suggests that the mechanism through which attachment, commitment, and involvement affect beliefs may differ from that in the West – and while the findings are too recent for the designers of the TYP survey to have added variables to the survey, they still have implications for both the social context of the analysis and the selection of other variables less proximate to deviance that can also serve as controls in light of these new findings. Using the Taiwan Youth Survey, Wu and Jou (2009) demonstrate that moral beliefs as a construct are different in Chinese culture compared to Western culture. Key to the authors' argument is the Chinese cultural belief that each person is situated in a network of relationships similar to concentric circles centered around one's family, an idea known as *chaxu geju* (差序格局) and central to the concept of filial piety. Confucian values contrast with Western moral values that are dominated by equality under the law and “minimal virtue,” a perspective that any action that does not impinge upon the rights of another should be allowable in public (Wu and Jou 2009). They instead center on positive actions to promote the well-being of one's immediate family, only then followed by one's own self, extended family, school, workplace, city, province, and so on. Indeed, the reason for competition with one's peers in Chinese culture is to bring honor to one's family and ancestors, opening up

the possibility of committing crime for the purpose of furthering one's family's interests – as long as one does not bring dishonor to the family in doing so (Hsu 1948).

This cultural difference suggests a refinement in Chinese culture of Hirschi's theory that all social bonds affect moral beliefs in the same way. Instead, attachment to the family should have the greatest effects on beliefs and thus crime, with attachment to institutions such as the workplace mattering much less. Wu and Jou (2009) thus hypothesize that the greater the likelihood that someone socially close to one's immediate family would find out about or be impacted by a deviant action, the less likely that person is to actually do it. In this context the prevalence of deviant behavior such as music piracy, for example, makes sense because it is difficult to prove and the victim is a large corporation that is socially distant from the offender. They then suggest that moral beliefs can be separated into two domains that they refer to as "private virtue" (私德) and "public virtue" (公德), and measure virtue in these domains using eight TYP survey items about how wrong respondents perceive various types of deviant behaviors to be.<sup>44</sup> Because students in Taiwan interact with the same set of classmates all day in junior high school and high school and become quite close, the authors theorize that acting

---

<sup>44</sup> To measure private and public virtue in Taiwan, the authors cite the results from an online survey of moral values that Taiwanese residents consider most important (Yu 2003), and sort the results into four categories that have "public" and "private" dimensions: benevolence (仁愛), duty (責任), honesty (誠信), and temperance (節制). They then identify questions asked of TYP respondents during high school concerning how wrong they consider each of eight behaviors that touch upon the "public" and "private" side of the above virtues. In particular, for actions violating benevolence they use not showing filial piety to one's parents ("private") and not yielding one's seat to an elderly passenger on a bus ("public"); for duty, they use skipping class ("private") and breaking traffic laws ("public"); for honesty, they use forging a parent's signature or using their chop to stamp a document ("private") and using pirated software ("public"); and for temperance, they use having sex ("private") and cutting in line ("public"). A limitation of the available data is that a single item measures each component of public and private virtue. Additionally, these various actions differ in their severity, such that a respondent's propensity to state how wrong a particular action would be may have less to do with whether the action is "public" or "private" and more to do with the level of harm it could potentially cause, regardless of how socially close the recipient of such harm is to the respondent. Future research methods could therefore potentially be improved by the use of IRT methods as presented in Chapter 2. The authors find in a confirmatory factor analysis that each of the four questions loads well onto latent factors for public and private virtue. (They do not, however, report the results of a test comparing whether separating virtue into "public" and "private" explains more of the variance in moral beliefs than simply using one factor ("virtue") measured by eight items.)

morally towards one's classmates and teacher (in addition to one's family) is part of the "private" component of virtue. Next, they examine the effect of social control on beliefs,<sup>45</sup> and in a multivariate regression analysis they find that net of control variables, both public and private virtue are positively predicted by family attachment, family cohesion, and attachment to one's teacher, while classroom cohesion does not explain moral beliefs (either public or private). Notably, Wu and Jou are able to explain twice as much of the variance in private virtue as in public virtue, and the effect sizes of these social control variables are larger. The authors then conclude that competition within the classroom may hinder the ability to learn pro-social values from one's classmates, as supported by in-depth interviews of TYP respondents and their families that highlight the unpopularity of students with the highest grades in their classes (Pan and Yi 2011:377).

Given the problems with the measurements the research must be replicated, but it still breaks significant new ground despite its flaws. The importance of this work for the present research is in the implication that moral beliefs in Taiwan may be multidimensional, with social bonds to school and the family explaining more about beliefs than peer attachment. Ideally the selection models for work and marriage in the next two chapters would directly include moral beliefs as a time-varying covariate, but questions involving moral beliefs are not included often enough in the Taiwan Youth Survey often enough to be time-varying; as moral beliefs themselves are a consequence of deviance (Matsueda 1989), it would be inappropriate to treat them as time-invariant. Accordingly, because these differing dimensions of beliefs all appear to be primarily affected by school and family attachment and due to a paucity of time-varying data on beliefs in the TYP survey, the selection model omits beliefs but captures school and family attachment and

---

<sup>45</sup> Interested as they are in moral beliefs in Taiwan as a topic for study, the authors do not state that they are testing Hirschi's theory, although that is what they are doing.

commitment with some rigor. Wu and Jou (2009) also link family cohesion to moral beliefs, so family cohesion is also included in the selection model as a partial proxy for beliefs as fostered by the family.<sup>46</sup>

### *Other Differences Between Social Control in Taiwan and the West*

While the studies reviewed so far demonstrate a link between school and family social control and lower delinquency in Taiwan, social control may not be as important a factor in Taiwan as Chinese cultural roots would suggest: Fukushima and her colleagues (Fukushima, Sharp, and Kobayashi 2009), although comparing delinquency in the U.S. to Japan and not Taiwan, provide an important counterpoint to the common argument that bonds to the family, school, and peers are stronger in Asian countries with traditionally Confucian views. They find that Japanese college students are much less strongly bonded to society than their American counterparts using as measures parental monitoring and attachment, school involvement, peer attachment, commitment to conventional social goals such as graduating from college, and involvement in social organizations. Only school attachment is stronger among Japanese students than their American counterparts; the authors posit that a sense of social malaise among Japanese youth and rising levels of inequality following its economic recession may be to blame, a situation mirrored in Taiwan.<sup>47</sup>

The above research into social control in Taiwan and the PRC suggests that school and family attachment and commitment both serve as strong sources of social control associated with

---

<sup>46</sup> See Appendix E for the specific questionnaire items used to capture family cohesion in the TYP survey.

<sup>47</sup> Studies involving cross-cultural comparisons are difficult to design. The authors compare white students in an American university to non-foreign-born Japanese students in a Japanese university in a location with a similar level of urbanization, and translate and back-translate the surveys to ensure that the questions are the same. The wording of the questions themselves are based on Hirschi's control theory questions. As the authors do not make use of techniques such as anchoring vignettes for conducting cross-cultural research, it is still possible that the differing findings for Japan and the US are a result of different cultural interpretations of the survey items.

lower crime and should be controlled for in selection, while peer attachment may be associated with lower crime but the findings are complicated by imprecise measurement designs and studies that do not control for peer deviance, and the findings for school and family involvement are sparse and mixed. The selection models include covariates for peer attachment, school involvement, and family involvement out of an abundance of caution. Moral beliefs are trickier to study: one study (Deng and Roosa 2007) finds that beliefs may have their own independent effects on deviance aside from the attachment, commitment, and involvement components of social control, with some gender-specific effects of beliefs as well; some evidence also suggests that moral beliefs are in fact a two-dimensional construct in Taiwan, with beliefs involving deviant behavior that is more likely to affect one's family and friends more closely linked to school and family social control (Wu and Jou 2009). Beliefs are not included in selection models in the present case due to lack of available data. Additionally, findings from Japan suggest that social bonds in Taiwan may be weaker than the focus on filial piety in Confucian thought may imply. One covariate available in the TYP survey but which is not mentioned in the literature on causes of delinquency involves the parents' aspirations for how far their child will go in school, net of the child's own aspirations. Because the child's aspirations are so predictive of both educational advancement and lower delinquency in Taiwan and because of the importance of parental attachment as well, the selection models also include parental aspirations for the child's educational attainment.<sup>48</sup>

---

<sup>48</sup> Note that parental aspirations for the child's educational attainment were at first included in the selection models, but excluded from the results presented in this dissertation due to both a high percentage of missing values for this covariate and the fact that those remaining cases predicted neither selection into work nor selection into marriage.

## Strain Theory

U.S. findings showing a link between strain and delinquency (Agnew 2006a; Agnew 2006b) have been reproduced in China and Taiwan. Strain theory, developed originally by Merton (1938; 1957) and extended to describe individual behavior by Agnew (Agnew 1992; Agnew 2006b), holds that sources of emotional pressure, either chronic (such as poverty) or immediate (such as getting cursed), can break through a person's mental discipline and render him or her temporarily vulnerable to learning definitions favorable to crime from delinquent peers. These findings have partially been borne out in Chinese research. One of the most relevant variables involving strain for the present research is related to pressure from school. A cross-sectional test of Agnew's General Strain Theory (GST) among 615 Mainland Chinese youth in grades 8-11 (Bao, Haas, Chen, and Pi 2014) finds that negative treatment by teachers and parents is associated with a weakening of attachment to school and the family, and a related increase in delinquent friends and propensity to commit delinquent acts. An earlier study with the same sample (Bao, Haas, and Pi 2007) showed that social support mediates the effect of strain among these students: parental social support protected from delinquency students whose teachers treated them badly, while teachers' social support in turn protected from delinquency those students whose parents treated them badly.<sup>49</sup> Contact with deviant peers actually amplified the criminogenic effects of strain caused by negative relations with either parents or teachers, as well as with peers in the case of boys. Liu (2003), in a cross-sectional study of Taiwanese students in grades 5-9 in Nantou County, finds positive effects of school-related strain on deviance net of family, school, and peer social control; strain, however, affects only minor forms of deviance (such as smoking and drinking) but not moderate or more serious crimes. His measure of strain

---

<sup>49</sup> By "social support," the authors refer to assistance in overcoming hardship, either in terms of emotional support or social network resources.

is, however, a composition of several questions that seem to cover self-control and academic performance as well as actual strain (perceived unfair treatment by teachers at school), so the findings are suspect.<sup>50</sup> They do imply, however, that strain due to school may have direct effects on deviance.

Results from two longitudinal studies suggest negative outcomes of strain at school, but also show that students are able to acclimate to strain over time (as long as they stay in school). An analysis of the effect of strain on depression using the TYP survey (Yi, Wu, Chang, and Chang 2009), which measures strain as perceived unfair treatment by teachers, shows that depression in general peaks during the years that students are taking their high school and college entrance examinations. Students in both junior high school and high school who are under strain experience higher baseline levels of depression, yet the subsequent rate of change (linear slope) of depression is *negative*, suggesting that students may actually acclimate to their teacher's authoritarian style, and possibly even appreciate it given the teachers are trying to help them pass the exams.<sup>51</sup> Chang (2011), studying juvenile delinquency among 415 Taiwanese 7<sup>th</sup>-graders surveyed three times over an academic year, uses one of the best measures of strain in this review<sup>52</sup> and finds no main effect of strain on delinquency in an event history analysis – but does

---

<sup>50</sup> Liu (2003) constructs a scale for “pressure from coursework” using the following eight Likert-style survey questions:

1. “I will definitely prepare whatever my teacher asks me to prepare for class.”
2. “My teacher often praises my academic performance.”
3. “I am able to complete my homework on time.”
4. “I am interested in my textbooks.”
5. “I like my classes in school.”
6. “I work hard in studying for classes at my school.”
7. “I am able to relax when playing games and focus when doing homework.”
8. “I actively plan out my study time.”

<sup>51</sup> A quadratic term for the effect of the “unfair teacher” variable on depression is positive, however, suggesting that as the examination date nears, students’ levels of depression either decline at a smaller rate of change or actually reverse course and increase again.

<sup>52</sup> In Chang’s study, strain is a latent variable obtained from four questions having to do with fears about an inability to perform well in school, and is based on General Strain Theory (Agnew 2006b) and previous questionnaires administered in Taiwan (Chang 2011:54)

find significant interaction effects between survey wave and a latent variable for strain, such that a one-unit increase in strain for a student in the fall compared to the same increase for a student in the winter is associated with a 147% increase in the odds of deviance!<sup>53</sup> Overall, only a few high-quality findings show a relationship between strain at school and deviance, while others are poorly designed and their findings are therefore suspect. The results, however, are all in the same general direction demonstrating the relevance of strain to the “zero-sum” hypothesis for the effect of work on crime (e.g. Greenberger and Steinberg 1986), so this variable is included in the selection model.

A second variable relevant to strain and included in the selection models in Chapters 4 and 5 that explains delinquency in Taiwan is related to stressful life experiences such as the loss of a family member, breaking up with a boyfriend or girlfriend, or having a parent lose their job. Auerbach et al. (2009) find that in a high school in Hunan Province, stressful life experiences completely mediate the effect of materialism on delinquency (although the materialism itself may cause the strain in the first place). Lin Wen-Hsu and his colleagues (2014), comparing two groups of about 1,500 adolescents in Taiwan and the U.S., use structural equation models with longitudinal data to find that in both the U.S. and Taiwan, victimization and stressful life events are directly associated with delinquency even after controlling for anger and depression, whereas strains associated with an inability to achieve one’s goals or a perception of one’s life circumstances as unfair affect delinquency only indirectly by increasing both anger and depression. The results for the U.S. and Taiwan sample are in fact so close that in some cases

---

<sup>53</sup> Chang uses a proportional discrete time event history model, which assumes that the covariates have the same multiplicative effects on the probability of initial deviance in the second wave compared to the first wave as they do for the third wave. As this assumption has no theoretical basis, as a sensitivity check he relaxes the assumption and runs a non-proportional discrete time model, but oddly reports a different referent category for the main effect of wave, so the interaction effects for strain and wave cannot be compared across the two models. For the non-proportional discrete time model, a one-unit increase in strain in the fall compared to a similar increase in the spring is associated with a 95% increase in the likelihood of committing a first delinquent act in the fall.

their structural equation models are more parsimonious if instead of estimating separate effects of strain on delinquency for each sample, they constrain the coefficients to be exactly the same! They do not control for peer deviance, but Lin Kuei-Hsiu and Li (2013), also using the Taiwan Youth Project survey, do find small but significant direct effects of stressful life events on delinquency net of delinquent peers.<sup>54</sup>

The literature also points to several mediators of strain. Huang (2006a) finds that family social control may mediate strain: children from intact families who experience conflict with their teachers due to problem behavior in school, instead of expressing defiance (the type of resolution theorized to produce the most strain for the student and teacher), are more likely to resolve the conflict through cooperation, obedience, compromise, or at least avoidance of the topic. The implication is that either children learn better coping strategies from their parents, or else the parents cooperate with the teachers to deter defiant behavior on the part of the child. Additionally, Cheung and Cheung (2010), using a cross-sectional sample of 1,015 Hong Kong students aged 14-19, find that self-control mediates the effect of strain on delinquency, but the effects are gender-specific; self-control shields girls but not boys from stressful life events, negative school experiences, and poor relationships with their peers, and shields boys but not girls from a sense of an inability to achieve one's educational goals. Bao, Haas, and Pi (Bao, Haas, and Pi 2007), in a paper focusing on strain and delinquency in China, also find that moral beliefs unfavorable to crime mitigate the criminogenic effects of strain caused by parents and teachers for both boys and girls, although the research is limited by its cross-sectional nature.

---

<sup>54</sup> The procedure that Lin and Li use to measure stressful life events is not ideal, in that when creating a scale for stressful life events, the authors simply use an unweighted sum of indicator variables associated with whether respondents experienced each of several traumatic events in the past year, which range in severity from mild events such as the death of a pet to serious event such as the death of a parent. Improving on this method is outside the scope of the present research, so I use the same approach, despite the comical nature of assigning the same weight to the loss of a pet and the loss of a parent.

Oddly, coercive parenting actually led to lower delinquency for the boys in the sample, and this effect was stronger for boys with low self-control, suggesting that the authoritative parents shielded their sons (but not daughters) from crime more effectively if the sons had poor self-control. Together, these studies make a strong case for stressful life events being included in the selection models in this dissertation. Cheung and Cheung's study (2010) also makes a weak case for including harsh and erratic discipline (cf. Glueck and Glueck 1950) and a respondent's perception of low socioeconomic status as proxies for coercive parenting and educational goal blockage, so these two variables are also included out of an abundance of caution. Based on the above research, this dissertation uses stressful experiences at school, stressful life experiences, perceived poor socioeconomic status, and harsh or erratic parental discipline as variables to control for in selection models for work and dating. Auerbach et al. (2009) also call attention to the potential criminogenic effects of materialism, which they find is mediated by stressful life events but that appears relevant to both the decision to work or go out on dates and the decision to offend. Materialism is therefore controlled for in the single wave in which it appears in the survey (Wave 9), and while limitations of the data prevent its inclusion in previous waves, the inclusion of stressful life events as a confounder in the selection model for previous waves may reduce the impact of its omission.

Overall, studies of strain in Taiwan to date show only a weak relationship between strain and deviance. Those showing a positive relationship either use poor measures of strain that are not true to Agnew (2006b), or else are cross-sectional in nature; longitudinal studies show students acclimating to strain at school over time, while other research shows that self-control and parental social control also mediate strain. Chapters 4 and 5 include measures of strain in the selection models, but they are not likely to be critical predictors of deviance, both due to the

weak findings for strain in the literature related to Taiwan and as these measures are themselves are not as good measures of strain as those proposed by Agnew.

### *Social Learning Theory*

While social learning theory is rich with a number of theorized mechanisms for how youth learn definitions favorable to crime from their peers, the TYP survey only contains measures of peer deviance, so the current review sticks to research that employs peer deviance as an explanatory variable for deviance. Several high-quality studies from Taiwan and the PRC show strong effects of peer deviance on juvenile delinquency that are consistent with those in the West and provide strong evidence for the need to control for it in the present research. Wang and Jensen (2003) include peer delinquency in Taiwan<sup>55</sup> in a joint model with social control, and find strong, positive effects of peer deviance that explain about 16.5% of the variance in general delinquency and hold for both boys and girls. Another high-quality study (Huang and Lin 2009), this time using the TYP data and fixed effects models that control for unobserved time-invariant within-individual heterogeneity, also found strong positive effects of peer smoking, drinking, and drug use on individual smoking, drinking, and drug use. A second investigation using the TYP found lagged effects of peer smoking or drinking at age 16 on smoking at age 20 and drinking at age 20 (Chiao, Yi, and Ksobiech 2014). And Lee (2011), fitting a latent growth model of deviance for 4,163 Taiwanese junior high school and high school youth observed over four years using the Taiwan Education Panel Survey (TEPS), found strong, positive effects of both the intercept and slope of peer deviance on respondents' own deviance, after controlling for self-control, school and family attachment, educational aspirations, and grades. A robust set of findings therefore shows that peer deviance predicts deviance in Taiwan.

---

<sup>55</sup> Their measure of peer delinquency is association with street gangs.

Social learning in a Chinese context can also weaken the effects of social control that steer people away from crime. Zhang and Messner (1995) find that parental deviance can erode the bond between parent and child in China, creating the space for deviant peers to pass values favorable to deviance on to the youth. Drawing on cross-sectional data gathered from 800 youth in northeastern China, the authors hypothesize that because family ties are so strong in China, the deviance of parents will be particularly salient in explaining that of their children. They first follow Sampson and Laub (1993) in hypothesizing that parents who are deviant are more likely to discipline their children harshly and erratically, eroding the attachment that children may feel for their parents. This weaker parental attachment may directly lead to greater juvenile delinquency due to an erosion of social control, and could also do so indirectly by allowing the children's values to be more easily shaped by those of delinquent peers. Using official delinquency as their dependent variable, they find that weak parental attachment is indeed associated with greater juvenile delinquency, but it is then completely mediated by peer delinquency, while parental delinquency itself retains a strong positive effect on juvenile delinquency. They conclude that consistent with social learning theory, parental deviance can erode the bond between parent and child in China, creating the space for deviant peers to pass values favorable to deviance on to the youth. Additionally, they find that youth can learn deviance directly from their parents through a separate social learning process. They also find that net of deviant peers, parents who do not discipline their children harshly and those who do not spoil their children have less deviant children, implying that social control still has a role to play net of social learning. Given that parental deviance is unavailable as a control variable in the TYP survey, their findings reinforce the importance of controlling for parental discipline and

parental attachment as potential mediators of the relationship between parental deviance and juvenile delinquency in the selection models in the present research.

Zhang and Messner's findings would be stronger had the data been analyzed using a structural equation framework, but they nonetheless are pioneering. Bao, Haas, Chen, and Pi (2014) essentially replicate their work while bringing in school attachment and using a structural equation model. They find that school and family attachment only act on delinquency indirectly by affecting involvement with delinquent peers. The latter finding provides some evidence, despite the limitations of the cross-sectional study, that delinquency is explained by social controls steering youth away from deviant peers, rather than by deviant peers eroding the bond between youth and their parents and teachers.

In a pioneering but ultimately inconclusive study, Wu (2011) uses dynamic social network analysis to investigate the social learning of deviant behavior in Taiwan and calls into question some assumptions about how behavior is imitated in a collectivist culture. He takes advantage of the fact that as the TYP survey is a stratified cluster sample, for each junior high school that was sampled, every student in each of two classrooms was surveyed. Accordingly, Waves 1-3 of the J1 cohort contain information about peer network effects on deviance. Wu uses the SIENA package (Snijders, van de Bunt, and Steglich 2010) to examine the dynamics of a single classroom of 28 students over three years, and to determine whether a youth's deviance or the deviance of a potential friend impact their propensity to become friends. In this model predicting the change between one wave and the next in the probability of a friendship occurring between the respondent and a potential friend, or "alter," a significant effect of the respondent's own deviance would suggest that deviant students seek out more friends; a significant effect of the alter's deviance would suggest that deviant students are more popular; and a significant

interaction effect between a respondent's own deviance and an alter's deviance indicates a social learning effect: net of the main effects of the respondent's deviance and alter's deviance on the odds that they become or remain friends, youth with deviant peers are more deviant.

Surprisingly, despite the consistent results in the literature that show a positive effect of peer deviance on individual delinquency, *Wu finds all three effects to not be significant*. In other words, students in the classroom he studied do not take deviance into account at all when choosing or retaining friends, deviant students are just as popular as other students, and students do not appear to be learning to be deviant from their friends. Rather than concluding that social learning does not take place in Taiwan, however, Wu concludes that the social learning of deviance is not a dyadic process. Instead, the number – or the proportion – of a Taiwanese junior high school student's deviant friends may be more salient in predicting whether he or she would choose to be deviant, a conclusion in keeping with the collective nature of Chinese culture. This study relies on just one classroom, however, and should be replicated using other classrooms observed in the Taiwan Youth Project study and elsewhere. After all, the small sample size, together with the low prevalence of self-reported deviance in the TYP, could be driving the null findings.

In a strong follow-up study, Wu and Lei (2013) group all J1 cohort respondents to Wave 1 of the TYP survey into classrooms, and construct classroom-level deviance measures from self-reported deviance. Next, in a multilevel framework they find that classroom-level delinquency predicts individual delinquency. These results confirm their suspicion that social learning does take place in Taiwan but is not a dyadic process. They further find that respondents with deviant peers are more deviant, and parental attachment does not predict lower deviance in the presence of this deviant peer variable (providing some support that social learning theory is more

predictive than social control theory in Taiwan). But what perhaps is most remarkable is the significant, positive cross-level interaction between classroom-level deviance and deviant peers that they find. In other words, *youth are most prone to deviant behavior when they interact with deviant peers in a classroom full of deviant students*. The main effect of classroom deviance in this model, however, is no longer significant at the 95% level, suggesting that students who do not have deviant friends are unmolested by the overall classroom context. This last finding suggests that perhaps youth can express agency in selecting their friends within a classroom to avoid deviance, even if they have less direct control over the classmates they have.

In summary, the articles reviewed here demonstrate the high importance of controlling for peer deviance in selection models for work and dating. They also suggest potential interaction effects between classroom cohesion and peer deviance (Wu and Lei 2013) and between peer deviance and parental attachment (Zhang and Messner 1995). While the selection model is already complicated enough and these interactions are not included in this dissertation, they denote possible avenues for future research. The implication of the findings concerning social learning with respect to work and crime in Taiwan is that young people can potentially learn deviant behavior from their coworkers – behavior that they could then emulate given stressful circumstances and weak social bonds. This fact is particularly important in high school because in Taiwan's educational system, secondary school students usually attend all their classes with the same group of 30 or so people, limiting the number of peers with whom they interact compared to Western schools. Accordingly, coworkers could affect youth more than in the West because friendships with them may constitute a larger portion of a young person's overall social network.

### *Social Disorganization Theory*

While strain, social control, and social learning constitute the bulk of research into delinquency in Taiwan, some solid initial research into effects of community-level variables on delinquency has been conducted. A long history of research into the effect of neighborhoods on crime states that neighborhood-level characteristics including population density, ethnic heterogeneity, mobility, and mean income have direct effects on crime, and also indirectly affect delinquency via residents' involvement in conventional activities and peer deviance (Sampson 2012; Shaw and McKay 1942:228; Yang and Hoffmann 1998). This section considers the use in selection models of TYP survey variables measuring respondents' attachment to their neighborhoods, involvement in neighborhood activities, and their perceptions of their neighborhoods as unsafe. It begins with one investigation of community-level effects on delinquency among respondents to the TYP survey that is particularly relevant to the present research (Chen and Wu 2004), then covers empirical tests of social disorganization theory in Taiwan.

Wu and Lei's study of peer deviance and classroom-level delinquency (2013) actually is based on a proposition that classroom contexts matter more than neighborhood contexts in Taiwan, a situation different from the West because students spend more than eight hours each day in the same classroom, with the same set of classmates. Accordingly, they generate three-level hierarchical models, nesting classrooms within neighborhoods because students in the same classroom live in the same neighborhood. After demonstrating that variance across classrooms but not neighborhoods predicts delinquency, they conclude that social disorganization theory should be reformulated for the case of Taiwan to reflect the importance of classroom dynamics

and not neighborhood dynamics. The fact that after the TYP respondents begin high school, classroom-level deviance is no longer captured in the survey is therefore quite a shame.

Yang and Hoffman (1998) also provide an important empirical test in Taipei of social disorganization theory, as originally stated by Shaw and McKay (1942). According to Shaw and McKay, residential mobility, population density, and ethnic heterogeneity should positively predict crime. Taipei was too ethnically homogeneous when the authors gathered their data in 1991 to test whether ethnic diversity predicts crime, but the authors use census data and a cross-sectional stratified random sample of 1,704 high school students in Taipei to test the rest of the theory. Family mobility and income affect deviance indirectly by predicting involvement in community social activities and peer deviance, but have no direct effects on individual delinquency; family disorganization<sup>56</sup> is, however, directly associated with individual deviance. Concerning neighborhood effects on crime, consistent with social disorganization theory household income as measured at the community level is negatively associated with individual deviance, and population density is positively associated with deviance.<sup>57</sup> Oddly, however, residential mobility is *negatively* associated with individual delinquency. The authors posit that these unusual results could be due to affluent residents being transferred to other parts of the city, but do not explain why transfers would necessarily cause people to move when the city itself is rather small; they also suggest that the discrepancy could simply be due measurement error in the data they used to capture migration.<sup>58</sup>

---

<sup>56</sup> For Yang and Hoffman, family disorganization is a scale based on whether both parents are in the household and whether they fight.

<sup>57</sup> A related time series analysis at the macro level (Denq, Lee, and Vaughn 1996) links the Gini coefficient to the number of drug arrests per 100,000 residents, suggesting that income inequality predicts crime in Taiwan. This finding occurs at such a macro level, however, that it only hints at the effect of household income at the neighborhood level on individual deviance.

<sup>58</sup> Yang and Hoffman's findings concerning residential mobility may not be an accident, however: using a 2004 household survey of 2,500 adult residents of Tianjin, China, Zhang, Messner, and Liu (2007) also find residential mobility to be negatively related to the incidence of burglary in a community, even while at the individual level, the

Social disorganization at the neighborhood level also influences deviance in Taiwan by affecting parenting practices. Following the *parental buffering* hypothesis, effective parenting practices protect children in neighborhoods with high social disorganization from delinquency, while the *evaporation* hypothesis posits that neighborhoods with high social disorder decrease the effectiveness of parenting in preventing juvenile delinquency (Simons, Lin, Gordon, Brody, Murry, and Conger 2002). To determine which of these hypotheses hold for Taiwan, Chen and Wu (2004) use government statistics corresponding to the villages in which youth respondents to the Taiwan Youth Project lived during junior high school to estimate community disadvantage<sup>59</sup> and collective socialization (presence of informal social capital, particularly adult role models, in the community),<sup>60</sup> then interact these variables with a Wave 1 “parenting behavior” scale comprised of parental attachment, parental monitoring, and harsh/erratic discipline (which is reverse-coded). They then use these covariates to predict delinquency in Wave 3, controlling for Wave 1 delinquency.<sup>61</sup> Their results support the evaporation theory: an interaction between parenting behavior and community disadvantage is positive in predicting delinquency, meaning

---

length of time living in a community is negatively associated with burglary risk. The authors argue that “newer neighborhoods have access to more elaborate security arrangements” (p. 932). One possibility for this discrepancy between Western and Chinese findings concerning residential stability is that criminal gangs such as the *jaotou* in Taipei more effectively operate in neighborhoods with low mobility, as they “are reluctant to recruit people who are not from their neighborhoods and communities” (Chin 2003:26).

<sup>59</sup> Community disadvantage is a sum of six standardized indices: the percentage of women over 15 who are divorced; the ratio of old to young; the ratio of dependents to total population; and, respectively, the percentage of the population age 15 or older that have a college degree, are disabled, and are in the labor force.

<sup>60</sup> The authors use three subscales for collective socialization: community cohesion (e.g. how often students interact with others in their community), participation in community activities, and “community strength” (students’ evaluation of the cleanness, safety, and “convenience” of the community).

<sup>61</sup> A multilevel structural equation model controls for biased standard errors associated with the fact that many respondents are grouped into the same classrooms. The authors then inspect the intra-class correlations for delinquency and find no significant evidence that students in the same classrooms are more likely to have similar levels of deviance than random chance would predict; they retain their multilevel approach, however, as some slight clustering does take place for the school performance variable (suggesting that Taiwanese children are placed into classrooms with classmates of similar abilities even in seventh grade). The implication for the present research is that I can safely ignore classroom-based effects on delinquency for respondents to Wave 3 and earlier, and once students move into different high schools and colleges, clustering effects by school would be further diluted because of the large diversity of high schools and colleges in Taiwan (and the fact that a visual inspection of the list of school codes suggests that students are attending different schools).

that parents who discipline, monitor, and maintain an emotional bond with their children are less effective at steering them away from crime when their neighborhoods are disadvantaged; similarly, high-quality parenting behavior predicts higher school engagement, but parenting is less effective in disadvantaged communities.<sup>62</sup> The interaction term between parenting and collective socialization is also negative, indicating that parents are less effective at protecting their children from crime in neighborhoods with low collective socialization. Additionally, they find that students who are less engaged in school have higher delinquency when the classrooms themselves are less cohesive.<sup>63</sup> Taken together, these findings suggest that for youth in Taiwan, social capital and role models matter both at the classroom and neighborhood level in deterring deviance, and work through reinforcing students' participation in conventional activities (in the case of the school) and parental monitoring, discipline, and affection (in the case of the community). Additionally, collective disadvantage appears to weaken the effects of social control.

In summary, these studies provide evidence that social disorganization, as operationalized by perceptions of a neighborhood as dangerous, may predict individual deviance. Some evidence suggests a possible interaction between social disorganization and parental attachment in affecting deviance, and Yang and Hoffman (1998) have found some direct effects of involvement in neighborhood activities on lowering deviance. The review uncovered no research into the effects of an adolescent's attachment to the community on deviance in Taiwan, but that variable is included in the selection model for work in lieu of direct information about the

---

<sup>62</sup> The authors use for school engagement a scale comprised of three summed binary indicator variables for whether students were leaders in school clubs, class leaders, or teaching assistants; five summed binary indicator variables regarding whether students had participated in various school clubs; and a Likert scale-based measure based on class rank.

<sup>63</sup> Classroom cohesion is a scale comprised of questions such as "Classmates always help each other," asked of each student in a classroom and aggregated into class-level covariates for the multilevel model.

number of adults in the community that an adolescent knows personally – a variable that could both reduce deviance and help the person find a job. Unfortunately, neighborhood effects are not available in the TYP surveys for the waves after most respondents have finished high school, and are thus excluded from models of selection into marriage, constituting a limitation of the present research.

### *Self-Control Theory*

Gottfredson and Hirschi's general theory of crime (Gottfredson and Hirschi 1990) assumes that society has a dominant moral order to which its members conform to a greater or lesser extent, and a person's level of self-control determines the extent to which they conform and explains the vast majority of crime. A significant shortcoming of the TYP survey is that measures of self-control are absent from the questionnaires. Gottfredson and Hirschi (1990, esp. 108-111) make a case, however, that low-severity crimes such as skipping class and cheating on tests themselves could be used as measures for self-control, a fact that demonstrates exactly why a person's previous involvement in crime is one of the best predictors of crime in the future. Therefore, given that the current research is not a test of Gottfredson and Hirschi's general theory of crime but rather seeks to eliminate variables that confound the relationship between work and crime, using lagged deviance as a proxy variable for self-control is not an implausible approach. Because actual measures of self-control would of course be preferable to using lagged delinquency to capture this theory, however, this section of the review covers explicit tests of self-control theory in Taiwan to investigate its explanatory power in a Chinese context and assess the potential risk associated with the omission of a direct measure of self-control from the selection model.

A few studies in Taiwan find support for Gottfredson and Hirschi's general theory of crime (Chang 2002; Sheu and Mon 1997), but even they often find that social control or social learning are more powerful predictors of deviance. In a cross-sectional study, Sheu and Mon (1997) gather data on parental monitoring and attachment, school teacher attachment and monitoring<sup>64</sup>, self-control<sup>65</sup>, and delinquency from a high-deviance group of 720 children with cases in Taiwan's juvenile court system and from a control group of 630 elementary and junior-high school students from across Taiwan. As they predict, the youth with cases in the juvenile justice system are more deviant and report lower self-control, parental attachment, and school attachment than those in the control group. They posit that following Gottfredson and Hirschi's theory, school and family attachment should only act on deviance indirectly by affecting self-control in a structural equation framework, but they find that structural equation models with direct effects of parental attachment on deviance fit the data better and completely eliminate the effects of self-control on deviance (for the control group) or else yield parameter estimates for the effect of parental attachment on deviance that are greater than that of self-control on deviance (for the high deviance group).<sup>66</sup> While the authors conclude that their research supports Gottfredson and Hirschi's general theory of crime, most of the support for this conclusion relies on their comparison of baseline characteristics of parental and school attachment, self-control, and delinquency for the two survey groups and not on the structural equation models. The most robust findings from the study in fact favor social control.

---

<sup>64</sup> True to Hirschi (1969), the authors measure parental and school attachment as a combination of monitoring by parents or teachers and affection that the parents or teachers have for the youth, as reported by the children.

<sup>65</sup> Self-control is measured by three latent factors for low impulsivity, perseverance, and focus. The high-deviance group has lower perseverance and higher impulsivity than the control group, but the same levels of focus as the control group.

<sup>66</sup> This discussion treats the authors' parameter estimates as standardized, although the authors did not appear to include an explicit statement of whether the parameters are standardized or not.

In a second cross-sectional study of 961 junior high school students in southern Taiwan, Chang (2002) tests for the effects of self-control on self-reported deviance<sup>67</sup> in a model that also controls for social control variables in the form of parental, school, peer, and sibling attachment, as well as for social learning variables via peer and sibling deviance. She finds that a one-standard deviation increase in self-control is associated with a decrease in self-reported delinquency of .21 standard deviations – a significant finding – but adding self-control to a model with social control and social learning variables only explains 4% more variance in delinquency (52% instead of 48%) than if self-control were left out. In contrast, following social learning theory, both sibling and peer deviance have massive, positive effects on self-reported deviance – a one-standard deviation increase in sibling delinquency and peer delinquency, respectively, is associated with an increase in self-reported deviance of 0.27 and 0.42 standard deviations, respectively, and these effects are the strongest in the model. Peer attachment and deviance explain 39% of the variation in self-reported delinquency, and sibling attachment and deviance explain 27%. Additionally, as she is conducting a cross-sectional survey, the weak findings for self-control cannot be explained by the inclusion of lagged delinquency.

Applying a multivariate latent curve model (MLCM) to four waves of the Taiwan Education Panel Survey corresponding to middle and high school, Lee (2011) conducts a joint test of social control theory, social learning, and self-control theory.<sup>68</sup> She estimates the initial value and constant rate of change of deviance with the initial values of each of her independent variables, and estimates the rate of change of deviance with the constant rates of change of each

---

<sup>67</sup> Her measure for self-reported deviance includes 21 items that encompass all the TYP deviance items as well as a few additional items such as carrying a hidden weapon, conflict with teachers, participating in gangs, trespassing, extortion, and (controversially, especially for Western researchers) having sex, and she reports an alpha reliability of 0.85.

<sup>68</sup> Chang's measures of social control include educational performance, educational aspirations, parental attachment, and school attachment; social learning is modeled with peer deviance.

of the independent variables. Net of the social learning and social control variables, the slope and intercept terms for self-control explain neither initial deviance nor the slope of deviance; only when she runs separate models for gender does she find that the intercept for self-control predicts a lower intercept for deviance for males only, with no effect of self-control on change in deviance. Further, the effect of self-control on initial deviance for males is about half the magnitude of the effect of school attachment, and one-third the effect size for social learning.<sup>69</sup> Lee's latent growth curve methodology is solid and produces convincing results, but she argues that measurement error may also have produced the null findings for self-control. In summary, these three studies indicate that self-control appears to have only mild effects on deviance in Taiwan compared to social control and social learning theory, so using lagged deviance as a proxy measure for low self-control in selection models for work and dating is likely to be adequate for the purposes of the current research.

#### *Other Relevant Background Variables*

This section covers a number of additional control variables including family intactness, parents' education, gender, and self-esteem. Lower deviance has been associated with intact families in Taiwan (Liu 2003; Wang and Jensen 2003), and in particular students living with both biological parents are less likely to resist correction from their teachers when school problem behavior leads to conflict with their teachers in the classroom (Huang 2006a), suggesting that parents may be better able to coordinate discipline with teachers when both parents are present. A child's mother's education is also associated with lower deviance in some

---

<sup>69</sup> All effect sizes are standardized.

studies (Wang and Jensen 2003),<sup>70</sup> as is the father's education (Wang and Jensen 2003)<sup>71</sup>, although Liu (2003) finds that the father's education is only associated with serious delinquency.<sup>72</sup> The father's education has also been linked to increasing levels of depression for Taiwanese middle and high school students, suggesting that children with well-educated fathers feel more keenly the pressure to pass the high school and college entrance examinations (Yi, Wu, Chang, and Chang 2009); given the strong association between depression and delinquency among Taiwanese students (Wu and Lei 2003), the father's years of education may therefore also indirectly predict increasing delinquency as children progress through school. Indeed, for TYP respondents between 16 and 18 years old, in a fixed effects framework,<sup>73</sup> fathers with more years of education have children with elevated levels of smoking, drinking, and drug use, while maternal education has no effect (Huang and Lin 2009).<sup>74</sup> Also, studies in both China and Taiwan consistently demonstrate that boys commit more crime than girls (e.g. Bao, Haas, and Pi 2007; Chang 2011; Cheung and Cheung 2010; Jessor et al. 2003; Lin et al. 2014; Lin and Yi forthcoming; Liu 2003).

Finally, using the Taiwan Youth Project (TYP) data, Tzeng and Yi (2009) find a nonlinear relationship between self-esteem and delinquency. In particular, in a random effects framework that allows them to focus on within-individual heterogeneity, they find that net of lagged

---

<sup>70</sup> Wang and Jensen (2003) also find that the mother's employment is unrelated to delinquency, while fathers who are unemployed are more likely to have delinquent children. Wang and Jensen's findings do not necessarily contradict those of Huang (2007), who finds a positive relationship between the lagged monthly male labor force participation rate and the juvenile delinquency rate in Taiwan, as long as the father's employment status is only weakly or not at all correlated with parental monitoring by the father (Wang and Jensen's posited mechanism for how the father's unemployment affects delinquency). The current research does not control for parental employment status, using household income instead.

<sup>71</sup> Wang and Jensen also control for peer delinquency in their cross-sectional study of 995 students aged 12-13 in Taipei, so these background effects are particularly robust in their research.

<sup>72</sup> For Liu (2003), serious delinquency is a factor comprised of drug use, making threats, sex, participating in street gangs, and vandalizing cars and motorcycles.

<sup>73</sup> Huang and Lin's model also controls for family income, parents' marital status, educational track, lagged deviance, and peer deviance.

<sup>74</sup> These findings suggest a different cultural context for the study of deviance than that of most Western studies, where maternal education matters more than paternal education in predicting delinquency.

delinquency, age, sex, and whether the respondent lives in Taipei city, Taipei county, or Yilan county, the relationship between self-esteem and delinquency is U-shaped, such that respondents in general commit less delinquency the higher their self-esteem, but the respondents with the highest self-esteem are slightly more deviant than those with moderate levels of self-esteem. For their dependent variable, however, they measure delinquency using an IRT model estimated using maximum likelihood, which (as presented in the previous chapter) violates the underlying assumption of unidimensionality in the latent construct for deviance for the TYP data set. Their findings may therefore be tentative, but nonetheless, controlling for self-esteem (and possibly squared self-esteem) in a selection model seems prudent.<sup>75</sup>

#### *Summary of Findings Concerning the Etiology of Juvenile Delinquency in Taiwan*

Taken as a whole, the literature on the etiology of juvenile delinquency in Taiwan and Mainland China suggests that interaction with delinquent peers constitutes a crucial, proximate cause of delinquency, and mediates the effect of family, school, and neighborhood bonds on crime. Social control, in addition to directly reducing deviance, may also lower it indirectly by strengthening moral beliefs and conventional behavior, thus protecting youth from the influence of delinquent peers. The most important types of social control in Chinese culture include the strong, assertive role of parenting and the cultural emphasis placed on education and the authority of teachers. School plays a role in social learning as well, as classroom-level effects on delinquency in Taiwan are even stronger than neighborhood-level effects. One implication of the cross-sectional nature of the studies that integrate social learning and social control variables is

---

<sup>75</sup> In practice, the models in Chapters 4 and 5 often require lagging self-esteem by more than one year as they use different survey waves than those which Tzeng and Yi use. Accordingly, self-esteem is not squared in this dissertation in predicting selection into work and marriage, as the lag removes some of the explanatory power of the variable.

that instead of peers imparting delinquent beliefs on Taiwanese youth, a young person's decision to offend may affect his or her beliefs and subsequent choice of friends (see Matsueda 1989). Concerning strain theory, some types of strain, such as traumatic life events, victimization, and depression, appear to directly affect deviance net of social control variables; they constitute important controls for a model of selection into part-time work given their potential to also influence work intensity and duration. Additionally, absent direct measures of respondents' levels of anger, additional strains such as their beliefs about the extent to which life is fair and the extent to which they are able to reach their own goals if they work for it may also affect crime (Agnew 2006b; Lin et al. 2014). Following Bao et al. (2014), the strain of work in particular could also either directly increase deviance or else do so indirectly by eroding attachment to school.

Based on the above review, Table 3.1 presents the Taiwan Youth Project variables that may affect juvenile delinquency in Taiwan and are available to model selection into work in the next chapter. Appendix E provides a detailed description of each of these measures, along with an explanation of which variables are lagged by more than one year to be included in the model. (The next chapter draws on the J1 Taiwan Youth Project cohort for its analysis, and a similar set of variables drawn from the J3 cohort surveys are used to model selection into marriage in Chapter 5 and are discussed in Appendix J.) The left column provides each variable name, and the rest of the columns correspond to waves containing work variables for which the given variable is used to model selection into work (the same waves are used to predict selection into dating). The table also provides the theory in the previous review corresponding to each variable.<sup>76</sup> A majority of these variables are related to control theory and reflect a focus on social

---

<sup>76</sup> Note that the columns in Table 3.1 indicate the waves for which work is modeled using *lagged* covariates, and the cells in Table 3.1 also contain the survey waves from which these lagged covariates are extracted.

institutions, such as the school and family, that can influence individual behavior in Taiwan. Including all these measures of social control theory in the same model would likely weaken the magnitude of the coefficients associated with measures of a theory able to significantly explain delinquency, but as the current objective is merely to capture selection into work and marriage rather than test the individual theories, this fact does not present an obstacle for the present research.

**Table 3.1.** TYP variables used to model selection into work that are based on a review of the criminology literature with respect to Taiwan. Columns representing waves denote the wave for which work is being predicted, and the actual measures are usually lagged by one or two waves (see Appendix E for more information). An “X” in a cell indicates that the indicated covariate is measured in the survey year corresponding to the column label.

<b>Variable</b>	<b>J1W4</b>	<b>J1W5</b>	<b>J1W6</b>	<b>J1W8</b>	<b>J1W9</b>	<b>Relevant Theory</b>
<b>Time-Invariant Covariates</b>						
Respondent’s gender	X					Control variable
Area (Taipei city, Taipei county, Yilan county)	X					Control variable
Respondent’s birth year	X					Control variable
Mother’s education	X					Control variable
Father’s education	X					Control variable
<b>Time-Varying Covariates</b>						
School / teacher attachment	X	X	X	X	X	Social control
School commitment: Class rank, studying hard, and time spent on homework <sup>a</sup>	X	X	X	X	X	Social control
Parental attachment	X	X	X	X	X	Social control

(Continued on next page)

**Table 3.1 (cont.)** TYP variables used to model selection into work that are based on a review of the criminology literature with respect to Taiwan. Columns representing waves denote the wave for which work is being predicted, and the actual measures are usually lagged by one or two waves (see Appendix E for more information). An “X” in a cell indicates that the indicated covariate is measured in the survey year corresponding to the column label.

<b>Variable</b>	<b>J1W4</b>	<b>J1W5</b>	<b>J1W6</b>	<b>J1W8</b>	<b>J1W9</b>	<b>Relevant Theory</b>
<b>Time-Varying Covariates</b>						
Parental monitoring	X	X	X	X	X	Social control
Live with parents	X	X	X	X	X	Control variable
Peer deviance	X	X	X	X	X	Social learning
Prior deviance	X	X	X	X	X	Self-control
Peer attachment	X			X		Social control, social learning
Stressful life events	X	X	X	X	X	Strain
Educational aspirations	X	X	X	X	X	Social control
Harsh and erratic discipline	X	X	X	X	X	Social control
Social Disorganization	X	X	X	X	X	Social disorganization
Neighborhood attachment	X	X	X	X	X	Social control, social disorganization
School involvement	X	X	X	X	X	Social control
Neighborhood involvement	X	X	X		X	Social control, social disorganization
Family involvement	X	X	X			Social control
Class cohesion	X	X	X	X	X	Social control
Family cohesion	X	X	X	X	X	Social control
Parental educational aspirations	X	X	X	X	X	Social control
Self-esteem	X	X	X	X	X	(Derived from depression research)

(Continued on next page)

**Table 3.1 (cont.)** TYP variables used to model selection into work that are based on a review of the criminology literature with respect to Taiwan. Columns representing waves denote the wave for which work is being predicted, and the actual measures are usually lagged by one or two waves (see Appendix E for more information). An “X” in a cell indicates that the indicated covariate is measured in the survey year corresponding to the column label.

<b>Variable</b>	<b>J1W4</b>	<b>J1W5</b>	<b>J1W6</b>	<b>J1W8</b>	<b>J1W9</b>	<b>Relevant Theory</b>
<b>Time-Varying Covariates</b>						
Adverse SES	X	X	X	X	X	Strain
Strain from school	X	X	X			Strain
Materialism					X	Strain

<sup>a</sup> Work in Wave 4 is predicted with a single scale for school commitment comprised of class rank, studying hard, and time spent on homework. School commitment for Wave 5 consists only of studying hard, and the variable for Wave 6 consists of class rank and time spent on homework included in the model separately. School commitment for Waves 8 and 9 consists only of time spent on homework.

<sup>b</sup> A lagged measurement from Wave 5, interpolated using responses in Waves 4 and 6.

The next chapter contains an empirical analysis of the causal effect of part-time work on deviance in Taiwan. It begins with a discussion of why Taiwanese youth work part-time, building on the review in this chapter to establish a selection model for work. The chapter then presents empirical findings, using the delinquency variables from Chapter 2 in the analysis.

## **Chapter 4: Empirical Analysis of the Effect of Work on Delinquency for Youth of High School and College Age**

The relationship between employment and crime has long been of interest to sociologists and policy makers, because work is such a ubiquitous conventional activity that has been shown to be associated with lower rates of crime for some groups. Working both demonstrates conformity to social norms and gives the worker a new set of relationships and commitments that he or she could jeopardize by breaking the law (Hirschi 1969). In keeping with the life course theories of crime introduced previously, a number of studies have focused on the effect of employment on delinquency, and have found effects related to desistance that are age-graded (Laub and Sampson 2003; Uggen 2000). In the United States for example, the National Supported Work Demonstration, the largest randomized experiment ever conducted to assess whether a program to help ex-offenders find stable jobs after leaving prison would lower delinquency, reduced the rate of recidivism for older offenders but had no effects for youth (Manpower Demonstration Research Corporation 1980). One reason for this result may be that high-quality jobs (usually available to those with more education or experience) are associated with lower recidivism after controlling for selection into these jobs (Uggen 1999), and another may be due to family bonds: ex-offenders involved in intimate relationships are more likely to be employed after incarceration, and those who are employed are less likely to re-offend (Berg and Huebner 2011). This age-graded result indicates the salience of life course theories of crime in explaining the protective effect of work, as well as the fact that turning points during the life course may operate differently for people of different ages.

In addition to working, going to school constitutes another stage of the life course that affects delinquency. Education takes place during the life course prior to work, but also overlaps with work for many people. School is particularly interesting to criminologists because it is a social institution with which most youth are frequently engaged during the peak years of offending for every population in the world, roughly ages 13 to 22 (Gottfredson and Hirschi 1990). As people of this age are usually expected to be in school, an important question for criminologists is the effect of concurrent work on steering people away from crime during a period in the life course when working is at a minimum seen as optional and may even be seen as indicative of neglecting one's studies.

The current chapter investigates the causal effect of part-time work on delinquency among youth of high school and college age in Taiwan. In focusing on Taiwanese youth, this chapter also examines a cultural group for which expectations about conforming to social norms are strong. The goal is to determine if work still has the potential to lead these youth away from crime, if it actually causes more deviance, or if it has no effect on deviance. Taiwanese sociologists have also studied the relationship between employment and delinquency in Taiwan, but have lacked longitudinal measures of juvenile delinquency and work patterns to derive causal inferences. The new longitudinal measures of deviance in northern Taiwan generated in Chapter 2 provide new information that can be used to address this question.<sup>77</sup>

The empirical research into the effect of employment on crime among youth focuses on three main hypotheses, known as the “zero-sum,” “agentic,” and “precocious maturation” hypotheses (Staff, Mortimer, and Uggen 2004). As very few studies in Taiwan have focused on the relationship between work and crime due to the aforementioned lack of data, this chapter

---

<sup>77</sup> Deriving true causal inferences from observational data is notoriously difficult for sociological research, and the inferences reached in the present chapter are of course only valid provided that the model assumptions are correct and the phenomena of interest are well measured by the data.

begins with a review of the support for these three hypotheses in Western studies, then continues with a critical examination of the reasons why Taiwanese youth choose to work part-time, linking the findings back to the hypotheses and comparing these motives to those for youth in the West. This review also covers some of the consequences of part-time work for students in Taiwan, as a person's decision to work may be based on an appraisal of returns from working or not working in previous years, as well as on observations of the people in his or her social network (a factor potentially more important in a collectivist culture such as Taiwan than in the West). The selection of other background factors in the survey that might affect selection into work is based on this literature review.

The analyses in this chapter employ three methods: fixed effects, random effects, and marginal structural models, and a brief discussion of the utility of each are in order in order to lay out the direction of the analyses in this chapter. Fixed effects and random effects models are strong methodologies for describing the relationship between variables in panel data, as they capture the between-individual variance and within-individual variance of the dependent variable separately. Random effects models capture unobserved differences between individuals as random variables and specify the probability distribution for them, while fixed effects models control for all differences between individuals and unobserved time-invariant characteristics of individual respondents by including dummy variables for respondents themselves in the regressions (or, equivalently, by partialling the information associated with such dummy variables out of the estimating equations for the model). Fixed effects models thus allow these unobserved confounders to be correlated with the regression parameters, but restrict analysis to respondents for whom change in the dependent variable is observed. Random effects models make use of stronger assumptions that the unobserved variables are uncorrelated with the

observed variables, but generally allow for more observations to be included in a model and therefore higher statistical power (Allison 2009). Causal models using inverse probability treatment weighting (IPTW) impose a different set of assumptions that are described in greater detail later in this chapter, and of allow researchers to make counterfactual inferences if the assumptions hold.

Following the literature review at the beginning of this chapter, the hypotheses tested in this chapter are presented, followed by a description of the variables related to work and selection into work in the TYP survey, including how the Chinese wording of the survey questions may influence the analyses and how the variables are prepared for analysis. To be consistent with previous research, the present chapter then presents results from fixed and random effects models of work and deviance, which provide initial findings that are relational and not counterfactual in nature. Preliminary analyses with fixed and random effects models also demonstrate a need for methods to address two types of missing value problems in the data: item nonresponse and unit nonresponse.<sup>78</sup> The chapter first presents the justification for correcting for missing values, and then addresses item nonresponse for the fixed and random effects analyses, and unit nonresponse for the causal models. With the causal effect of work, measured as a binary outcome, then determined, the chapter concludes with an analysis of how the number of hours per week and months per year that respondents work is related to their deviance, followed by a discussion.

### **Three Hypotheses for the Effect of Youth Employment on Delinquency**

Research into whether youth employment reduces, fosters, or has no effect on delinquency have to date produced conflicting results, as the particular causal mechanisms through which

---

<sup>78</sup> Item nonresponse refers to missing values for which a respondent completed the survey for a particular year but for some reason did not answer a particular question, while unit nonresponse refers to missing values for which a respondent did not complete a survey at all for a given survey year.

these effects occur relate to specific criminological theories and may be strong or weak depending on context. The lion's share of the research is connected in some manner to rational choice theory (Becker 1968), which assumes that in choosing whether to commit a crime instead of engaging in another activity such as working, people maximize personal benefit while minimizing risk and effort. Recent findings in this vein favor an "agentic" relationship between employment and crime, such that working creates positive incentives not to offend for people better attuned towards delayed gratification. A few well-designed studies, however, suggest that selection effects explain most or all of the relationship between employment and crime among youth.

*The "Zero-Sum" Hypothesis: Working Long Hours Increases Delinquency*

The "zero-sum" hypothesis, dominant in the mid-1980s and 1990s (Apel, Bushway, Brame, Haviland, Nagin, and Paternoster 2007), holds that a zero-sum game exists between study time and hours worked, such that those who work have less time to complete homework and prepare for tests, and those who study have less available spending money. According to this hypothesis, lower grades resulting from the choice to work intensively may cause youth to lose interest in both school and related activities such as organized sports, and thus set them on a path towards delinquency and drug use (Greenberger and Steinberg 1986). Indeed, intensive work has been associated with more crime (Steinberg, Fegley, and Dornbusch 1993), lower grades (Steinberg and Dornbusch 1991; Steinberg, Greenberger, Garduque, and McAuliffe 1982), and dropping out of high school among white male sophomores and white female juniors (D'Amico 1984). Intensive work also is associated with aggressive behavior (Bachman and Schulenberg 1993), alcohol use (Mortimer, Finch, Ryu, Shanahan, and Call 1996; Schulenberg and Bachman 1993),

and marijuana and tobacco use (Schulenberg and Bachman 1993). Under the zero-sum model, individuals have the free will to choose to work intensively or not, but once they make that choice, the different social organizations in which they then participate then affect their subsequent level of delinquent behavior.

The zero-sum hypothesis is consistent with several theories. The first is social control theory (Hirschi 1969). As put by Matza (1964), only when bonds to social institutions such as school or the family are weak does an individual then have the opportunity to make a rational choice to offend or not based on costs and benefits. Work could thus degrade a young person's bonds with his or school or family by reducing time for involvement in related activities. A second theory compatible with this hypothesis is strain theory. According to Agnew's General Strain Theory (Agnew 1992; Agnew 2006b), people commit crimes when sources of pressure, either chronic (such as poverty) or more immediate (such as the need for better clothes or makeup to fit in with peers) leads them to break a rule to quickly ease the strain. Work-induced strain is common in Taiwan: 55.7% of vocational students in Miaoli County cite "not enough sleep" as the chief disadvantage of working, and another 48.8% cite "frustration caused by pressure at work" (Li and Qiu 2004:36). In the U.S., policy makers have drawn on studies linked to the "zero-sum" conceptual model to devise labor laws intended to limit employment during high school (Committee on the Health and Safety Implications of Child Labor 1998). The "zero-sum" model appears to be an idea commonly accepted by Taiwanese parents who, prioritizing educational attainment for their children, often discourage them from working part-time while in school (Huang, Yi, and Yang 2008a).

*The “Agentic” Hypothesis: The Effect of Work on Crime Depends on a Person’s Intentions*

A second hypothesis concerning the effect of work on delinquency posits a more “agentic” approach to work, meaning that some people express agency by deciding how much to work based on a plan they have for their lives, and those who are “planful” are less likely to risk committing crime for short-term gain. Under this hypothesis, for those who plan ahead employment is an investment, and some people may select intensive work experience and social capital over developing human capital through education, or else treat work as a supplement to their education (Mortimer 2003; Mortimer et al. 1996). D’Amico (1984) finds that those who work fewer than 20 hours per week are more likely to complete high school than those with no job at all. Mortimer et al. (1996) similarly show that students working fewer than 20 hours per week report higher GPAs than non-workers and “intensive” workers, those who work more than 20 hours per week. The latter group also does less homework than non-workers, but their GPA is no lower. Schulenberg and Bachman (1993) suggest that job quality may explain why some work patterns are more beneficial than others, most likely in the form of “good” jobs exercising a form of social control. Like the “zero-sum” hypothesis, this theory predicts that low-intensity work results in less crime than high-intensity work, but differs from the zero-sum theory in its emphasis of the role of choice. It suggests that some people who work do so to supplement their education, and in terms of Merton’s typology of adaptation to cultural norms (1957) this group of people are in fact conformists who are using an active, culturally-accepted strategy to reach their goals, and thus do not experience very much strain at all.

The agentic hypothesis could be interpreted as one of heterogeneity of treatment effects, such that the effect of work intensity (hours per week) and extensity (number of weeks worked) on crime varies based on “planfulness” (Mortimer 2003). Planful people working a few hours per

week for many weeks tend to have high educational aspirations and diversify their time use between work and other activities, while those working intensely for many weeks may desire skills and pursue more stressful but rewarding jobs. Creating a typology of work patterns as either “occasional” (low-intensity, low-extensity), “steady” (low-intensity, high-extensity), “sporadic” (high-intensity, low-extensity), “most invested” (high-intensity, high-extensity), and “no work”, Mortimer (2003:190, 199) found that students who had either had “steady” or no work during high school were more likely than students falling under the other three categories to graduate from college four years later, and were less likely to marry early or have children by age 24-25. Looking at contemporaneous problem behavior (2003:162), she also found that high school seniors with no work or “occasional” work drank less during high school than “most invested” workers, with alcohol use for the “steady” and “sporadic” categories not statistically different from those who were “most invested.” A replication and extension of Mortimer’s work that applied an IPTW model to a longitudinal survey of 1,500 high school youth in Denver (Callahan 2010) confirmed this posited heterogeneity in treatment effects: net of selection effects, moving from high-intensity to low-intensity, long-duration work lowered general crime but not economic crime, suggesting that stable work for youth in Denver reduces crime due to non-pecuniary factors.

This hypothesis emphasizing the role of agency is also consistent with social learning theory (Akers 1998). Youth with a sense of control over their own lives may be less susceptible to peer pressure, or may choose to work in settings that happen to have a low percentage of deviant colleagues. In Taiwan, the same survey in Taiwan’s Miaoli County that suggested that youth who work feel pressure and also lose sleep (Li and Qiu 2004) also found that in keeping with the agentic model, the top four self-reported benefits of part-time work for vocational students were

“learning to perform different types of job roles” (90.4% of respondents), “broadening my perspective” (86.8% ), “learning to control my emotions” (83.3%), and “increasing my self-confidence” (79.5%). All these benefits relate to the degree that a person expresses agency.

### *Precocious Maturation: The Effect of Work on Crime is Spurious*

Finally, the relationship between intensive work during adolescence and deviant activity may be spurious. The most prominent model to explain spuriousness is “precocious maturation” (Newcomb and Bentler 1988), a condition in which some individuals, possibly because they are less able to delay gratification, engage at an early age in antisocial behavior that may include criminal activity, early sex, and especially drug use (McGee and Newcomb 1992; Newcomb 1995; Newcomb and Bentler 1988). Newcomb and Bentler’s model closely relates to Jessor and Jessor’s problem behavior theory, which holds that drug use is one part of a larger syndrome characterized by a single latent predilection for deviance (1977; 1978). Much support for a syndrome of precocious maturation comes from the substance use literature (e.g. Newcomb 1995). This hypothesis is consistent with Gottfredson and Hirschi’s General Theory of Crime (Gottfredson and Hirschi 1990). In other words, the same people who work only lightly or not at all are also able to delay gratification, while those who work intensively often desire the immediate pocket change over the long-term benefits of a strong education – and are likely to commit crime for similar reasons.

### *Precursors and Outcomes of Part-Time Work in Chinese Research Settings*

The basic reasons why youth seek part-time jobs do not seem culture-specific. Using the Youth Development Survey in the US, Mortimer (2003:115-128) finds that youth in 12<sup>th</sup> grade

who are asked about why they wish to work generally would do so in order to buy things (81.6% of respondents), save for other purposes (73.0%), save for their future education (57.7%), gain work experience (41.3%), and learn new skills (26.5%). A similar survey of Taiwanese youth aged 15-24 that only allows youth to check one category indicates that they would work primarily to buy things (55.3%), supplement their family's income (16.2%), gain "experience in society" (11.2%), learn new skills (10.8%), contribute to their family business (2.8%)<sup>79</sup>, get out of the house (1.8%), participate in an apprenticeship (1.3%), or because their friends are working (0.5%) (國勢普查處[National Status Investigatory Division] 2010). Thus, despite the wording of the questions about saving money reflecting cultural norms concerning the expression of filial piety in the Taiwan survey and individual autonomy in the American survey, the overall reasons why youth work in both surveys are similar.<sup>80</sup>

Another way to determine why Taiwanese youth choose to work is to ask those who do not work why they choose not to. When youth between 15 and 24 years old who did not look for work in the past year are asked why they chose not to, 72.2% say they are concerned that working would interfere with their schooling, 13.0% do not believe they need to work, and 8.8% believe their parents or teachers would not approve of their working (國勢普查處[National Status Investigatory Division] 2010).<sup>81</sup> Accordingly, commitment to school, family

---

<sup>79</sup> Because fewer than 3% of Taiwanese youth reported that they worked primarily to support the family business, despite the prevalence of family-owned enterprises in Taiwan the present research does not pay special attention to the mechanisms through which social control from the family may lead to the decision to work for the family's company.

<sup>80</sup> Because the U.S. survey allowed respondents to check multiple categories and the Taiwan survey only allowed respondents to check one, the best way to compare the responses is to examine the relative rankings of the categories in both surveys. In particular, given a mild assumption that Taiwanese families would use their child's contribution to the family savings to pay for the child's future education or other expenses, then the "supplement my family savings" category for Taiwanese youth would map to the "save for other purposes" and "save for my future education" categories for the U.S. survey, making the relative category rankings exactly the same.

<sup>81</sup> Additionally, 2.4% say they worry they would not be able to find a job, 1.2% say their chores take up too much time, 1.2% are too sick to work, and 0.6% are males who plan to fulfill their mandatory military service first in order to not later disrupt their job.

socioeconomic status, and the extent to which parents and teachers approve of their children or students working would seem to predict selection into work.

If materialism motivates crimes such as theft, then it could explain a positive association between work and crime. If indeed 82% of youth who wish to work desire money to buy things, yet 9% are concerned that their parents or teachers would not approve and a further 72% are concerned about their grades being adversely affected, then crime becomes a possible solution to the dilemma.<sup>82</sup> One study in Mainland China (Auerbach et al. 2009) did in fact link materialism to deviance (although the authors find that materialism was completely mediated by stressful life events), but sadly the TYP survey only contains questions about materialism in Wave 9. Absent such a set of questions, controlling for the number of stressful life events in the last year in a statistical model for selection into work could potentially mitigate part of the confounding effect of materialism, if it is indeed associated with crime in Taiwan as it is in China.<sup>83</sup>

The preceding review of causes of delinquency in Taiwan touched upon the pressures of schooling, but as the structure of Taiwan's educational system may influence when and why Taiwanese students might decide not to jeopardize their education by working, as well as for why their parents might discourage them from working, a brief review of research covering the role that educational tracking and cram school play in the lives of Taiwanese youth is also necessary. Central to how tracking influences work is the extent to which students are planful in deciding how and when to work, as well as what type of school to attend – so to frame the

---

<sup>82</sup> Merton (1938) makes exactly this case, that those who feel social pressure to attain some object or achieve some goal, yet for one reason or another perceive themselves as unable to do so through legal means, will often turn to crime in order to realize their goal.

<sup>83</sup> Note that the “stressful life events” as measured in the Taiwan Youth Project survey tend to be events that happen to the youth (such as losing a family member), while Auerbach et al. (2009) use events, such as arguing with one's parents, that imply a greater degree of agency on the part of the respondent. Controlling for “stressful life events” as phrased in the TYP questionnaires therefore may not fully mitigate the effect of materialism, and that is a limitation of the present research.

discussion, the next section begins with how part-time work affects a teenager's happiness, sense of independence, and self-efficacy.

### *Consequences of Work in Taiwan*

Using random effects models controlling for a broad array of youth and parental characteristics, Huang and Chien (2013) test whether working part-time is related to depression among TYP respondents. Their findings suggest that youth of college age are happier when they are out of school and working. They divide the SCL-90-R questions included in the Taiwan Youth Project survey into three factors – sleep disorders, physical disorders, and depression – and use each as a separate dependent variable. Work is represented as a categorical variable with four levels: working while in school, in school but not working, working but not in school, and neither working nor in school. Three dependent variables for depression and a categorical explanatory variable with four levels, and separate analyses for youth of high school age and of college age, create a great deal of detail that is not completely covered here, but the most important effects of work on well-being are for youth who are not in school compared to those who are, and are age-specific. Those aged 16-18 who are not in school and not working lose much more sleep than students who are in school and not working, and those not in school and working lose only a little more sleep than students in school and not working. Students not in school, however, experience no more or less depression or physical disorders compared to those in school and not working. For the respondents aged 19-24, compared to those in school (college for the vast majority of respondents) and not working, those who are working and not in college are much less depressed and experience fewer physical aches and pains, but lose more sleep.

Because the youth who are not in school are happier, the findings cast some doubt on whether the zero-sum theory applies in Taiwan.

Young Taiwanese who work also appear more likely to believe that they can act on their own and make their own decisions. In an analysis particularly important to the present research that speaks directly to the “agentic” model, Wu and Yeh (2013) investigate the extent to which part-time work affects individual autonomy. Recent research in Taiwan (Yeh, Liu, Huang, and Yang 2007) has shown that autonomy – namely a combination of self-awareness, emotions, and actual ability associated with the capability of governing oneself, making one’s own choices, and demonstrating the free will to do so – is a two-dimensional trait for Taiwanese youth, comprised of *individuating autonomy* and *relating autonomy*. The first emphasizes individual capacity to overcome external limitations and achieve personal goals, while the second involves attaining autonomy through achieving a common identity with others by means of maintaining interpersonal relationships and moving beyond one’s limited viewpoint (Wu and Yeh 2013:5). Yeh and Yang (2006) find that individuating autonomy is associated with lower depression and a slight increase in delinquency and aggression, while relating autonomy is associated with lower aggression and delinquency; they theorize that those with high relating autonomy have lower deviance as a result of greater control over their interactions with delinquent peers, while those with high individuating autonomy have lower depression due to a greater ability to regulate their own emotions.

Wu and Yeh (2013) also find that these two dimensions of autonomy are not mutually exclusive, and theorize that those who are strong in both dimensions use both to their advantage when appropriate. One way to understand how these two types of autonomy might be used is that in Chinese culture, individuals must locate themselves within a hierarchy and have the social

skills to acknowledge and respect authority, putting the goals of the group ahead of their own, while at the same time competing with their peers – often those in other groups, but sometimes those in the same group – to get ahead (Hsu 1948:276). Essentially they must navigate two interpersonal domains, much as those inner-city youth in the U.S. who are the most upwardly mobile are those who can successfully navigate both the “mainstream” culture of the U.S. educational system and the more aggressive “code of the street” (Anderson 1999). The implication for the “agentic” model is that the “planfulness” referred to by Mortimer (2003) is mostly individuating autonomy, and in Chinese culture another entire dimension of autonomy associated with a willingness to accept direction from those of higher status may also create salutary effects of work on deviance.

Drawing on a set of questions in the Taiwan Youth Project survey related to individuating and relating autonomy, in a structural equation framework Wu and Yeh (2013) find that among those who work when they are about 20 years old, autonomy in the workplace is associated with greater individuating autonomy two years later, while the effect on relating autonomy is insignificant.<sup>84</sup> Compared to those who do not work, those who work and express high workplace autonomy also have higher individuating autonomy after two years. Looking at the effects of parenting, the authors also find that lagged paternal care and support affects only relating autonomy, whereas lagged maternal care and support affects both individuating and relating autonomy, with an effect for relating autonomy that is much larger than both that of the

---

<sup>84</sup> In the TYP survey, workplace autonomy is operationalized by two Likert scale questions, “I cannot decide how I go about doing my work” and “I often lack control over how I make progress at work,” reverse coded and converted to a scale. Individuating autonomy is characterized by two questions: “I am able take control of the situation around me to achieve my goals” and “I always have confidence in the decisions I make,” respectively corresponding to components of autonomy associated with ability and emotions. Similarly, relating autonomy is operationalized by two questions: “When I make a decision, I am able to take into consideration my parents’, teacher’s, and personal views” and “When I take into consideration my parents’ and teacher’s views on a matter, I have more confidence in my decisions” (questions that again correspond to ability and emotions, respectively).

father's and the mother's effect on individuating autonomy.<sup>85</sup> The implications for the present research are twofold: first, given Yeh and Yang's findings (2006) that relating autonomy and not individuating autonomy is associated with lower deviance, then work might not lower deviance because it affects only individuating autonomy and not relating autonomy. Thus, if deviance takes place when a youth is uninhibited by social bonds restricting his or her behavior (Matza 1964), then work experience would not improve a person's ability to resist such pressure through proactive management of the relationship, and the additional exposure to deviant peers might even raise deviance. The second implication is methodological: the assumptions of the causal model used in this research require that all components of an individual's decision making process that affect both the choice to work and the choice to offend and are also *dynamic* must be explicitly included in an initial model for selection into work for each wave (Bjerk 2009). Accordingly, the preceding research suggests that both individuating and relating autonomy should be controlled in models for selection into work, in order to identify independent effects of work on deviance (which may or may not act through affecting individual autonomy).<sup>86</sup>

Acting independently can also have negative consequences. One prospective study of the effect of internet use on smoking and drinking using the TYP data (Chiao, Yi, and Ksobiech 2014) finds that 16-year-old youth who visit internet cafés and who search for non-school work when they do so are more likely to smoke and drink at age 20, suggesting that internet café use might be a variable affecting both selection into work and deviance, and should therefore be

---

<sup>85</sup> These effects hold after controlling for age, gender, parental education, family income, time of work (day or night), and salary. Because the questions about workplace autonomy are of course only asked of the 84% of respondents with work experience, the authors cannot model the effect of workplace autonomy on individuating autonomy directly in a structural equation model. They therefore conduct an ANOVA test and find that respondents with no work experience express the same levels of individuating autonomy as those with less than the median level of work autonomy, an amount significantly less than that in workers experiencing work autonomy greater than or equal to the median. They conclude that high levels of workplace autonomy foster individuating autonomy.

<sup>86</sup> Unfortunately autonomy is not well-measured in every TYP survey wave, but the available questions are used to their fullest extent possible.

included in a propensity score model. The selection models in this chapter therefore include measures of time spent online and of internet café use. The next section, in contrast to the preceding discussion of the role of work with respect to autonomy, discusses some constraints on the daily lives of Taiwanese youth, examining the state of the literature on how Taiwan's educational tracking and cram school systems impact work and deviance.

### *Educational Tracking in Taiwan and its Relation to Work and Deviance*

Taiwan has a multi-track educational system. This section outlines the importance of educational tracking in Taiwan in order to introduce tracking as an important control variable, but first describes the system itself as many Western readers may be unfamiliar with Taiwan's educational system. Since 1968, Taiwan's educational system has included nine grades of compulsory education (Huang, Yi, and Yang 2008b), and its educational system features general and vocational tracks for both high school and general and tertiary education. Students themselves decide which track they would like to enter, but higher scores are generally required on the high school and college entrance examinations for admission to general-track schools (Huang and Chien 2013). Youth attending general-track high schools generally focus on academic topics such as English, Chinese literature, mathematics, and the natural sciences, although they also choose a concentration within the track (such as a "natural science" concentration that emphasizes chemistry, physics, and mathematics or a "social science" concentration that emphasizes writing). The vocational track is further branched, such that students entering this track in 9<sup>th</sup> grade must decide whether to enroll in a three-year vocational high school or a five-year junior college known as *wuzhuan* (五專) that is essentially the

equivalent of senior high school and an associate's degree.<sup>87</sup> The *wuzhuan* schools were much more prevalent during Taiwan's industrialization period and in recent years have gradually been replaced by vocational high schools and institutes of science and technology (four-year colleges with a mandatory technical orientation), such that during the administration of the Taiwan Youth Project survey, only about 10% of respondents attended *wuzhuan* schools. Those who graduate from a three-year vocational high school face a subsequent decision to either complete a two-year associate's degree, or else attend a science and technology institute.<sup>88</sup> Among respondents to the Taiwan Youth Project survey, 74% of academic-track high school graduates enter four-year general colleges, while only 10% enter science and technology institutes; another 13% delay their college entrance for a year, most likely to score higher on the college entrance examination, and 2% enter the labor market. Among vocational high school students, only 4.7% enter four-year colleges, 57% go to universities of science and technology, 17% delay their education, and 22% enter the labor market (Huang, Yi, and Yang 2008b).<sup>89</sup>

---

<sup>87</sup> Even at an *wuzhuan* institute, a cultural separation between high school and university life remains, such that after completing 12<sup>th</sup> grade, students at the same institute often wear different uniforms and follow an academic calendar more in keeping with other universities in Taiwan than with the slightly longer high school calendar year.

<sup>88</sup> Occasionally vocational high school students may also enter a general-track college, but achieving high scores on the college entrance examinations is more difficult for students on this track because in vocational high schools they take specialized courses that are analogous to professional certificates in the West and which cover skills not on the general college entrance examination. Getting the high scores therefore involves learning extra material such as history and geography not covered in the general vocational school curricula, and also involves reaching a higher level of proficiency in subjects such as English than required at most vocational schools. It is also difficult for students to go from general-track high schools to science and technology institutes, as those who wish to test into science and technology institutes have to learn a particular skillset and take an examination specific to their desired major in addition to demonstrating a strong grasp of Chinese, English, and mathematics. This demand for picking up additional knowledge outside traditional schools is one factor that drives the prevalence of extracurricular "cram schools" in Taiwan.

<sup>89</sup> Recent years have seen a trend in Taiwan towards integrating general-track and vocational high school programs into one building or campus, such that students can achieve more of a customized education – general-track students can apply to specific schools to pick up the equivalent of hard-to-find professional certifications while still in high school, while vocational-track students can find it simpler to prepare for the general college entrance examination without having to pay a lot to attend private cram schools in the afternoon. Taiwan's educational reforms have been quite rapid, and these integrated high schools were still a rarity at the time that TYP respondents were in high school.

Perhaps because graduating from a general high school is generally considered more prestigious than graduating from vocational high school, the demand for seats in general high schools, both among parents and their children, exceeds the supply of available spots in high schools (Huang and Wu 2010). Additionally, vocational high schools in Taiwan are more expensive to attend and provide fewer services than general-track high schools (Liao and Liu 2004). This demand creates a great deal of pressure for children taking the high school entrance examination. The disparity between supply and demand was particularly strong around the mid-1990s, when positions in high schools were allotted for only about 34% of Taiwanese youth (1998), while 61% of Taiwanese youth desired to enter high school, and a whopping 66% of parents wanted their children to attend general-track high schools. By 2006, the number of high schools had increased such that positions were available for 52% of Taiwanese youth, but the pressure still remains (Huang and Wu 2010).

Pressure to pass the high school and college entrance examinations is evident among the Taiwan Youth Project survey respondents. Depression peaks during 9<sup>th</sup> grade and again in 12<sup>th</sup> grade, right as students are preparing to take these tests (Yi, Wu, Chang, and Chang 2009). In fact, the authors find that depression is higher during middle school than in high school, and highest during 9<sup>th</sup> grade, suggesting that the competition over limited seats in good high schools is even more severe than the competition to enter college. Depression, moreover, is linked to tracking: Chang and Yi (2013) conduct a group-based trajectory analysis of depression among TYP respondents using Daniel Nagin's methodology (2005), and find that among respondents who tend to have consistently low depression, an increase in depression is associated with attending a general-track school as opposed to attending a vocational-track school, joining the work force, or otherwise delaying education by a year. That said, respondents most likely to map

onto a trajectory of individuals with moderate but consistent levels of depression (26.5% of the population) are slightly more likely to be depressed if they are in the workplace as opposed to in an academic-track school, suggesting that the effect of tracking on depression depends on one's history of depression.<sup>90</sup> A follow-up study that brings family and school cohesion into the model, examines trajectories of youth depression as a piecewise linear growth model, and models depressive trajectories for different tracks net of a variety of controls (Yi, Fan, and Chang 2013), finds that tracking does not begin to affect depression until age 17, and while students who graduate from general-track high schools are more depressed at age 18 than their vocational-track counterparts, after age 20 the college students (both general and vocational track) are less depressed than youth who do not attend college. The implication is that as far as depression goes, sticking through college is more rewarding in the long run.

Given the tremendous pressure that the hyper-competitive nature of educational tracking and the examination system creates for children, perhaps it is no wonder that cram schools are so popular. Cram schools are after-school classes taught by private companies that review material taught in schools and, most importantly, help to prepare students for examinations. For each year that respondents in the J1 cohort for the TYP study were in junior high school, more than half of the cohort attended cram school (Chang 2013). Huang, Yi, and Yang (2008b) find that respondents who live in Taipei City as opposed to outside the city limits are also more likely to enter an academic-track high school at grade 10, a result which may be related to the fact that residents of urban areas have more access to cram schools (Chang 2013). Cram schools also contribute to widening economic inequality, as families with higher incomes are more likely to

---

<sup>90</sup> The authors also identify four other depression trajectories but inferences concerning those trajectories are less certain, as none comprise more than 4% of the population, the authors made a rather subjective decision when selecting the number of groups, and they did not conduct sensitivity checks to ensure that their findings regarding the smaller groups were robust to the number of groups they used in their models.

send their children to cram schools (Pan and Yi 2011), and students attending cram school are more likely to place into general-track high schools than either students who do not participate in after-school study programs or students who participate in free after-school study programs sponsored by their junior high schools (Chang 2013:149). Chang's methodology that predicts placement into high school using multiple logistic regression, however, does not correct for the endogeneity associated with the fact that students may choose to either attend cram school or not based on their expectations about future track placement. Xie and Zhou (2013) go a step further, using a marginal structural model with inverse probability of treatment weighting (IPTW) methods to show that for TYP respondents, attending cram school during 9<sup>th</sup> grade (the third year of junior high school) causes students to score higher on the high school entrance examination and place into academic-track schools, and attending cram school during 11<sup>th</sup> grade causes students to score higher on the college entrance examination and get accepted into public, four-year (academic-track) colleges.

A young person in Taiwan's likelihood of working part-time during high school varies by educational track. Vocational high school students are more likely to work during every grade and every summer and winter break than general high school students. That said, students in both tracks are more likely to work during grades 10 and 11 than during their senior year, and also more likely to work during summer or winter breaks than during the academic year, suggesting that concern for the trade-off between grades and work hours is shared by students from both tracks (Huang, Yi, and Yang 2008b).

Using the TYP data, Huang, Yi, and Yang find both general and track-specific effects of part-time work on 12<sup>th</sup>-grade college entrance examination scores and on the ultimate decision to enter a four-year academic institution, institute for science and technology, delay their education

for a year, or enter the labor force.<sup>91</sup> Students in both tracks who work during high school are more likely to enter the labor force after graduation and less likely to attend a four-year general-track college, potentially bearing out the concerns of young Taiwanese that working could impact their academic performance, and likely also justifying some parents' active discouragement of their children from working (國勢普查處[National Status Investigatory Division] 2010). But the whole picture is not that simple, as some of the effects of work are track-specific: For students in vocational high schools, consistent with the “zero-sum” hypothesis, working is associated with lower college entrance scores, and the number of years worked is associated with a lower propensity to enter an institute of science and technology and a higher propensity to delay one's education. The implication is that both the job and the choice of a science and technology institute could be related to a young person's specific plan for his or her life (in accordance with the agentic hypothesis). For students in academic-track high schools, however, working does not directly affect college entrance scores, and years of part-time work are associated with a *higher* probability of entering an institute of science and technology and no difference in the odds of delaying one's education. Still, perhaps examining the number of years worked ignores the fact that only working during breaks may not negatively impact academic performance, so Huang, Yi, and Yang (2008b) run additional models that estimate effects for work during the semester for each year and during the summer or winter breaks for each year. Their results show that working during the academic year for 12<sup>th</sup> grade is also associated with a greater propensity for both general and vocational high school students to enter the labor market

---

<sup>91</sup> To reach their conclusions concerning track-specific effects of working on test scores, the authors use endogenous switching regression models to adjust for the fact that students may choose a track in 9<sup>th</sup> grade based on expectations about their college entrance examination scores. In contrast, they use multinomial logit regressions to model whether respondents attend a general-track or vocational college, enter the workplace, or delay their education after their senior year in high school. To simplify their analysis and because students in *wuzhuan* schools only constitute about 10% of the sample, the authors do not examine students attending *wuzhuan* schools.

after high school graduation. But most of the remaining findings of note are track-specific. For students in academic-track high schools, even working during holidays during the 11<sup>th</sup>-grade winter break and the summer break following 11<sup>th</sup> grade is associated with a lower probability of attending a general-track university. But for students in vocational high schools, there is no relationship between holiday work at any time during vocational high school and the probability of entering general university, while part-time work during the academic year for grade 12 is associated with a lower probability of entering an institute for science and technology and a higher probability of delaying education for vocational high school students compared to general-track high school students (for whom both relationships are spurious).<sup>92</sup>

Huang, Yi, and Yang's findings are consistent with both a zero-sum theory of work and a theory of agentic behavior; students who work as college entrance examinations approach may already be planning to enter the labor force, or their competing priorities may drive down their test scores and cause them to choose to work full time instead of going to college. One way to gain some purchase on which of these theories has greater explanatory power may be to examine the authors' model for the decision to delay one's education rather than to enter a particular college or the workforce, given that delay implies an inability to achieve one's goals and hence lack of agency; the fact that work during the academic year for any grade in high school has no

---

<sup>92</sup> Lest Huang, Yi, and Yang's findings be explained away as due to differing family backgrounds, the authors control for gender, father's years of education, family income, family intactness at grade 12, and whether the respondents lived in Taipei city, Taipei county, or Yilan county at grade 9. Family background variables also interact with track in predicting college entrance: for general high school students, fathers with more years of education have children who are more likely to enter four-year academic universities, while family intactness has no effect; for vocational high school students, family intactness predicts a greater likelihood of entering a four-year college, while father's education does not matter. Interestingly, father's education predicts a lower propensity to enter institutes of science and technology for general high school students and a higher propensity to enter a science and technology institute for college students, suggesting a possible link between the father's education and purposive, planful behavior. Note that the authors control for neither the mother's education nor either parent's degree of care and support offered to the child as was done by Wu and Yeh (2013), so father's education and family intactness may be obliquely capturing the effect of parents' care and support on individuating and relating autonomy, which in turn affects both the probability that a child deliberately enters into and stays on a particular academic track, and that he or she is ultimately successful.

effect on the probability of delaying education for general high school students, but work during 12<sup>th</sup> grade is associated with a greater probability of delaying education for vocational high school students, may imply that general high school students are more proactive than their counterparts in vocational high school.<sup>93</sup>

Tracking also can affect delinquency by influencing attachment and commitment to school, and through reinforcing economic inequality. Yogan (2000) argues that students in lower tracks form weaker social bonds with their schools through a symbolic interaction process. Even though Yogan writes about the U.S. educational system, this impact makes sense for Taiwanese students in vocational schools who may perceive themselves as less capable due to having not made the cut for the general track and who therefore may not be satisfied with their experience in high school – especially in light of the cultural emphasis on the accumulation of prestige in Chinese culture in order to honor one’s parents and ancestors (Hsu 1948). Tracking also, however, has been linked in the West to differences in socioeconomic status, in that students from families with wealthier backgrounds are more likely to be placed into academic tracks; among other things, these students are then more likely to go to college, have higher self-esteem, have more intrinsic motivation to study, are more involved in school conventional activities, and have friends who also like school (Vanfossen, Jones, and Spade 1987). This effect of tracking on widening inequality has also been demonstrated to take place in Taiwan (Tsai 2004).

Lin and Yi (forthcoming), using structural equation models, investigate the effects of high school tracking on delinquency for the J1 cohort of the Taiwan Youth Project survey, using as their dependent variable logged sums of the four delinquency items that did not change between Waves 4 and 6 (10<sup>th</sup> and 12<sup>th</sup> grade). They find that students in the vocational track as

---

<sup>93</sup> Accordingly, fixed-effects methods that control for unobserved time-invariant within-individual confounding variables such as self-control would yield further insight into how part-time work affects track placement for college.

sophomores in high school are more delinquent as both sophomores and seniors; they also have more delinquent peers and are less attached to their schools two years later, even after controlling for selection into track. Most notably, tracking has no direct effect on deviance during the senior year of high school net of indirect effects through Wave 4 deviance, Wave 6 school attachment, and Wave 6 delinquent peers. Importantly, this study mirrors others mentioned in this review that suggest that peer deviance and school attachment are proximate determinants of crime among Taiwanese youth. Lin and Yi are limited in their ability to examine change in delinquency over time, however, by the changing nature of the delinquency items in the Taiwan Youth Survey, addressed in the first chapter of this dissertation. Using a broader range of years, the present study treats tracking as time-varying – potentially increasing its contemporaneous explanatory power – and includes it as a control variable in the selection models for work and marriage.

In summary, tracking is related to work and to deviance in Taiwan, and cram school has strong effects on tracking. Less research is available concerning the effects of cram school on work, although a person could infer that students attending cram school would have less time to work. Additionally, little is known about the independent social control effects of cram school on delinquency, net of school attachment, commitment, and involvement. How cram school affects deviance is also unclear. The present review did not uncover any research on effects of time use on deviance in Taiwan, aside from a macro-level study that found no causal effect of unemployment on crime or of crime on unemployment in Taiwan (Chen 2009). Accordingly, out of an effort to control for as many confounders for work and deviance as possible, the selection models in this chapter predict work with cram school attendance, just in case cram school does

affect delinquency net of school and tracking effects, possibly by reducing students' opportunity to commit crime.<sup>94</sup>

In summary, cram school attendance and educational tracking both may confound the relationship between youth employment and crime. The mechanisms through which tracking and cram school affect delinquency may be related to social control, social learning, or strain, but to the author's knowledge the characteristics of these mechanisms are still open research questions. Similar in form to Table 3.1, Table 4.1 presents the Taiwan Youth Project variables that may affect selection into work in Taiwan and are based on the above review of the causes and consequences of youth employment in Taiwan, along with the discussion of the educational tracking system and cram school. Appendix E provides a detailed description of each of these measures, along with an explanation of how the variables are lagged or interpolated.

### **Data and Hypotheses**

This section presents the hypotheses tested in the current chapter and the nature of the Taiwan Youth Project survey items related to work, with a more detailed presentation of the survey items in Appendix E. The current chapter makes use of the J1 cohort responses for the TYP for two reasons. First, other researchers working with the TYP work variables have tended to favor the use of this cohort, so the present research findings can be compared to previous research in a more meaningful way. Second, the Cronbach's alpha reliabilities for the dependent variable for delinquency are slightly higher in Table A6, and even though the discussion in Chapter 2 indicates that alpha reliabilities constitute a lower bound on reliability estimates that are likely to be substantially lower than the actual reliability in the case of a nonlinear measure of

---

<sup>94</sup> Indeed, the selection models demonstrate that those in cram school are less likely to work during the next year; this result, while not a focus of the present analysis, may constitute a modest contribution to the literature on work in Taiwan.

general delinquency constructed from a set of heterogeneous survey items, all else being equal it is still a good idea to work with a lower bound for reliability that is higher.

**Table 4.1.** TYP variables used to model selection into work that are based on a review of the causes and consequences of youth employment in Taiwan. Columns representing waves denote the wave for which work is being predicted, and the actual measures are usually lagged by one or two waves (see Table 4.3 and Appendix E for more information).

Variable	J1W4	J1W5	J1W6	J1W8	J1W9
<b>Time-Invariant Covariates</b>					
Respondent's gender	X				
Area (Taipei city, Taipei county, Yilan county)	X				
Respondent's birth year	X				
Mother's education	X				
Father's education	X				
<b>Time-Varying Covariates</b>					
Household income	X	X	X	X	X
Tracking (contemporaneous)	X	X	X	X	X
Self-esteem	X	X	X <sup>a</sup>	X	X
Autonomy	X	X	X	X	X
Allowance	X	X	X		
If respondent has student loans				X	X
Cram school	X	X	X	X	
Depression	X	X	X	X	X
How often online	X	X	X	X	X
Internet café use		X	X		

<sup>a</sup> A lagged measurement from Wave 5, interpolated using responses in Waves 4 and 6.

This chapter tests the following hypotheses, all of which make use of binary indicator variables except for the last, which uses a categorical variable with four levels:

*H4.1.* Whether a young Taiwanese person works is associated with increased delinquency.

*H4.2.* Working during the summer and winter breaks is associated with less delinquency than working during the academic year.

*H4.3.* Hypothesis 4.1 notwithstanding, part-time work does not actually cause increased delinquency, but rather the relationship between the two variables is spurious in a causal framework.

*H4.4.* Heterogeneity of treatment effects: for young people experiencing low social control (as operationalized by parental monitoring only in order to include students not in school in the model), part-time work lowers deviance, while it has no effect on deviance for those with high social control.

*H4.5.* Following Mortimer (2003), low-intensity, high-duration work in college is associated with less deviance than no work, high-intensity work, or low-intensity, low-duration work. (Due to limitations of the available data, this hypothesis can only be tested for students of college age.)

In the West, recent findings from longitudinal surveys have favored the “agentic” hypothesis, which suggests that students self-select into high-quality or low-quality jobs and jobs with varying numbers of work hours, and the “good” jobs then steer them away from crime. Hypotheses 4.1 and 4.2 follow from the fact that so many Taiwanese parents believe that any work is deleterious for youth whose time would be better spent preparing for placement exams. In such an environment, a “zero-sum” relationship between working and success in school may be a *fait accompli*: if only the students with weak bonds to their schools and parents (i.e. those not persuaded by their parents and teachers not to work) decide to get a job, then a positive relationship between work and deviance could exist due to selection effects alone. Hypothesis 4.3 states that after controlling in a selection model for the confounders mentioned in the literature review in this chapter and the previous one, work would have no effect on crime. Hypothesis 4.4 then posits that work may provide a form of social control, steering people with low social control away from crime. One reason for this could be that the job buffers students from the influence of deviant peers, but the present research does not test that specific

mechanism. Finally, Hypothesis 4.5 states that Mortimer's findings (2003) concerning the protective effect of low-intensity, high-duration work hold in Taiwan as well as in the United States.<sup>95</sup>

Hypothesis 4.5 seeks to replicate research in the West that steady work during the academic year is associated with the most beneficial outcomes for youth (Mortimer 2003) and, in particular, that it lowers crime (Callahan 2010). Such research explicitly models work intensity and extensity during the academic year, so this chapter uses a data filter, described in Appendix D, that fills in gaps in the available information about work intensity and extensity for Wave 8. The appendix goes into greater detail concerning issues such as inferring when the same respondent in Waves 8 and 9 refers to the same Wave 8 job, and resolving starting and ending dates in such cases. Another issue related to the measurement of work in the TYP survey involves the Chinese phrasing of the word for "work" and its interpretation by respondents, and is discussed next.

#### *Description and Coding of the TYP Variables Concerning Work*

When youth are asked about whether they worked in Chinese-language surveys, the exact phrasing of the word "work" is often more varied than on surveys in English, and the different phrasings carry with them slightly different cultural meanings. Respondents to the TYP survey are asked if in the last year they worked full-time (*quanzhi gongzuo*, 全職工作) or part-time,

---

<sup>95</sup> A related hypothesis to Hypothesis 4.4 would hold that in a causal framework part-time work lowers crime for those of college age (19-22) but not for those of high school age (16-18), because the job is more important to youth of college age, and hence the social control-related forces influencing behavior are stronger. Unfortunately that hypothesis cannot be tested with the present data set, due to an artifact of the vertical scaling process. The problem has to do with the fact that a hypothetical person reporting zero deviance in more than one wave who experiences no change in criminality between these two waves would still be assigned ability scores that are slightly different after the scores are placed on the same scale, simply because the minimum difficulty of the survey items changes from wave to wave. Correcting for this problem requires the inclusion of dummy variables for wave in the regression models, which are collinear with dummy variables corresponding to high school and college years and thus prevent the testing of a hypothesis comparing effects for high school and college.

using one or both of two Chinese expressions for “part-time” work, *dagong* (打工) or *jianzhi gongzuo* (兼職工作). These two expressions capture slightly different meanings for work: *jianzhi gongzuo* generally refers to a formal job that is related to one’s future career plans in some way, such as clerical work or working as an apprentice mechanic. *Dagong*, in contrast, refers to a more temporary part-time job, such as a cashier or a vendor; one example for a high school student might be selling holiday specialty foods to passers-by in a busy Taipei night market during the two-week Lunar New Year holiday. In practice, a great deal of overlap exists between these two terms. Someone working as a restaurant waiter might label the job with either term, and in completing a survey his or her interpretation of the questions might depend on other factors such as whether he or she is in school at the time (and thus more likely to view the position as a temporary situation, and hence use the term *dagong*), or whether he or she is contributing to the family business (and hence more likely to use *jianzhi gongzuo*). The survey design makes a few assumptions about how youth might view their jobs, potentially introducing measurement error. Students in early waves, including high school, are generally asked if they worked during the previous academic year or summer and winter breaks using the phrasing *dagong*; the term *gongzuo* is generally used when youth not in school are asked if they worked.

Table 4.2 lists the available information related to part-time work for the TYP J1 cohort. Several features of the survey design for this cohort impact the analysis. First, in contrast to Western surveys, the TYP survey includes questions about part-time work during winter vacation as well as summer vacation. Taiwanese students in high school have a two-to-three-week holiday for Chinese New Year, and college students usually have a month off from school after end-of-semester examinations in January. The break is thus long enough that many students

**Table 4.2.** Distribution of work variables during J1 survey waves for high school and college.

<b>Variable</b>	<b>J1W3</b>	<b>J1W4</b>	<b>J1W5</b>	<b>J1W6</b>	<b>J1W7</b>	<b>J1W8</b>	<b>J1W9</b>
Whether worked during previous year	X	X	X*	X	X	X	X
Whether worked during previous summer	X	X		X**	X***	X***	X***
Whether worked during previous winter break			X	X**			
Whether worked during previous summer or winter break			X	X**	X†	X†	X†
Whether worked during previous academic year	X	X	X*	X**	X	X	X
Number of hours worked / week during summer	X			X**	X	X	X
Number of hours worked / week during winter break				X**	X	X	X
Number of hours worked / week during academic year	X			X**	X	X	X
Number of weeks worked during summer				X**	X	X	X
Number of weeks worked during summer or winter breaks				X**	X	X	X
Number of weeks worked during academic year				X**	X	X	X
Total number of weeks worked during year				X	X	X	X

\* Respondents to J1W6 were only asked if they currently had a full-time or part-time (*jianzhi gongzuo*) job, so as the survey was conducted in June, this response is used here as a proxy for having worked during the previous year and for having worked during the academic year.

\*\* Obtaining these variables would have involved writing an additional data filter, and as delinquency variables were not available for this wave, the current research does not filter these work variables.

\*\*\* These responses may include a handful of respondents who had jobs that started and ended within the winter break and who did not work during the previous summer (see the discussion in Appendix D). A total of 42 respondents had jobs that started and ended within the winter break, but as many of these respondents likely worked during the previous summer, this number constitutes an upper bound and the number of actual respondents that might be treated as working during the previous summer when they did not due to this coding decision is likely about a dozen or so.

† This variable was coded in two different ways; see Appendix D for more information.

work part-time during that time period, and most of the survey years therefore contain separate questions about work during the summer and winter vacations.<sup>96</sup>

Second, while other survey waves generally contained questions about whether respondents had worked during the previous year, the J1W6 work questions used a different point of reference, asking respondents if they currently had a full-time or part-time (*jianzhi gongzuo*) job (regardless of whether they were currently in school). The questionnaire was administered in June of most respondents' senior year in high school, so in order to use the work variables from this wave, the present research assumes that respondents to the survey who state that they were working at the time of the survey also were working at some point during the preceding academic year. An assumption about measurement is also necessary for the present analysis, namely that respondents who had a job that they considered an informal part-time (*dagong*) job, when not given the option to use that wording to refer to their status, indicated that they had a *jianzhi gongzuo* (part-time) job instead. That assumption is not without its problems – especially since the amount of such measurement error cannot be easily measured<sup>97</sup> – but it is unavoidable.

Third, determining the number of weeks worked and hours worked per week during the full year, academic year, and summer and winter breaks for Waves 8 and 9 involved writing a complex data filter that also introduces measurement error and involves a few assumptions. For these survey waves, respondents were generally asked to describe several jobs that they had held,

---

<sup>96</sup> A few additional features of the data affect measurement error. First, for Waves 4 and 5, survey respondents were only asked if they had a part-time job (phrased as *dagong*) if they were one of the 96% of Wave 4 respondents or 96% of Wave 5 respondents who were in school, and only respondents in school were asked if they had jobs during the summer or winter break. Respondents not in school were asked if they formally worked full-time (*quanzhi gongzuo*) or part-time (*jianzhi gongzuo*). Measurement error might therefore cause a few respondents not in school to erroneously be coded as not having worked if they considered their part-time job to count as *dagong* but not quite as formal as *jianzhi gongzuo*. Given the small percentage of respondents who were not in school and the fact that most youth appear to refer to *dagong* as temporary work usually associated with schooling, the amount of such measurement error is likely to be limited.

<sup>97</sup> To answer the counterfactual question of whether a respondent who answered that they did not have a part-time job (*jianzhi gongzuo*) would have stated that they did have a job had they been asked using the *dagong* phrasing of “part-time work,” the same person would need to be interviewed twice using different survey instruments, and that never happened.

and for each one to state how many hours they worked per week, the month and year that they started the job, and (if not still working there) the month and year that they left. Waves 8 and 9 differ, however, in terms of the specific jobs that respondents who had ever held more than one job were asked to recall. Such respondents to the Wave 8 survey, conducted in June 2007, were asked during the main body of the survey to describe their “main” job, based on either the length of time at their job or the total amount of money they earned at their discretion. Then at the end of the survey, respondents who had ever worked at more than one job were asked to provide additional information about the first job they had ever had.<sup>98</sup> In many cases, respondents described separate jobs when responding to both questions, such that a position that began halfway through the previous year was listed as the “main” job, and the “first” job began and ended prior to the year covered by the Wave 8 survey (i.e. the job ended prior to July 2006). Such circumstances opened up the possibility of a systematic underestimation of work intensity and extensity because the respondent may have had a third job during part of Wave 8, but that information is missing because the survey did not ask about enough separate jobs in Wave 8. Possibly for this reason, the Wave 9 survey, conducted two years later, includes questions asking respondents to describe up to three jobs that they had held during the previous three years. Of course, some respondents may have held more than three jobs by the time of the Wave 9 survey, so the instructions associated with the first job asked respondents to state the job that they either had the longest or at which they earned the most, and respondents who worked at more than one job then filled in a table to describe characteristics of up to two additional jobs, placing priority on “1) the job you currently have or 2) the job you have held the longest.”

---

<sup>98</sup> A priming effect may have caused respondents to be more likely to list their current jobs as their main job, even if the current job was not the one they had held the most or at which they had earned the most money. The effect is due to the fact that in both Waves 8 and 9, respondents first were asked to state if they had a full-time job, a part-time job (worded as “either *dagong* or *jianzhi gongzuo*,” or had no current job but previously had a full-time or either type of part-time job).

### Description and Coding of Covariates Affecting Selection into Work

The Taiwan Youth Project survey contains a rich set of background covariates, but only a few questionnaire items are consistently included in every survey year for several reasons. First of all, the survey contains fewer questions during years in which respondents complete telephone surveys instead of written questionnaires. Additionally, some survey items, such as attachment to one's neighborhood or cohort of classmates in high school, become harder and perhaps less important to measure once respondents enter college and many move away from home. Third, quite often the survey questions measuring a particular concept do so more reliably in later survey waves, suggesting that the questionnaire design was treated as a work in progress while the research team took the wording of survey items tested first in the West and adjusted them for a Chinese setting. Moreover, some constructs, such as a two-dimensional measure for autonomy in Chinese culture (Yeh and Yang 2006), had not even been invented at the beginning of the survey. Finally, some questions are only asked of a subset of respondents in different survey waves, limiting their comparability across years.<sup>99</sup>

Table 4.3 lists the covariates, discussed in Chapter 3 and the preceding literature review in this chapter that the present research uses to model selection into work for the counterfactual models in this chapter. It is a composite of Tables 3.1 and 4.1, except that the cells indicate the wave from which the questions used to model selection into work for a particular wave are drawn. The columns in Table 4.3 correspond to the survey wave for which part-time work is being predicted, the rows are the covariates, and the cells indicate the particular wave used to obtain the covariates. In general the variables are nearly always lagged to strengthen inferences

---

<sup>99</sup> Information about how much money respondents receive as part of a regular allowance, for example, is obtained for all respondents in Wave 4 and only for respondents currently in school in Wave 6, and as a result allowance can only be studied as a time-varying covariate for respondents in school.

that they indeed are predictors rather than consequences of part-time work, with tracking as one exception because its effect on the decision to work is assumed to be contemporaneous.<sup>100</sup> Blank cells indicate that a covariate is not used to predict work in a particular wave. As the first five rows denote covariates treated as time-invariant, for these variables missing values in Wave 1 are replaced with responses measured in the earliest available later wave. The covariates used for the selection model corresponding to each column usually predate the year being modeled by a single wave, but a longer lag is sometimes necessary when a construct does not appear in every survey. Following Xie and Zhou (2013), in this chapter household income for waves in which it is not observed is calculated through linear interpolation using adjacent survey waves (or using waves that are further apart if necessary due to missingness).<sup>101</sup> In general, however, this approach is avoided for covariates that are more likely to themselves be influenced by part-time work. If work potentially raises self-esteem, for example, then interpolating self-esteem in Wave 5 using self-esteem in Waves 4 and 6 could mean that the interpolated self-esteem value becomes influenced by part-time work in Wave 6 for some respondents, thereby jeopardizing the validity of the causal inferences that can be drawn from the data. Therefore, in most cases covariates not contained in a survey for a given wave are lagged by one or two additional survey waves, no doubt attenuating their effects on work and deviance but still ensuring that the predictors of part-time work temporally precede the work variables they are predicting. The particular survey items used to measure each construct in Table 4.3, their wording and reliability, and additional coding details that someone seeking to replicate the present research would need to know are presented in Appendix E.

---

<sup>100</sup> In other words, this research assumes that most students already know what their school track will be when deciding whether to work during a given academic year or during the summer or winter breaks for that academic year.

<sup>101</sup> The linear interpolation procedure used here takes into account the fact that the survey waves are not equally spaced.

**Table 4.3.** Covariates used for the selection models into work, for each survey wave used for the work analysis.

<b>Variable</b>	<b>J1W4</b>	<b>J1W5</b>	<b>J1W6</b>	<b>J1W8</b>	<b>J1W9</b>
<b>Time-Invariant Covariates</b>					
Respondent's gender	J1W1	J1W1	J1W1	J1W1	J1W1
Area (Taipei city, Taipei county, Yilan county)	J1W1	J1W1	J1W1	J1W1	J1W1
Respondent's birth year	J1W1	J1W1	J1W1	J1W1	J1W1
Mother's education	J1W1	J1W1	J1W1	J1W1	J1W1
Father's education	J1W1	J1W1	J1W1	J1W1	J1W1
<b>Time-Varying Covariates</b>					
Household income	J1W3	J1W4, interpolated from J1W3 and J1W6	J1W5, interpolated from J1W3 and J1W6	J1W7, interpolated from J1W6 and J1W8	J1W8
Tracking (contemporaneous)	J1W4	J1W5	J1W6	J1W8	J1W9
School / teacher attachment	J1W3	J1W4	J1W5	J1W7	J1W8
School commitment: Class rank	J1W3		J1W5		
School commitment: time spent on homework	J1W3			J1W7	J1W8
School commitment: studying hard		J1W4			
Parental attachment	J1W3	J1W4	J1W5	J1W7	J1W8
Parental monitoring	J1W3	J1W4	J1W5	J1W6	J1W8
Live with parents	J1W3	J1W4	J1W5	J1W7	J1W8
Peer deviance	J1W3	J1W4	J1W5	J1W6	J1W8
Prior deviance	J1W3 <sup>a</sup>	J1W4	J1W5	J1W6	J1W8
Peer attachment	J1W3			J1W6	

(Continued on next page)

**Table 4.3 (cont.)** Covariates used for the selection models into work, for each survey wave used for the work analysis.

<b>Variable</b>	<b>J1W4</b>	<b>J1W5</b>	<b>J1W6</b>	<b>J1W8</b>	<b>J1W9</b>
<b>Time-Varying Covariates</b>					
Negative life events	J1W3	J1W4	J1W5	J1W6	J1W8
Self-esteem	J1W3	J1W4	J1W4	J1W6	J1W8
Educational aspirations	J1W3	J1W3	J1W3	J1W6	J1W8
Autonomy	J1W3	J1W3	J1W3	J1W7	J1W7
Harsh and erratic discipline <sup>b</sup>	J1W3	J1W4	J1W4	J1W6	J1W6
Allowance	J1W3	J1W4	J1W4 <sup>c</sup>		
If respondent has student loans				J1W7	J1W8
Cram school	J1W3	J1W4	J1W5	J1W7	
Social disorganization	J1W3	J1W3	J1W6	J1W6	J1W6
Neighborhood attachment	J1W3	J1W3	J1W3	J1W6	J1W8
School involvement	J1W3	J1W3	J1W3	J1W7	J1W8
Neighborhood involvement	J1W2	J1W2	J1W2		J1W8
Family involvement	J1W3	J1W4	J1W5		
Class cohesion	J1W3	J1W3	J1W3	J1W7	J1W8
Family cohesion	J1W3	J1W3	J1W3	J1W6	J1W6
Parental educational aspirations	J1W3	J1W4	J1W5	J1W6	J1W8
Adverse SES	J1W3	J1W3	J1W5	J1W6	J1W6
Strain from school	J1W3	J1W4	J1W5		
Depression	J1W3	J1W4	J1W5	J1W7	J1W8
How often online	J1W3	J1W4	J1W5	J1W7	J1W8
Internet café use		J1W4	J1W5		
Materialism					J1W8

<sup>a</sup>A raw scale is used for J1W3 delinquency as a predictor, rather than an IRT measure for latent deviance.

<sup>b</sup>When both measures are available, harsh and erratic discipline are combined to create a single scale.

<sup>c</sup>An attempt was made to interpolate allowance in Wave 5 using allowance in Waves 4 and 6, but partly because allowance in Wave 6 is observed only for respondents who were in school at the time, the item nonresponse rate for the resulting imputed variable was too high to justify using the interpolated measure.

For the propensity score models to calculate treatment weights, uniformity of wording and scales across survey waves makes the models easy to understand, but the main objective for the selection model is to control for confounding to the greatest extent possible. In the case of fixed effects models with time-varying covariates, however, the time-varying covariates must be on the same scale for each wave because they are directly included in the regressions. Among all the time-varying covariates in Table 4.3, the only variables that are measured in the same way across survey waves are lagged parental monitoring<sup>102</sup>, lagged household income (interpolated for some waves but still time-varying), lagged family intactness (as measured by whether the respondent lives with both parents), and contemporaneous educational tracking.

#### *Fixed and Random Effects Models of Delinquency and Work*

This chapter tests Hypotheses 4.1, 4.2, and 4.5 using fixed and random effects models, and Hypotheses 4.3 and 4.4 using causal models.<sup>103</sup> While fixed effects models are advantageous in that any unobserved person-specific characteristics do not bias the model parameters so long as they are time-invariant, they do suffer from the incidental parameters problem: when the number of model parameters approaches the sample size, regression coefficients can be biased, depending on the functional form of the model and the method in which coefficients are calculated (Allison 2009). Because the dependent variable is continuous and highly skewed, inverse Gaussian regressions would be a suitable choice for the modeling after transforming it to be positive for all waves. As fixed effects inverse Gaussian models are not commonly used in the

---

<sup>102</sup> For parental monitoring, two questions are consistently used across survey waves: “Does your mother know where you are when you are away from home?” and “Does your father know where you are when you are away from home?” Some additional questions are used to construct measures for the selection models (see the discussion in Appendix E), but the time-varying covariate used in the fixed effects models is a weighted average of these two questions.

<sup>103</sup> Hypothesis 3.5 is not tested with a causal model because the findings from the fixed and random effects models were null.

social sciences, whether this particular type of model is vulnerable to incidental parameters bias appears to have not yet been rigorously tested through simulation studies. Allison and Waterman (2002), however, show through simulation that negative binomial regression models do not suffer from incidental parameters bias when they are estimated unconditionally (that is, by including in the model  $n-1$  covariates corresponding to the fixed effects and a referent fixed effect, rather than by conditioning the fixed effects out of the likelihood during the estimation process). To avoid relying on statistical models with untested properties to reach model conclusions, the present research therefore converts the dependent variable to a pseudo “count” variable by converting every observation to a positive value through a linear transformation, multiplying by 1,000, and rounding to the nearest whole number.

The next section of the analysis yields some insight into the nature of attrition in the Taiwan Youth Project survey. Lin and Li (2013) examine the effect of missingness on models of deviance that use the TYP data. They predict Wave 1 deviance with a raft of Wave 1 covariates, splitting their sample into three groups corresponding to latent classes describing different “trajectories” for survey attrition between Waves 2 and 9 that they identified using group-based trajectory modeling (Nagin 2005). The parameter estimates for nearly every explanatory variable in each of their models do not differ by attrition group, suggesting that even though attrition is nonrandom in the TYP survey, the missingness patterns do not bias the inferences that researchers would draw when modeling deviance.<sup>104</sup>

---

<sup>104</sup> Were it possible to conduct, an important additional test would be to conduct the same analysis, but predicting *Wave 9* deviance with *Wave 9* covariates. Of course, such a test is not possible precisely because the *Wave 9* respondents are not representative of the original sample due to nonrandom survey attrition, and applying a weighting scheme to adjust for the nonrandom attrition would require making exactly the sort of assumptions about a mechanism for survey attrition that Lin and Li seek to demonstrate are unnecessary to make. In other words, carrying out the model for *Wave 9* would have required the authors to assume *a priori* exactly what they were hoping to prove, so logically they were forced to use *Wave 1* for their analysis.

**Table 4.4.** Fixed effects negative binomial (“FENB”) regressions of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months. (The *N*-1 fixed effects parameter estimates are not shown.)

	Model 1. Unconditional FENB Model – Full Sample		Model 2. Unconditional FENB Model – Lagged Covariates	
	coef.	SE†	Coef.	SE†
<b>Work</b>	<b>.039</b>	<b>.031</b>	<b>.0040</b>	<b>.032</b>
Wave 6	-.155 ***	.037	<b>-.173 ***</b>	<b>.038</b>
Wave 8	-2.93 ***	.032	<b>-2.94 ***</b>	<b>.035</b>
Wave 9	-.286 ***	.039	<b>-.241 ***</b>	<b>.042</b>
Intercept	8.44 ***	.453	8.66 *	5.17
N	2,120		1,762	
Person-years	7,071		5,632	

\*\*\* p<.01 \*\*p<.05 \* p < .10 † Outer product of the gradient.

**Table 4.5.** Random effects negative binomial regressions of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months. (The *N*-1 fixed effects parameter estimates are not shown.)

	Model 1. Random Effects NB Model – Full Sample		Model 2. Random Effects NB Model – Lagged Covariates	
	coef.	SE	Coef.	SE
<b>Work</b>	<b>.058 ***</b>	<b>.009</b>	<b>.048 ***</b>	<b>.009</b>
Wave 6	-.149 ***	.009	-.159 ***	.009
Wave 8	-2.76 ***	.022	-2.87 ***	.025
Wave 9	-.262 ***	.010	-.262 ***	.010
Intercept	2.68 ***	.021	2.87 ***	.022
N	2,120		1,762	
Person-years	7,071		5,632	

\*\*\* p<.01 \*\*p<.05 \* p < .10

Contrary to Lin and Li’s findings, Tables 4.4 and 4.5, which constitute initial tests of Hypothesis 4.1, suggest that perhaps missingness does in fact affect regression results for later TYP waves. These tables present fixed and random effects models for the effect of work on

deviance, using as the measures for deviance the Expected a Posteriori (EAP) estimates from Chapter 2. Each fixed and random effects model contains dummy variables for wave because as mentioned previously, the vertical scaling process created a quirk with the data: as some waves had questions that were particularly easy to answer, the average EAP deviance estimates for these waves were much lower than those for other waves, creating wave-specific effects on individual deviance due to the large prevalence of respondents reporting zero deviance.

Accordingly, the dummy variables for wave adjust for these artifacts of the data. The models are otherwise bereft of predictors, and the only difference between the left-hand and right-hand models in each table is the sample size (and for Table 4.4, the number of fixed effects in each model). The actual magnitude of the coefficients for work is difficult to interpret due to the several transformations performed on this variable, but the coefficient for work of .039 in the left-hand regression in Table 4.4 implies that going from no job to having a job is associated with an increase in deviance of  $e^{.039}$ , or about four percent, in a latent ability to commit delinquent acts as measured on the J3W4 scale. How this change in “ability” translates into an increase in delinquent behavior reported on a particular survey depends on the difficulty of the survey items. The magnitude of the effect of work in Table 4.5 is associated with an increase in latent deviance of about six percent ( $e^{.058} = 1.060$ ), still modest but not insignificant.

Model 1 on the left side of Table 4.4 contains all person-years associated with respondents for whom both work during the year and deviance is observed for at least two waves (as the fixed effects approach models within-individual change). Model 2 in Table 4.4, in contrast, takes the sample used in Model 1 and removes all person-years for which contemporaneous educational tracking, lagged family intactness, lagged household income, or lagged parental monitoring is not also observed. In other words, it represents the sample size for a model with these lagged

covariates included, but without the covariates actually in the model.<sup>105</sup> This approach eliminates 20.4% of the person-years, and as the rate of item nonresponse is low within survey waves, most of these missing cases are due to unit non-response during the wave when the lagged covariates are to be observed.

If survey attrition were a random process unrelated to work, deviance, or an unobserved time-varying confounder, then the point estimates for the effect of work on deviance in both models in Table 4.4 should be very close, and similarly the estimates should also be close in Table 4.5. Instead, the estimates in both models are attenuated for the sample with the approximately 20% of observations missing for tracking, family intactness, household income, and parental monitoring removed, suggesting that the relationship between work and deviance is weaker among respondents who complete the survey in successive survey years than among respondents who are first missing and then are interviewed the next year. Indeed, the effect of work in the fixed effects model on the right side of Table 4.4 is basically zero. Wu, Chiang, and Hwang (2013) observe that respondents with high social capital are less likely to be lost to follow-up, so the findings in Tables 4.4 and 4.5 would make sense if a positive relationship between work and crime exists primarily among youth with low social capital. Perhaps those with part-time jobs are less attached to school or their parents, and therefore less likely to be tracked down by the survey team for an interview.

It is likely that the type of missingness attenuating the model results in Tables 4.4 and 4.5 is unit nonresponse, as the item nonresponse rate is less than 2% in all survey years for parental monitoring, and nearly zero for educational tracking and family intactness. The item nonresponse rate for household income in Waves 4, 5, and 7 is higher (up to 13.4% in Wave 4), but as

---

<sup>105</sup> Normally the random effects model would have a slightly larger sample size than the fixed effects model as individuals observed for one person-year would remain in the random effects model, but here the sample sizes are exactly the same to aid comparability.

household income is actually imputed from other waves for these waves, item nonresponse in these cases merely reflects unit nonresponse in other survey waves. That said, the causal models each incorporate more than two dozen covariates in the selection models, and even missingness of a percent or two per covariate could potentially add up to a substantial loss of statistical power and potential bias. To illustrate this problem, Table 4.6 contains the item nonresponse patterns for all the time-varying covariates in the survey. Each cell in the table contains the percentage of respondents who completed a survey for a particular wave that did not respond to a survey question that they were eligible to complete (e.g. for parental attachment, only respondents with at least one living parent are included in the numerator and denominator).

**Table 4.6.** Item nonresponse patterns for variables affecting selection into work for the TYP survey. All covariates in this table are time-varying.

<b>Covariate</b>	<b>J1W1</b>	<b>J1W2</b>	<b>J1W3</b>	<b>J1W4</b>	<b>J1W5</b>	<b>J1W6</b>	<b>J1W7</b>	<b>J1W8</b>	<b>J1W9</b>
Household income			1.43	11.8	8.23		7.44	0.69	
Tracking (contemporaneous)				0.30	0.15	0.11		0.29	0.32
School / teacher attachment			0	0.71	0.67		0.09	0.46	
School commitment: class rank	0.48		0		6.8				
School commitment: time spent on homework			0				2.69	0.08	
School commitment: studying hard				0.22	0.56				
Parental attachment			0	0.21	1.23		0.17	0.23	
Parental monitoring			3.50	1.66	1.13	0.33		0.12	
Live with parents	0.19		0	0	0		0.06	0.06	
Peer deviance			3.75	14.2	8.52	0.33		0.06	
Prior deviance			0.04			0.06			

(Continued on next page)

**Table 4.6 (cont.)** Item nonresponse patterns for variables affecting selection into work for the TYP survey. All covariates in this table are time-varying.

<b>Covariate</b>	<b>J1W1</b>	<b>J1W2</b>	<b>J1W3</b>	<b>J1W4</b>	<b>J1W5</b>	<b>J1W6</b>	<b>J1W7</b>	<b>J1W8</b>	<b>J1W9</b>
Peer attachment			0.11			1.48			
Stressful life events			0	0.17	0.94	0		0.06	
Self-esteem			0	0.13		0.05		0	
Educational aspirations	1.4		0			6.20		0.52	
Autonomy			0				0.11		
Harsh and erratic discipline <sup>b</sup>			3.53	1.83		0.77			
Allowance			0.94	4.07		1.48			
If have student loans							0.09	0.08	
Cram school			0.08	0.22	0.77		0.46		
Social disorganization			0.26			0.05			
Neighborhood attachment			0.11			0		2.59	
School involvement	0.15		0				0.17	0.15	
Neighborhood involvement	0.04	0.04						0.12	
Family involvement			0.34	0.08	1.08				
Class cohesion			0.04				0.09	0.46	
Family cohesion	0		0			0			
Parental educational aspirations			22.2 <sup>a</sup>	8.04 <sup>a</sup>	7.84 <sup>a</sup>	15.5 <sup>a</sup>		10.7 <sup>a</sup>	
Adverse SES			0		1.28	0			
Strain from school			0	0	0.61				
Depression				0.08	1.13		0	0	
How often online			0.34	0.04	1.18			0	
Internet café use				0.09	1.18		1.49		
Materialism								0	
Deviance	0.26		.04	0.21	0.94	0.33		0.69	0
Worked during summer break 1				0.13	1.07			0.91	0.87
Hours worked during the week during summer break 1								1.56	2.08

(Continued on next page)

**Table 4.6 (cont.)** Item nonresponse patterns for variables affecting selection into work for the TYP survey. All covariates in this table are time-varying.

<b>Covariate</b>	<b>J1W1</b>	<b>J1W2</b>	<b>J1W3</b>	<b>J1W4</b>	<b>J1W5</b>	<b>J1W6</b>	<b>J1W7</b>	<b>J1W8</b>	<b>J1W9</b>
Months worked during summer break 1								1.56	2.08
Worked during summer break 2				0.13	1.07			0.60	0.72
Hours per week worked during summer break 2								2.55	2.29
Months worked during summer break 2								2.55	2.29
Worked at any time during year					1.07	0		0.46	0.59
Hours worked during the week during the year								3.22	2.23
Months worked during the year								3.22	2.23
Work during the academic year					1.07	0		0.83	0.72
Hours worked per week during the academic year								2.67	2.01
Months worked during the academic year								2.67	2.01
Mortimer work typology								2.34	1.88
Percentage answering all questions	97.5	99.96	90.19 <sup>b</sup>	71.19 <sup>c</sup>	78.30 <sup>c</sup>	91.66 <sup>c</sup>	90.18 <sup>d</sup>	93.35 <sup>e</sup>	97.83 <sup>f</sup>

<sup>a</sup> Denotes a covariate that was is not used in the selection models and also not used to compute the percentage of respondents answering all questions.

<sup>b</sup> Percentage of all respondents with both parents answering all questions, and excluding parental educational aspirations

<sup>c</sup> Percentage of all respondents who are in school, have at least one living parent, and answer all questions, excluding parental educational aspirations where that row is listed

<sup>d</sup> Percentage of all respondents who are in college, have at least one living parent, and answer all questions

<sup>e</sup> Percentage of all respondents who are in college, have at least one living parent, and answer all questions except for hours worked per week and weeks worked per year (which require having a job as a prerequisite for answering), and excluding parental educational aspirations

<sup>f</sup> Percentage of all respondents who are in college and answer all questions except for hours worked per week and weeks worked per year (which require having a job as a prerequisite for answering).

In general, the TYP data are remarkably clean. The most significant sources of nonresponse occur in Waves 4 and 5, when respondents state that they do not know how deviant their peers are and when household income is interpolated using information from neighboring waves. In short, item nonresponse is remarkably low, and in cases where it is affected, it is in part due to a spillover effect of unit nonresponse onto item nonresponse due to the interpolation procedure used to compute household income.

A strategy for addressing item nonresponse is necessary for two reasons. First, as household income is a theoretically important variable affecting selection into part-time work, it should not be excluded from analysis. Second and more importantly, the procedure for reweighting survey observations to account for selection bias involves multiplying together survey weights from previous survey years. The real cause for concern in Table 4.6, therefore, is the percentage of respondents who completed questionnaires for all survey waves relevant for the causal models but whose information is omitted from those models because they did not respond to at least one question in at least one survey year. The most egregious example of such a problem is Wave 4, for which about 29% of individual responses would be lost without imputing values that are missing due to item nonresponse.

A reasonable method to address item nonresponse, given the abundance and diversity of covariates, is to first assume that each covariate for a given survey wave is missing at random (MAR) conditional on the other observed covariates in Table 4.3 for that wave, and then to impute missing values due to item nonresponse using multiple imputation techniques. All models that follow use this approach to handle item nonresponse, and the causal models also address unit non-response using “inverse probability of missingness weighting,” or IPMW (e.g. Fewell, Hernán, Wolfe, Tilling, Choi, and Sterne 2004), a procedure for weighting observations by the

inverse of their probability of being observed that is analogous to inverse probability of treatment weighting (IPTW) and often conducted jointly with the latter. Details concerning the multiple imputation procedure, conducted with the `ice` multiple imputation extension in the `Stata` software package, are provided in Appendix F. The procedure produces 33 multiply-imputed data sets, and fitting the resulting regression models involves fitting 33 separate models and calculating a single summary value for each regression coefficient and standard error pooled from all 33 fits using methods initially proposed by Rubin (1987).<sup>106</sup>

### *Findings: Fixed and Random Effects Models of Delinquency and Work*

Tables 4.7 and 4.8 present the results of repeating the regressions on the right side of Tables 4.4 and 4.5, with 33 multiply-imputed data sets.<sup>107</sup> Once again, fixed and random effects negative binomial models are run for the contemporaneous effects of work during the year on deviance, controlling only for wave but using the same observations that would be used if time-varying covariates were included in the model. The results are essentially the same as those in Tables 4.4 and 4.5, suggesting that item non-response is not driving the difference between the fixed and random effects models.

---

<sup>106</sup> In particular, each point estimate in the combined analysis is the mean of the corresponding individual point estimates, and the variance of each point estimate is a weighted sum of the within-imputation variance and between-imputation variance for the corresponding point estimates for each data set (White, Royston, and Wood 2010).

<sup>107</sup> The models in this section (covering Tables 4.7-4.10) are based on multiple imputations that include all covariates in Table 4.6, including the aspirations that parents have for their child's educational attainment, which are the child's assessment of these aspirations for waves in which the parent was not interviewed or in cases where the parent survey was administered but the parent could not be reached. The youth respondents often stated that they simply did not know what their parents' aspirations were. For such cases multiple imputation is ideal, as an actual observable value of the response exists but could not be observed, and the fact that the value was not observed is plausibly unrelated to the actual value of the response were it to have been observed. Accordingly, the fact that imputations for covariates used in these tables make use of a variable with a somewhat high item nonresponse rate is unlikely to bias the findings. Nonetheless, the rest of the analyses in this chapter take a more conservative approach and are based on multiple imputations and selection models that omit this variable from the analysis.

**Table 4.7.** Fixed and random effects negative binomial regressions of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months (after multiple imputation). These are the regressions shown on the right half of Tables 4.4 and 4.5.

	Unconditional Fixed Effects NB Model – Lagged Covariates			Random Effects NB Model – Lagged Covariates with Multiple Imputation		
	coef.		SE†	Coef.		SE
<b>Work</b>	<b>.005</b>		<b>.031</b>	<b>.048</b> ***		<b>.009</b>
Wave 6	-.058 ***		.012	-.053 ***		.002
Wave 8	-.730 ***		.008	-.714 ***		.006
Wave 9	-.049 ***		.008	-.053 ***		.002
Intercept	8.66 *		4.89	2.90 ***		.022
Multiple Imputations	33			33		
N	1,818			1,818		
Person-years	5,801			5,801		

\*\*\* p<.01 \*\*p<.05 \* p < .10 † Outer product of the gradient.

**Table 4.8.** Fixed and random effects negative binomial regressions of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months, controlling for all time-varying covariates of theoretical interest measured in the same way in Waves 4, 5, 7, and 8.

	Unconditional Fixed Effects NB Model – Lagged Covariates With Multiple Imputation			Random Effects NB Model – Lagged Covariates with Multiple Imputation		
	Coef.		SE†	Coef.		SE
<b>Work</b>	<b>.008</b>		<b>.031</b>	<b>.044</b> ***		<b>.009</b>
Wave 6	-.056 ***		.012	-.053 ***		.002
Wave 8	-.738 ***		.010	-.716 ***		.006
Wave 9	-.059 ***		.010	-.058 ***		.002
Vocational track <sup>a</sup>	-.039		.056	.012		.008
Not in school (track) <sup>a</sup>	-.043		.063	.011		.015
Lagged parental	-.044 ***		.015	-.021		.004
Lagged household income	-.0004		.0009	.00025 **		.00013
Lagged living with parents	-.043		.036	-.014		.010
Intercept	8.91		7.62	2.97 ***		.029
Multiple Imputations	33			33		
N	1,818			1,818		
Person-years	5,801			5,801		

\*\*\* p<.01 \*\*p<.05 \* p < .10 † Outer product of the gradient.

<sup>a</sup> Referent category: general track

Leaving the matter of unit non-response for the causal analyses, the next question, then, is whether the fixed or random effects model more closely reflects the “truth” in each table, and the conventional approach for testing whether the assumption of the random effects model that the time-varying covariates of interest are all uncorrelated with the fixed effects for individuals would be to perform a Hausman test (Hausman 1978). Such a test, however, cannot be conducted for models containing time fixed effects, as in the present case, and an alternative is to conduct an auxiliary random effects regression that includes as control variables the average value over all time points for each time-varying covariate (Schmidheiny 2014). The random effects assumption then holds if an  $F$  test of a null hypothesis that all such covariates are jointly equal to zero then constitutes a test of the random effects assumption. Table 4.9 presents the results of the auxiliary regression, and includes the time-varying covariates in Table 4.8.

To perform an  $F$  test that the six parameter estimates in Table 4.9 associated with the means of work, parental monitoring, household income, living with parents, and the two dummy variables for educational tracking are all jointly equal to zero, I use the `testparm` function as implemented by Stata’s `mim` package, which implements a procedure developed by Li, Raghunathan, and Rubin (1991) for jointly testing hypotheses involving vectors of parameters in the case of multiple data sets created through multiple imputation. The test fails to reject a null hypothesis that all six parameters are equal to zero, suggesting that the random effects model in Tables 4.7 and 4.8, which suggest a positive relationship between work during the year and deviance, is reasonable in the present case. Finally, because random effects models can also include observations associated with respondents only observed for a single person-year, Table 4.10 brings in these additional 401 person-years of observations. Table 4.10 additionally presents

**Table 4.9.** Auxiliary random effects negative binomial regression of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months.

	Auxiliary Random Effects NB Model With Lagged Covariates	
	coef.	SE
<b>Work</b>	<b>.046</b>	<b>.035</b>
Work (mean)	.030	.048
Wave 6	-.052	*** .002
Wave 8	-.535	*** .014
Wave 9	-.069	*** .005
Vocational track <sup>a</sup>	-.035	.064
Vocational track (mean)	-.002	.071
“Not in school” track <sup>a</sup>	-.043	.063
“Not in school” track (mean)	.025	.086
Parental monitoring	-.055	*** .018
Parental monitoring (mean)	-.037	* .082
Household income	.00009	.00079
Household income (mean)	.00091	.00079
Living with parents	-.058	.042
Living with parents (mean)	.065	.048
Intercept	8.99	.075
Multiple Imputations	33	
N	1,818	
Person-years	5,801	

\*\*\* p<.01 \*\*p<.05 \* p < .10

<sup>a</sup> Referent category: general track

both negative binomial and inverse Gaussian models<sup>108</sup>, as incidental parameters bias is not a concern for random effects models.

<sup>108</sup> The inverse Gaussian model presented here uses a log link function instead of the inverse-squared canonical link function (McCullagh and Nelder 1989), due to problems achieving convergence with the canonical link function in Stata. The effect sizes of the parameter estimates are interpreted as follows: the coefficient is exponentiated, and the result is the multiplicative effect of the covariate of interest on the mean level of deviance for respondents with the other covariates held fixed.

**Table 4.10.** Random effects negative binomial regression of delinquency on whether respondents to Waves 5, 6, 8, and 9 had a part-time job during the previous 12 months, including respondents only observed for one person-year and incorporating additional time-invariant covariates.

	Random Effects Negative Binomial Model With Time-Invariant and Lagged Time-Varying Covariates		Random Effects inverse Gaussian Model With Time-Invariant and Lagged Time-Varying Covariates			
	coef.	SE	Coef.	SE		
<b>Work</b>	<b>.055</b>	<b>***</b>	<b>.0086</b>	<b>.258</b>	<b>**</b>	<b>.113</b>
Wave 6	-.053	***	.0028	-.043	***	.0064
Wave 8	-.716	***	.0062	-.76	***	.026
Wave 9	-.059	***	.0021	-.14	***	.020
Vocational track <sup>a</sup>	.011		.0078	-.091		.112
“Not in school” track <sup>a</sup>	.021		.014	-.193		.157
Lagged parental monitoring	-.021	***	.036	-.253	***	.043
Lagged household income	.00019		.00013	.0032	*	.0018
Lagged living with parents	-.023	**	.0091	-.101		.111
Gender (female=1)	-.092	***	.0072	-1.10	***	.091
Taipei County <sup>b</sup>	-.0080		.0085	-.152		.126
Yilan County <sup>b</sup>	-.022	**	.0095	-.283	**	.141
Mother’s years of education	-.0018		.0016	-.0088		.021
Father’s years of education	.0017		.0015	.025		.025
Respondent’s birth year	.0077		.0074	.168	**	.079
Intercept	3.02	***	.034	3.39	***	.300
Multiple Imputations	33					
N	2,219					
Person-years	6,202					

\*\*\* p<.01 \*\*p<.05 \* p < .10

<sup>a</sup> Referent category: general track

<sup>b</sup> Referent category: Taipei City

Regardless of the functional form of the random effects model, work is positively associated with deviance, affirming Hypothesis 4.1.<sup>109</sup> Because the analyses thus far have raised the possibility that unit nonresponse may be biasing the results for the fixed and random effects models, the flow of the chapter now proceeds to a test of Hypothesis 4.3 – whether work during the year actually causes deviance. Hypotheses 4.2 and 4.4, concerning work during holidays and an interaction between social control and work, will then be tested in a causal framework such that possible bias due to unit nonresponse is first corrected.

#### *Assumptions and Specification of Marginal Structural Models for Causal Inference*

This section makes use of marginal structural models for causal inference (MSM) that weight observations based on the inverse of the probability of receiving the treatment in order to derive causal inferences from observational data (Hernán, Brumback, and Robins 2000; Robins 2000). Such models, in which observed variables that would confound the effect of a treatment on the outcome are explicitly used to weight the observations in the survey but are not directly included in regressions, avoid biases currently found in structural equation models for time-varying treatments (Hernán, Brumback, and Robins 2000; Robins, Hernán, and Brumback 2000). Inverse Probability of Treatment Weighting (IPTW) denotes a type of marginal structural model in which individuals unlikely to receive the treatment receive higher weights. After the reweighting has taken place, differences between the group receiving the treatment and not

---

<sup>109</sup> A goodness-of-fit test of the two random effects models would normally be achievable by comparing the BIC scores for the two models (Raftery 1995), but in the present case, the random effects inverse Gaussian model is fit using a generalized estimating equation (`-xtgee-`) in Stata, and likelihoods are not available for such models. The negative binomial model in Table 4.10 is fit using Stata's `-xtnbreg-` procedure; an alternative is to fit both models using generalized estimating equations, then compute the quasi-likelihood information criterion, or QIC (Pan 2001), for both models. The QIC for the negative binomial model with random effects, however, is too large to be credible, suggesting a problem with the estimation. Given the findings for the effect of work on deviance for both models in Table 4.10 are significant and positive, I choose not to examine the fit for the two models in any greater detail.

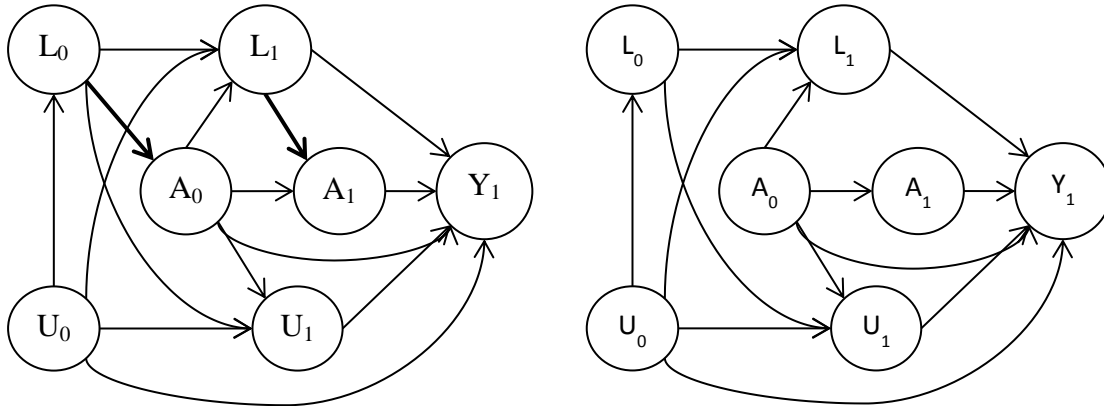
receiving the treatment are assumed to be statistically random, thus approximating a randomized controlled trial and solving the missing data problem at the heart of counterfactual inference.<sup>110</sup>

The weights can then be applied within a multilevel modeling framework using a statistical package that supports weighting, with the effect of the treatment on the outcome interpreted as causal, as in Sampson et al. (2006).

This class of models contains its own set of assumptions, as indicated graphically in Figure 4.1. In the graph, a set of unobserved time-varying confounders  $U$  are assumed to have no contemporaneous effect on a time-varying treatment  $A$ , conditional on the observed confounders  $L$ . Reweighting observations in the sample based on selection into the treatment then eliminates the effect of the observed confounders on the treatment (the bold arrows in the left panel), allowing for a causal interpretation of the parameters associated with a regression of the outcome  $Y$  on the time-varying treatment  $A$ . The counterfactual interpretation of the treatment effect is of course violated as soon as a single unobserved confounder  $U$  directly affects both the treatment and the outcome net of the observed confounders, making the literature review to identify potential confounders so important.

---

<sup>110</sup> Counterfactual questions are at their core a “missing data” problem because they involve inferring what would have happened to some outcome variable  $Y$  if some treatment  $A$  had been  $B$  instead. Now if the treatment could somehow be observed as both  $A$  and  $B$  for every subject at the same time, and the outcome associated with each treatment were also simultaneously observed, then the result of changing  $A$  to  $B$  could be modeled explicitly via a regression of the outcome on the treatment. Of course, at best the treatment can only be one of these two hypothetical values (and often the treatment is far more complex than a binary outcome). Counterfactual inference can be seen as analogous to a missing data problem in which at least half the data are missing.



**Figure 4.1.** A model for the causal effect of a time-varying treatment (A) on an outcome (Y), at two time points,  $t_0$  and  $t_1$ . L denotes a vector of observed confounders, and U denotes unobserved confounders. Arrows denote causal effects. The left panel shows the model before reweighting, and the right panel shows the panel after reweighting.

Following the assumptions of marginal structural models, covariates corresponding to causal effects are unbiased under two additional conditions. The first, known as “experimental treatment assignment,” or ETA (Wang, Bangsberg, Petersen, and van der Laan 2006), and alternatively referred to as the positivity assumption (Cole and Hernán 2008), requires that the selection model never estimates the probability of treatment as exactly 1 or exactly 0 for any combination of covariates in the selection model. In order for this assumption to hold, every combination of covariate values must be observable (or else the probability of treatment would be exactly zero for that combination of covariates). One situation in which positivity could have been violated for the present study involves the fact that in the Waves 8 and 9 surveys for the J1 cohort, military service is measured as a type of educational tracking separate from being otherwise not in school. Republic of China law mandates a year of military service for males only, such that no female respondents to the TYP survey serve in the military. In order to avoid a situation in which the probability of part-time work is zero for female respondents serving in the

military, the educational tracking variable used in this dissertation is coded by collapsing together the two categories corresponding to military service and otherwise not being in school, thereby losing some information but preserving the validity of the ETA assumptions.

Additionally, in flagging situations in which levels of one categorical variable perfectly predict levels in another, the `-ice-` package for `Stata` used for multiple imputation also serves as a tool for verifying that positivity is not violated (see Appendix F for more details). The second assumption, known as the “sequential randomization assumption” or SRA (Wang, Bangsberg, Petersen, and van der Laan 2006), or alternatively as exchangeability (Cole and Hernán 2008), requires that the model for selection into treatment is properly specified. Unlike experimental treatment assignment, which, can be tested empirically, sequential randomization cannot be tested and can only be considered in light of theoretical knowledge of the selection process at hand.<sup>111</sup>

Computations of weights based on the inverse of the probability that individual  $i$  receives treatment at a particular time  $t$ ,  $w_i(t)$ , are based on the probabilities of treatment at all times prior to  $t$  under the following formula:

$$w_i(t) = \prod_{k=1}^t \frac{1}{f[A_i(k) | \bar{A}_i(k-1), \bar{V}_i, \bar{L}_i(k)]} . \quad (\text{Equation 4.1})$$

In Equation 4.1,  $f$  denotes the probability that the  $i$ th person receives treatment at time  $k$ , conditional on treatment at time  $k-1$ , a vector of time-varying confounders  $\bar{L}$ , and a vector of time-invariant baseline confounders  $\bar{V}$ . In the case of work, a dichotomous treatment,  $f$  is a

---

<sup>111</sup> One way to relax this assumption is to make use of double robust estimation methods, which simply require that either the selection model or the model of the treatment on the outcome is specified in order for the regression results to be consistent (Kang and Schafer 2007). This dissertation does not use double robust estimation because not enough theoretically-relevant control variables are consistently available across waves of the Taiwan Youth Project Survey to justify an assumption that a regression of work during the year on deviance is properly specified.

logistic probability density function when modeling the presence or absence of work. The confounders, such as family intactness or peer delinquency, may affect both crime and selection into work and are observed at or before time  $k$ . Time  $t = 1$  denotes the year in which the treatment is initiated; in the present case, it is the first year in which respondents are asked about work.  $A$  denotes the treatment that the respondent receives. The product sign in Equation 4.1 captures the concept that the type of treatment received in one survey year may depend on an entire history of treatments and observed confounders. The procedure thus requires computing the propensity scores for selection into treatment (work patterns) for each year of the study, then multiplying the inverse of each propensity score together to compute weights for each person-year. The researcher's aim is to capture all the major theoretical reasons why an individual would select one kind of work over another.<sup>112</sup>

Sequential ignorability, in fact, is a strong assumption, especially in the case of life course research in which people may be subjected to a wide variety of changing environments as they age, and the set of environmental conditions that are confounders in one survey year may be replaced by a different set in the next year. Meanwhile, the surveys themselves may remain the same, and thereby fail to record important confounders. The fact that the TYP survey questions

---

<sup>112</sup> The intuition for multiplying probabilities together in Equation 4.1 might be best understood with a metaphor. Imagine that an unbalanced sample used for counterfactual inference is like a hospital bed with  $t$  sheets, and wrinkles in the sheets represent bias. A researcher attempting to adjust for confounding in a longitudinal sample is analogous to a slightly obsessive hospital orderly seeking to remove all the wrinkles in the bed. When  $k=t$  in Equation 4.1, a computation of  $1/f(A_i(k))$  is analogous to the orderly smoothing the topmost sheet. The sheet underneath the topmost sheet represents the set of confounders and treatments at time  $t-1$ . The topmost sheet will retain some wrinkles, however, if the sheet underneath the topmost sheet is also wrinkled due to the presence of confounders occurring prior to time  $t-1$ . In order to correctly remove all the wrinkles, the orderly must start with the bottommost sheet (time  $t = 1$ ) and proceed upwards, until all the sheets are smooth. The reason for calculating products in Equation 4.1 instead of sums has to do with least common multiples: assume the "sheet" corresponding to time  $t-2$  is smooth. If the  $i$ th individual must receive a weight of  $a_i$  in order for the "sheet" corresponding to time  $t-1$  to be "unwrinkled" by the confounders at time  $t-1$ , and a weight of  $b_i$  for the "sheet" corresponding to time  $t$  to be "unwrinkled" by the confounders at time  $t$ , then a weight of  $a_i b_i$  will ensure that all the sheets are smooth, and therefore the bed is wrinkle-free (assuming time  $t-2$  denotes the beginning of treatment). The assumption associated with this procedure is that the orderly has access to every sheet in the bed (i.e. all relevant data is collected) or that the bottommost sheet to which the orderly has access is smooth (i.e. no confounding exists between the treatment and the outcome at time  $t=1$ ).

change each year forces a researcher to change the set of confounders included in the regressions predicting treatment with each wave, and in so doing requires a reader to more directly confront the question of whether sequential ignorability is a plausible assumption than may be the case for research using other data sets. Does it really make sense that the mechanisms for selection into work vary slightly with each survey wave? Such an assumption has its drawbacks, but is still at least as plausible – if not more so – than one assuming that the mechanisms do not change as survey respondents age and enter different stages of their life course. Indeed, in any case the fact that individuals enter different stages of their life course during different times in their lives further complicates matters, so that in practice researchers use a variety of approaches to demonstrate that their findings are robust to variations in model specification and that the standard errors for parameter estimates associated with the treatment effects are not driven up by large weights. I make use of two techniques in this chapter and the next to address such problems: stabilized weighting (Hernán, Brumback, and Robins 2000) and weight truncation (e.g. Cole and Hernán 2008).

In practice, when the confounders predict selection into treatment well, most weights would be close to 1 but a few respondents may be observed receiving a particular treatment despite being extremely unlikely to do so. In such cases, these observations would receive enormous weights equivalent to dozens or even hundreds of observations. The problem tends to worsen during later waves of a survey, for as  $k$  increases in Equation 4.1, the denominator becomes smaller and the weights increase. To compensate, Robins, Hernán, and Brumback (2000) recommend the use of stabilized weights (Equation 4.2).

$$sw_i(t) = \prod_{k=1}^t \frac{f[A_i(k) | \bar{A}_i(k-1), V_i]}{f[A_i(k) | \bar{A}_i(k-1), V_i, \bar{L}_i(k)]} \quad (\text{Equation 4.2})$$

To Equation 4.1 is added a numerator that constitutes the notion of a baseline probability that the respondent receives a particular treatment. Equation 4.2 denotes the probability that the  $i$ th person receives treatment at time  $k$ , conditional on treatment at time  $k - 1$ , divided by the probability that the  $i$ th person receives treatment at time  $k$ , conditional on treatment at time  $k - 1$  and a vector of confounders  $\bar{L}$  (such as family intactness or peer delinquency.)<sup>113</sup> In the numerator of Equation 4.2,  $V$  denotes a vector of baseline covariates; in the present case, I use the time-invariant covariates in Table 4.1. Individuals who, given their age and treatment history alone, are unlikely to have received their observed treatment, do not receive an unusually high weight under this weighting system. In addition, weights do not increase as quickly for observations in later waves. The findings presented here use stabilized weights. Additionally, throughout this dissertation the weights are truncated to prevent extreme values for the weights from driving up the variance of the estimated treatment effects. Following Cole and Hernán (2008), all weights with values larger than the 99<sup>th</sup> percentile are recoded to equal the weight at the 99<sup>th</sup> percentile, and those with values smaller than the first percentile are recoded to the weight for the first percentile. Truncation occurs as a final step prior to fitting the causal models, and uses weights that correct both for selection into treatment and for survey attrition.

---

<sup>113</sup> In practice,  $A_i(k)$  in the numerator may also be conditioned on a few covariates, such as age, which may be considered part of a “baseline” model for selection into the treatment. In the present case, age is included in the numerator because the age-crime curve is so ubiquitous (e.g. Gottfredson and Hirschi 1990).

Inverse Probability Weighting to Address Unit Nonresponse

The fixed effects models for work and deviance presented earlier in this chapter demonstrated how for the present analysis, attrition in the TYP survey may lead to biased estimates of the relationship between work and deviance. Apart from multiple imputation methods, researchers also correct for bias due to missing data using inverse probability weighting methods (Seaman and White 2011). The approach is analogous to IPTW methods, with indicator variables for responding to each survey year replacing the time-varying treatment as the set of dependent variables for selection models. The probability of a TYP survey participant responding to the survey in a given year is modeled explicitly using observed variables from previous years, and weights for the observations in that year are then computed according to the inverse of the probability that the respondent was observed. Equation 4.3 gives the formula used in this dissertation for weighting observations based on a history of survey nonresponses, with stabilized weights:

$$w_{MISSING_i}(t) = \prod_{j=1}^t \frac{g[M_i(j) | Mprev_i(j)]}{g[M_i(j) | Mprev_i(j), V_{M_i}]} \quad (\text{Equation 4.3})$$

$g[M_i(j)]$  denotes a function predicting the probability that respondent  $i$  does not complete a survey at time  $j$ , and is modeled using a logistic regression. Let  $M_i(j)$  be the value of an indicator variable corresponding to responding to the survey in time  $j$  for person  $i$ . Then

$$Mprev_i(j) = M_i(1) \prod_{l=1}^3 M_i(j-l) \text{ for } j > 4 \text{ and } Mprev_i(j) = \prod_{l=1}^j M_i(l) \text{ for } j \leq 4. \text{ As one of the best}$$

predictors of nonresponse is lagged nonresponse, this definition for  $Mprev_i(j)$  indicates that the regressions predicting nonresponse in each wave being modeled control for nonresponse in all

prior waves through Wave 4, and in Wave 1 and the three most recent waves for Waves 5 and later.<sup>114</sup> In Equation 4.6,  $V_M$  denotes a vector of baseline predictors for nonresponse, observed at time  $t=1$ . Previous researchers in Taiwan have already tested hypotheses associated with the social mechanisms governing survey attrition in the TYP (Lin and Li 2013; Wu, Chiang, and Hwang 2013), so models including the variables they identify in predicting nonresponse stand a good chance of compensating for some of the bias due to unit nonresponse. As the regressions predicting missingness all use covariates from a single wave, when estimating stabilized weights I choose to leave only the indicator variables for response to previous survey waves in the numerator.

Also note that the counter for time uses a different letter in Equation 4.3 than in Equation 4.2, emphasizing the fact that the “clock” used to measure the first observation to predict unit nonresponse need not be the same as the “clock” for which treatment begins, as the theoretical models governing sample attrition and selection into treatment may require different start times (and do in this dissertation). In the present case, the first wave in which work is observed is Wave 4 (during summer vacation), so Wave 5 is treated as time  $t = 1$  in Equation 4.2 (with work during the previous summer used as a lagged predictor for work during the year in Wave 5). In Equation 4.3, however, Wave 2 is the first wave for which survey nonresponse can be predicted using lagged Wave 1 predictors, so it is the first wave for which weights are computed.<sup>115</sup> In

---

<sup>114</sup> The reason for not including an indicator variable for nonresponse in every previous survey wave is to avoid an accidental violation of the experimental treatment assignment assumption for marginal structural models, a situation which would occur if a particular combination of nonresponses were not observed for any survey respondents.

<sup>115</sup> Technically, following the notation in Equation 4.3,  $t=1$  corresponds to Wave 1. The nonresponse weights, however, are all fixed to 1 for respondents observed during that wave as there is no lagged nonresponse indicator variable to use as a predictor for Wave 1 nonresponse. But were I to conduct this analysis again, I would also use time-invariant covariates not specific to Wave 1 such as gender, parental education, birth year, and location (Taipei City, Taipei County, or Yilan County) to predict nonresponse at Wave 1, in order to compute nonresponse weights for Wave 1 as well. Given the fact that more than 95% of respondents responded to Wave 1 and gender, birth year, and location of residence were not predictive of nonresponse for Wave 2, however, the fact that this analysis did not do so is not likely to affect the findings.

order to obtain weights via Equation 4.3 for the 5.4% of respondents who did not respond to Wave 1 and for whom values of the baseline predictors  $V_M$  are not observed, those predictors are assigned arbitrary values (9999 here) for respondents who do not answer Wave 1, then interacted with the indicator variable for responding to Wave 1; the interaction effects are then incorporated into the two regressions in Equation 4.3 without a main effect for each of the baseline predictors. These interaction terms thus capture the effect of the baseline predictors for respondents answering Wave 1 of the survey.

Of course, the advantage of using inverse probability weighting to correct for unit nonresponse in the present research is that it is easily integrated with the IPW methods that are already being used to derive causal inference. Adjusting the IPTW approach to include a missingness model, the formula to compute stabilized inverse probability weights for the causal analysis corrected for nonresponse now becomes:

$$w_i(t) = w_{MISSING_i}(t) \bullet w_{IPTW_i}(t) . \quad (\text{Equation 4.4})$$

To date, Taiwanese researchers have conducted two studies of missingness in the TYP survey. Wu, Chiang, and Hwang (2013) propose that respondents with high social capital are more likely to have responded to surveys because Academia Sinica's research team was more likely to reach them each year via contacts in their schools and families. They conduct a rigorous set of 42 separate regressions, predicting nonresponse in Waves 4 through 10 with lagged covariates from Waves 1 through 9 (with predictors from only one wave used for each regression). Significantly, the effect of grades, living with parents, deviance, and location of the respondent (Yilan County, Taipei City, or Taipei County) in Wave 1 predict nonresponse in nearly every wave for nearly every variable – with fairly stable effect sizes. With respect to

school, they find that students with high grades and who commit themselves to small leadership roles in their classroom are less likely to be lost to follow-up. Concerning the family, respondents who express strong family attachment and live with both parents are much more likely to respond. Importantly for the present research, they also find that deviant respondents are more likely to be lost to follow-up. Finally, respondents living in Yilan County are more likely to be retained in the survey than respondents in Taipei City or Taipei County (possibly because residents of rural communities are less likely to move). The nonresponse models for the present chapter therefore use the Wave 1 variables that Wu and his colleagues (2013) found to be predictive of attrition.

Table 4.11 presents the results of the logistic regressions for the denominator in Equation 4.3 for each wave of the TYP survey; the magnitude and significance of the terms in the baseline regressions for the numerator (which only contain dummy variables for response to each wave) are similar and not shown. The findings replicate those of Wu, Chiang, and Hwang (2013): school grades and living with one's parents during seventh grade reliably predict survey response rates even through college, while to a lesser extent deviance in seventh grade predicts a lower

**Table 4.11.** Logistic regressions predicting missingness in the Taiwan Youth Project survey (J1 cohort).

Variable	Wave 2			Wave 3			Wave 4			Wave 5		
	Coef. <sup>a</sup>		S.E.	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.
Wave 1 <sup>b</sup>	1.48	*	.82	-4.84	**	.97	-2.64	***	.56	-2.24	***	.51
Wave 2 <sup>b</sup>				5.01	**	.29	1.15	***	.28	.87	***	.31
Wave 3 <sup>b</sup>					*		3.39	***	.25	1.37	***	.30
Wave 4 <sup>b</sup>										2.62	***	.14
Female <sup>b</sup> * J1W1	-.039		.21	-.21		.21	.17		.13	-.082		.11
Taipei County <sup>c</sup> * J1W1	-.042		.23	-.17		.24	.28	**	.14	-.046		.12
Yilan County <sup>c</sup> * J1W1	-.36		.28	.58	*	.33	.48	***	.18	.14		.15
Birth year * J1W1	.32		.20	.057		.21	.38	***	.12	.19	*	.11
Mother's years of education * J1W1	-.044		.044	-.0039		.044	-.031		.026	.0016		.022
Father's years of education * J1W1	-.069		.043	.022		.043	.019		.025	-.016		.022
Wave 1 family cohesion * J1W1	.304		.43	.28		.46	.15		.28	.34		.24
Wave 1 deviance (logged) * J1W1	-.083	**	.041	-.055		.043	-.059	**	.025	-.046	**	.023
Wave 1 Live with parents * J1W1	1.12	***	.22	.55	**	.25	.71	***	.15	.42	***	.14
Wave 1 school involvement * J1W1	-.409		.33	-.28		.35	-.015		.21	-.33	*	.18
Wave 1 school commitment * J1W1	.210	**	.10	.069		.10	.29	***	.063	.22	***	.054
Wave 1 educational aspirations * J1W1	-.014		.034	.026		.37	.022		.022	.058	***	.018
Intercept	.616	***	.17	1.50	**	.35	-2.91	***	.31	-3.58	***	.36

\* p < .10 \*\* p < .05 \*\*\* p < .01. Sample size for all models: 2,844.

<sup>a</sup> All coefficients and standard errors are based on regression estimates for 33 multiply-imputed data sets, combined using Rubin's rules.

<sup>b</sup> Denotes an indicator variable.

<sup>c</sup> Referent category: Taipei City.

**Table 4.11 (cont.)** Logistic regressions predicting missingness in the Taiwan Youth Project survey (J1 cohort).

Variable	Wave 6			Wave 7			Wave 8			Wave 9		
	Coef. <sup>a</sup>		S.E.	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.
J1W1 <sup>b</sup>	-.77	*	.45	-1.32	***	.47	.86	*	.47	-.59		.50
J1W3 <sup>b</sup>	1.91	***	.29									
J1W4 <sup>b</sup>	.84	***	.14	.94	***	.15						
J1W5 <sup>b</sup>	1.33	***	.11	1.26	***	.11	.86	***	.11			
J1W6 <sup>b</sup>				1.49	***	.10	1.22	***	.10	.79	***	.11
J1W7 <sup>b</sup>							1.40	***	.10	1.04	***	.11
J1W8 <sup>b</sup>										2.17	***	.11
Female <sup>b</sup> * J1W1	.023		.095	-.13		.10	-.0075		.099	-.089		.11
Taipei County <sup>c</sup> * J1W1	-.12		.11	.24	**	.11	-.17		.11	.18		.12
Yilan County <sup>c</sup> * J1W1	1.19	***	.15	.52	***	.14	.17		.14	-.093		.15
Birth year * J1W1	.054		.099	.066		.10	-.19	*	.10	.10		.11
Mother's years of education * J1W1	.012		.020	-.0026		.021	-.0042		.020	-.0049		.022
Father's years of education * J1W1	-.038	**	.020	-.0017		.020	-.037	*	.020	.0042		.022
Wave 1 family cohesion * J1W1	.31		.22	.27		.23	.091		.23	-.47	*	.25
Wave 1 deviance (logged) * J1W1	.013		.021	-.065	***	.021	-.030		.021	-.012		.023
Wave 1 Live with parents * J1W1	.27	**	.13	.35	**	.14	.25	*	.14	.47	***	.15
Wave 1 school involvement * J1W1	-.11		.16	-.17		.16	-.010		.16	.16		.18
Wave 1 school commitment * J1W1	.092	**	.047	.16	***	.050	.093	*	.049	.084		.054
Wave 1 educational aspirations *	.023		.017	.022		.017	-.024		.017	-.00049		.019
J1W1												
Intercept	-3.12	***	.33	-2.67	***	.23	-2.08	***	.22	-1.37	***	.22

\* p < .10 \*\* p < .05 \*\*\* p < .01. Sample size for all models: 2,844.

<sup>a</sup> All coefficients and standard errors are based on regression estimates for 33 multiply-imputed data sets, combined using Rubin's rules.

<sup>b</sup> Denotes an indicator variable.

<sup>c</sup> Referent category: Taipei City.

**Table 4.12.** Logistic regressions predicting part-time work at any time during the year in the Taiwan Youth Project survey (J1 cohort).

Variable	Wave 5			Wave 6			Wave 8			Wave 9		
	Coef. <sup>a</sup>		S.E.	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.
Lagged work during the year				1.64	***	.19	.45 <sup>e</sup>		.28	1.15	***	.14
Lagged work during the summer	1.26	***	.13									
Female	.29	**	.14	-.22		.20	.37	**	.16	.83	***	.14
Taipei County <sup>b</sup>	.048		.14	-.16		.20	-.27		.19	-.083		.15
Yilan County <sup>b</sup>	-.29		.18	-.58	**	.24	-.45	**	.20	.0055		.17
Birth year	-.14		.12	.10		.18	.023		.15	-.086		.13
Mother's years of education	-.015		.026	-.0025		.040	-.090	***	.035	-.026		.029
Father's years of education	-.022		.025	-.011		.038	-.020		.032	-.042		.027
Household income	-.096 <sup>j</sup>		.097	-.096 <sup>j</sup>		.20	-.021 <sup>j</sup>		.16	.099 <sup>j</sup>	*	.058
Vocational track <sup>c,d</sup>	.87	***	.15	1.66	***	.25	.26		.19	.178		.15
"Not in school" track <sup>c,d</sup>	1.5	***	.35	3.32	***	.48	.076		.26	1.16	***	.23
Parental attachment	.15	*	.08	-.056		.12	.091		.10	.047		.11
Parental monitoring	-.057		.061	-.089		.11	-.072 <sup>e</sup>		.071	.12		.094
Living with parents	-.26		.17	-.55	**	.24	.48	***	.17	.16		.14
Family involvement	-.16	***	.058	-.019		.082						
Family cohesion	-.010		.11				.031 <sup>e</sup>		.13			
Harsh / erratic discipline	-.10 <sup>j</sup>		.17				.043 <sup>e,j</sup>		.033			
Peer deviance	.34	*	.18	.24		.23	.12 <sup>e,j</sup>		.12	.034		.044
Prior deviance	.10 <sup>j</sup>	**	.027	.10 <sup>j</sup>		.072	-.30 <sup>j</sup>		.42	.065		.054
Stressful life events	.012 <sup>j</sup>		.027	.021 <sup>j</sup>		.031	.039 <sup>e,j</sup>	*	.020	.031	*	.018
Self-esteem	.074		.15				-.17		.17	.25	*	.15
Educational aspirations <sup>i</sup>	-.14 <sup>e</sup>		.097				-.056 <sup>e</sup>		.098	-.055		.10
Autonomy	.078 <sup>e</sup>		.15				.069		.077			
Social disorganization	.073 <sup>e,j</sup>		.17	-.038 <sup>c,j</sup>		.077	-.017 <sup>e,j</sup>		.075			
Neighborhood attachment	-.051 <sup>e</sup>		.13				-.028 <sup>e</sup>		.10	.0073		.11
Neighborhood involvement	.15 <sup>f</sup>	*	.074							-.060		.11
Depression	.36 <sup>j</sup>		.25	.23 <sup>j</sup>		.37	.11 <sup>j</sup>		.20	.056 <sup>j</sup>		.24
N	1891			1552			1178			1508		

\* p < .10 \*\* p < .05 \*\*\* p < .01.

(continued on next page)

**Table 4.12 (cont.)** Logistic regressions predicting part-time work at any time during the year in the TYP survey (J1 cohort).

Variable	Wave 5			Wave 6		Wave 8		Wave 9		
	Coef. <sup>a</sup>		S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	
In school <sup>g</sup>	-1.6	***	.48	-.76	.66	.39	.57			
School attachment * in school	.29	***	.10	.21	.13	-.12	.11			
School commitment (class rank) * in school	-.23	*	.13	.16	.39					
School commitment (effort) <sup>h</sup> * in school				-.13	.20	-.096	.085			
Allowance * in school	-.016 <sup>j</sup>		.014							
Cram school * in school	-.54	***	.17	-.72	**	.28	-.26	.21		
Student loans * in school						.91	***	.25		
School involvement * in school	.11 <sup>e</sup>	*	.058			.029	.022			
Class cohesion * in school	.20 <sup>e</sup>	*	.12			.040	.11			
School strain * in school	.034		.14	-.27	.19					
Perceived bad socioeconomic status	-.17	**	.084	.087	.093	-.027 <sup>e</sup>	.13			
Time spent online	.023		.058	.059	.087	.0047 <sup>i</sup>	.043	.089	.11	
Visiting cyber-cafes				-.16	*	.087				
Peer attachment						.24 <sup>e</sup>	.15			
Materialism								.14	.10	
Intercept	-1.37	***	.22	-1.57	1.25	2.2	*	1.2	-1.4	.95
N	1891			1552		1178		1508		

\* p < .10 \*\* p < .05 \*\*\* p < .01.

<sup>a</sup> Estimates are based on analyses for 33 multiply-imputed data sets, combined using Rubin's rules. Covariates are lagged by one wave unless otherwise indicated.

<sup>b</sup> Referent category: Taipei City.

<sup>c</sup> Denotes a contemporaneous measure.

<sup>d</sup> Referent category: not in school (i.e. either working full time, serving in the military, or staying at home – possibly studying to retake an entrance examination).

Note that educational tracking is not collinear with the “in school” variable because “in school” is lagged by one year.

<sup>e</sup> This variable is lagged by two years.

<sup>f</sup> This variable is lagged by three years.

<sup>g</sup> For variables measured in Waves 7 and later, “in school” refers only to respondents in college due to the questionnaire design. Also, all Wave 3 respondents were presumed to be in school, so this “interaction effect” is no different from a main effect for Wave 5 predictors lagged by two years.

<sup>h</sup> Wave 5 (predicting Wave 6 work): question about how hard respondents studied. Waves 7 and later: logged weekly hours spent doing homework or studying.

<sup>i</sup> For the model predicting work in Wave 8, the coefficients presented for this variable refer to that of an interaction with “in school” for Wave 6.

<sup>j</sup> Denotes a variable that is logged.

probability of responding to the survey. Respondents living in rural Yilan County are also less likely to be lost to follow-up, probably because they move less often.

The other component to the inverse probability weights is of course the selection models for work. Table 4.12 presents the findings from these models for the denominators in Equation 4.2. A trade-off had to be made in using lagged covariates to fit these models, as the inclusion of covariates lagged by more than one year inevitably reduces the sample size. Accordingly, a set of initial regressions based on the initial data set prior to performing multiple imputations progressively added lagged covariates to each model one wave at a time. F-tests then determined whether adding the new set of covariates explained significantly more variance in work than a model with them removed – and if not, those covariates were omitted from the model, increasing the size of the available sample. Accordingly, the model for Wave 8 has a smaller sample size than the model for Wave 9.<sup>116</sup> Unsurprisingly, one of the most important predictors of work is lagged work – except for college sophomores (Wave 8), for whom part-time work while seniors in high school is not very informative. Educational tracking is also extremely relevant to the decision to work: respondents not in school during Wave 6, moreover, are *27.7 times* as likely to be working as respondents attending a general-track school – suggesting both that the majority of respondents of high school age who are not in school are working. Additionally, respondents in vocational schools are 2.4 times as likely as their colleagues in general-track schools to have part-time jobs during 11<sup>th</sup> grade (J1W5), and 5.3 times as likely in 12<sup>th</sup> grade – suggesting that general-track students may be devoting more time to preparation for the college entrance examinations than their counterparts attending vocational schools, or that the vocational school students are integrating their work with their studies. The difference between these two tracks

---

<sup>116</sup> An alternative approach would have been to interact the lagged covariates with indicator variables for whether the respondent answered the survey for the wave corresponding to the lag. I decided not to take this approach in order for the interpretation of the probability weightings for unit nonresponse to be more straightforward.

disappears for college, however, suggesting that students attending institutes of science and technology are equally likely to focus on knowledge acquisition through study instead of work (or that general-track students begin their transition into work at this time). The effect of the college examinations on work is also evident through the coefficient for cram school, such that attending cram school in 11<sup>th</sup> grade is associated with a 51% reduction in the probability of working during 12<sup>th</sup> grade!

A few covariates have effects that seem to be age-graded. Neighborhood effects generally do not predict part-time work except for high school sophomores, for whom greater involvement in community activities may provide them with more social capital to find a job – these findings mirror Western findings that youth below the age of 15 find work through informal ties to friends (instead of advertisements), and are less likely to have a contract (Mortimer 2003). Additionally, respondents living with their parents are much less likely to work during their senior year of high school, but if they are still living with their parents when most respondents are freshmen in college, they are 62% *more* likely to work the next year. Unsurprisingly, student loans also matter for college sophomores – those who had a student loan as freshmen are 2.5 times as likely to be working the following year as those who did not. Stressful life events also predict work, but only for college students who may be choosing to work to support their families. And materialism turns out to not be a strong predictor of work for seniors in college, suggesting that perhaps the omission of this variable from the TYP survey prior to Wave 8 is not problematic for the present research. Finally, women are more likely to work part-time. One possible reason may be the requirement that all Taiwanese men serve in the military for one year around that time. As no women are required to do so, military service could not be included in the selection model to preserve the positivity assumption for inverse probability weighting – and so the effects of

military service on work are likely observed in the model through the coefficient for gender. Finally, respondents in rural Yilan County are less likely to work, probably because there are fewer available jobs for youth compared to Taipei.

With the selection and nonresponse models fit, the stage is set for a test of Hypothesis 4.3. Table 4.13 presents inverse Gaussian and negative binomial counterfactual models of the effect of part-time work during the year on self-reported deviance for Waves 5, 6, 8, and 9, with Huber-White robust standard errors. 27.9% of Wave 5 respondents, 17.2% of Wave 6 respondents, 77.1% of Wave 8 respondents, and 72.3% of Wave 9 respondents stated that they had worked part-time at some point during the previous year. The results of the counterfactual model appear to depend on model specification; the negative binomial model suggests that working part-time during the year causes a 6.6% increase in self-reported deviance as measured by a latent IRT variable, while the inverse Gaussian model indicates no significant effect of work on deviance. With conflicting findings from different models, a reasonable approach is to use the Bayesian Information Criterion (BIC). A fair way to perform the test for this model and for each of the models that follow is to fit both the inverse Gaussian model and the negative binomial model to the transformed count variable used for the negative binomial regressions. The test can only currently be performed on one multiply-imputed data set at a time, so I perform the BIC on a single data set and the results are convincing (Table 4.13): with the lower BIC value, the negative binomial model is the better fit to the data. Thus, Hypothesis 4.3 is not supported, and part-time work causes an increase in deviance of about 6.6% ( $e^{0.064}$ ).

**Table 4.13.** Counterfactual inverse Gaussian and negative binomial regressions of delinquency on work during the year for respondents to Waves 5, 6, 8, and 9, using multiple imputation methods to address item non-response and Inverse Probability Weighting methods to handle unit non-response and selection into work, and controlling for all time-varying and time-invariant predictors of theoretical interest measured in the same way in Waves 4, 5, 7, and 8. Coefficients and standard errors are estimates from 33 separate regressions, one for each multiply-imputed data set, combined using Rubin’s rules. Standard errors are calculated using clustering on individual respondents to compensate for autocorrelation across waves.

	Inverse Gaussian Model		Negative Binomial Model	
	coef.	SE†	Coef.	SE†
<b>Work During the Year</b>	<b>.109</b>	<b>.164</b>	<b>.064 ***</b>	<b>.032</b>
Wave 6	-.15 ***	.028	-.16 ***	.0081
Wave 8	-2.80 ***	.17	-2.15 ***	.068
Wave 9	-.47 ***	.11	-.33 ***	.025
Taipei County <sup>a</sup>	-.072	.15	-.030	.031
Yilan County <sup>a</sup>	-.26 *	.15	-.063 *	.035
Female	-1.0 ***	.15	-.29 ***	.024
Birth year	.23 ***	.087	.052 **	.022
Mother’s yrs. of education	.0076	.024	-.0013	.0056
Father’s yrs. of education	.021	.020	.0038	.0055
Vocational track <sup>a</sup>	-.053	.13	-.018	.026
Not in school (track) <sup>a</sup>	-.14	.20	.023	.066
Lagged parental monitoring	-.15 ***	.059	-.049 ***	.014
Lagged living with parents	-.15	.14	-.041	.037
Lagged household income (logged)	.060 **	.030	.0077	.014
Intercept	2.60 ***	.43	8.96 ***	.12
Multiple Imputations	33		33	
N	1,891		1,891	
Person-years	5,293		5,293	
BIC	131,988		93,077	

\*\*\* p<.01 \*\*p<.05 \* p < .10 † Outer Product of the Gradient.

<sup>a</sup> Referent category: Taipei City

<sup>b</sup> Referent category: general track

Next, Tables 4.14 and 4.15 present the causal models for work during the academic year and during the summer (the selection models are quite similar to Table 4.12, and are omitted for

these variables).<sup>117,118</sup> The models find no causal effect of work on deviance – either during the school year or during the summer – regardless of the functional form of the statistical model. These covariates are slightly more prone to measurement error than working during the year, due to the need to reconstruct the Wave 8 work variables using information from Wave 9 (see Appendix D). Just as with the causal model for work at any time during the year, these analyses include multiple imputations to address item non-response, and inverse probability weighting methods to compensate for both selection into work and attrition from survey waves. BICs are computed using a single multiply-imputed data set, with both the inverse Gaussian and negative binomial regressions fit to the same outcome variable: the rescaled delinquency items after they have been transformed to count variables. In all cases, the negative binomial model is a better fit to the data.

---

<sup>117</sup> The same covariates for the selection models for work during the year are used to model selection into work during the academic year and during breaks, as the present review could not locate any literature that was so specific as to suggest that some confounders would affect work during the academic year and not the summer, or vice versa. Note that only respondents currently in school answered questions about whether they worked during the academic year or during breaks. Therefore, in using lagged work to predict selection into work during the next year, the treatment is interacted with an indicator variable for being in school, and the indicator variable and interaction effect are included as predictors without a main effect for lagged work. The interpretation of the interaction effect for work during the school year with attending school, for example, then becomes the effect of work during the school year for those who were in school during the previous year, with no effect observed for those who were not in school.

<sup>118</sup> Note that for Wave 6, respondents were asked in June at the end of their senior years if they currently were working part-time, so this question is used both as a substitute for working during the academic year and for working during the current year. (The question is included in the analysis despite the measurement error because of the importance of measuring work during the year that respondents took the college entrance examinations.) Accordingly, the model for work during the academic year covers Waves 5, 6, 8, and 9, while the model for work during the summer or winter breaks covers Waves 5, 8, and 9 only.

**Table 4.14.** Counterfactual Inverse-Gaussian and negative binomial regressions of delinquency on work during the academic year for respondents to Waves 5, 6, 8, and 9, using multiple imputation methods to address item non-response and Inverse Probability Weighting methods to handle unit non-response and selection into work, and controlling for all time-varying and time-invariant predictors of theoretical interest measured in the same way in Waves 4, 5, 7, and 8. Coefficients and standard errors are estimates from 33 separate regressions, one for each multiply-imputed data set, combined using Rubin’s rules. Standard errors are calculated using clustering on individual respondents to compensate for autocorrelation across waves.

	Inverse Gaussian Model		Negative Binomial Model	
	coef.	SE†	Coef.	SE†
<b>Work During School</b>	<b>.121</b>	<b>.128</b>	<b>.054</b>	<b>.040</b>
Wave 6	-.17	*** .013	-.17	*** .0058
Wave 8	-2.73	*** .15	-2.14	*** .075
Wave 9	-.47	*** .10	-.32	*** .028
Taipei County <sup>a</sup>	-.17	.14	-.040	.030
Yilan County <sup>a</sup>	-.31	** .13	-.074	** .033
Female	-.91	*** .13	-.26	*** .022
Birth year	.20	** .088	.056	** .022
Mother’s yrs. of education	-.0062	.022	-.0014	.0051
Father’s yrs. of education	.024	.019	.0033	.0048
Vocational track <sup>a</sup>	-.12	.12	-.030	.026
Lagged parental monitoring	-.12	** .053	-.054	*** .014
Lagged living with parents	-.043	.13	-.034	.038
Lagged household income (logged)	.032	.047	-.0022	.019
Intercept	2.59	*** .38	9.01	*** .13
Multiple Imputations	33		33	
N	1,827		1,827	
Person-years	4,730		4,730	
BIC	117,262		82,071	

\*\*\* p<.01 \*\*p<.05 \* p < .10 † Outer Product of the Gradient.

<sup>a</sup> Referent category: Taipei City

<sup>b</sup> Referent category: general track

**Table 4.15.** Counterfactual inverse Gaussian and negative binomial regressions of delinquency on work during the summer or winter breaks for respondents to Waves 5, 8, and 9, using multiple imputation methods to address item non-response and Inverse Probability Weighting methods to handle unit non-response and selection into work, and controlling for all time-varying and time-invariant predictors of theoretical interest measured in the same way in Waves 4, 7, and 8. Coefficients and standard errors are estimates from 33 separate regressions, one for each multiply-imputed data set, combined using Rubin’s rules. Standard errors are calculated using clustering on individual respondents to compensate for autocorrelation across waves.

	Inverse Gaussian Model		Negative Binomial Model	
	coef.	SE†	Coef.	SE†
<b>Work During Breaks</b>	<b>.040</b>	<b>.13</b>	<b>.069‡</b>	<b>.042</b>
Wave 8	-2.77 ***	.14	-2.2 ***	.071
Wave 9	-.45 ***	.098	-.34 ***	.030
Taipei County <sup>a</sup>	-.18	.16	-.043	.041
Yilan County <sup>a</sup>	-.35 **	.16	-.095 **	.023
Female	-.99 ***	.14	-.34 ***	.031
Birth year	.20 *	.11	.069 **	.030
Mother’s yrs. of education	.0083	.026	.0028	.0073
Father’s yrs. of education	.038 *	.023	.0052	.0067
Vocational track <sup>a</sup>	-.064	.13	-.023	.035
Lagged parental monitoring	-.16 ***	.062	-.053 ***	.018
Lagged living with parents	.024	.14	-.066	.047
Lagged household income (logged)	.033 **	.043	.0026	.016
Intercept	2.51 ***	.43	8.97 ***	.14
Multiple Imputations	33		33	
N	1,827		1,827	
Person-years	3,312		3,312	
BIC	80,782		57,655	

\*\*\* p<.01 \*\*p<.05 \* p < .10 † Outer Product of the Gradient. ‡ p = .103

<sup>a</sup> Referent category: Taipei City

<sup>b</sup> Referent category: general track

Hypothesis 4.2 suggests differential effects of work on deviance depending on whether the job takes place during the academic year or during breaks. A casual inspection of Tables 4.14 and 4.15 suggests that the effects are likely to not be different if they are both null, but a formal test of this hypothesis is still necessary to be sure. The test requires a model that includes both

work during the academic year and work during vacations in a single causal model of deviance, such that both work effects can be interpreted as causal. To create such a model, I use weights for each observation corresponding to the product of the inverse probability weights for work during the summer, work during the academic year, and unit nonresponse; such an approach accounts for selection into both types of work along with bias due to sample attrition. Then, a test of whether the coefficients for holiday and academic-year work differ corresponds to a test of whether the two causal effects differ in magnitude. Table 4.16 shows the result of this regression.

The `mi testtransform` function in STATA (StataCorp. 2013) compensates for the fact that the coefficients are calculated using Rubin's rules, and the tests for each model in Table 4.16 fail to reject a null hypothesis that the academic-year and holiday break coefficients for work are the same for both the inverse Gaussian model ( $p = .94$ ) and the negative binomial model ( $p = .42$ ). The two models therefore yield causal effects of work during the academic year and work during the summer that are not significantly different from zero and not significantly different from each other, so Hypothesis 4.2 is not supported.

Next, Hypothesis 4.4 posits that work causes an increase in deviance for those with low social control and no effect for those with high social control, and is operationalized by an interaction between parental monitoring and work. Table 4.17 presents the results of such a test, using work at any time during the year as the treatment, as it is better measured than work during holiday breaks or during the academic year for respondents in school.

**Table 4.16.** Counterfactual Inverse-Gaussian and negative binomial regressions of delinquency on work during the academic year and on work during the summer or winter breaks for respondents to Waves 5, 8, and 9, using multiple imputation methods to address item non-response and inverse probability weighting methods to handle unit non-response and selection into work. The model also controls for all time-varying and time-invariant predictors of theoretical interest measured in the same way in Waves 4, 7, and 8. Coefficients and standard errors are estimates from 33 separate regressions, one for each multiply-imputed data set, combined using Rubin’s rules. Standard errors are calculated using clustering on individual respondents to compensate for autocorrelation across waves.

	Inverse Gaussian Model		Negative Binomial Model	
	coef.	SE†	Coef.	SE†
<b>Work During School</b>	<b>-.081</b>	<b>.13</b>	<b>.020</b>	<b>.051</b>
<b>Work During Breaks</b>	.093	.15	<b>.026</b>	<b>.045</b>
Wave 8	-2.79 ***	.16	-2.2 ***	.081
Wave 9	-.43 ***	.12	-.33 ***	.037
Taipei County <sup>a</sup>	-.20	.17	-.034	.042
Yilan County <sup>a</sup>	-.40 **	.17	-.087 **	.046
Female	-1.01 ***	.14	-.34 ***	.032
Birth year	.21 *	.12	.079 **	.032
Mother’s yrs. of education	-.0050	.026	.0018	.0070
Father’s yrs. of education	.037	.023	.0038	.0072
Vocational track <sup>a</sup>	-.14	.14	-.045	.037
Lagged parental monitoring	-.11 *	.063	-.065 ***	.021
Lagged living with parents	-.035	.15	-.041	.048
Lagged household income (logged)	.038	.041	-.0025	.020
Intercept	2.60 ***	.43	9.04 ***	.17
Multiple Imputations	33		33	
N	1,827		1,827	
Person-years	3,312		3,312	

\*\*\* p<.01 \*\*p<.05 \* p < .10 † Outer Product of the Gradient. ‡ p = .103

<sup>a</sup> Referent category: Taipei City

<sup>b</sup> Referent category: general track

**Table 4.17.** Counterfactual inverse Gaussian and negative binomial regressions of delinquency on work during the year for respondents to Waves 5, 6, 8, and 9, with an interaction between parental monitoring and work. The models use multiple imputation methods to address item non-response and inverse probability weighting methods to handle unit non-response and selection into work, and control for all time-varying and time-invariant predictors of theoretical interest measured in the same way in Waves 4, 5, 7, and 8. Coefficients and standard errors are estimates from 33 separate regressions, one for each multiply-imputed data set, combined using Rubin's rules. Standard errors are calculated using clustering on individual respondents to compensate for autocorrelation across waves.

	Inverse Gaussian Model			Negative Binomial Model		
	coef.		SE†	Coef.		SE†
<b>Work During the Year</b>	<b>1.22</b>	<b>***</b>	<b>.47</b>	<b>.316</b>	<b>***</b>	<b>.117</b>
Lagged parental monitoring	.033		.087	-.022		.014
Lagged parental monitoring *	-.28	**	.12	-.068	**	.029
Wave 6	-.17	***	.032	-.16	***	.0082
Wave 8	-2.83	***	.16	-2.15	***	.068
Wave 9	-.44	***	.10	-.35	***	.027
Taipei County <sup>a</sup>	-.16		.15	-.031		.031
Yilan County <sup>a</sup>	-.33	**	.15	-.063	*	.035
Female	-1.07	***	.16	-.30	***	.024
Birth year	.25	***	.091	.054	**	.022
Mother's yrs. of education	.0015		.024	-.0016		.0056
Father's yrs. of education	.023		.020	.0039		.0054
Vocational track <sup>a</sup>	-.097		.14	-.018		.027
Not in school (track) <sup>a</sup>	-.19		.21	.022		.066
Lagged living with parents	-.12		.14	-.043		.037
Lagged household income (logged)	.056	*	.032	.0067		.014
Intercept	1.99	***	.43	8.86	***	.12
Multiple Imputations	33			33		
N	1,891			1,891		
Person-years	5,293			5,293		
BIC	131,884			93,071		

\*\*\* p<.01 \*\*p<.05 \* p < .10 † Outer Product of the Gradient.

<sup>a</sup> Referent category: Taipei City

<sup>b</sup> Referent category: general track

The results in Table 4.17 are compelling to say the least. Comparing Table 4.13 to Table 4.17, the effect of parental monitoring on reducing crime in the model without an interaction effect is completely absorbed by the interaction term, such that the main effect for parental monitoring in Table 4.17 is not associated with deviance. Appendix E describes how parental monitoring is measured: the question is essentially Hirschi's measure of parental monitoring (1969), a Likert-scale question concerning whether a respondent agrees with a statement that his or her parents know where he or she is every day, averaged for both parents and ranging from 1 ("very much does not apply to me") to 5 ("very much applies to me.") Thus, in the negative binomial model above (which fits the data better than the inverse Gaussian model according to the BIC statistic), respondents who have no part-time job, and who describe their parents as "very much" not knowing their daily whereabouts, would commit 28% more crime ( $e^{.316-.068} = 1.28$ ) were they to get a part-time job. But for those respondents with high levels of parental monitoring, those with no part-time job who describe their parents as "very much" *knowing* their daily whereabouts would actually commit 2.4% *less* crime were they to get a job ( $e^{.316-5(.068)} = .976$ ) – a change that is statistically no different from zero given the standard error for the main effect of work. In other words, part-time work only raises a young person's level of deviance in the absence of social control as measured by parental monitoring – and in the presence of such social control, it is not harmful at all. The lesson for Taiwanese parents is thus that part-time work does not raise crime, provided that children are aware that their parent knows where they are and what they are doing. Moreover, the main effect of parental monitoring in Table 4.17 is not significant, suggesting that perhaps the entire role of parental monitoring in explaining deviance is accounted for by the fact that the parent is aware of where his or her child is working.

This last inference, however, refers to a causal mechanism that is beyond the scope of the present research to explore.

#### *Analysis of the Effect of Work Intensity and Extensity on Deviance*

The significant interaction between work and parental monitoring provides evidence in support of social control in Taiwan, and as such seems to support the hypothesis that the positive association between work and deviance is part of a syndrome of precocious maturation (Newcomb and Bentler 1988), in the sense that Taiwanese youth who take on adult roles without supervision by their parents then proceed to engage in delinquent behavior. Such a conclusion seems to suggest that delinquency is the product of a person's circumstances. As mentioned previously, however, Mortimer (2003) makes a compelling case that American youth who have a plan for their lives and act on it tend to be steady workers – that is, they work for relatively few hours per week, and for more weeks per academic year – and this pattern of work results in more positive outcomes. A replication of Mortimer's typology of work with the Denver Youth Survey (Callahan 2010) also found that steady work reduced general delinquency compared to high-intensity work (for more than 20 hours per week) or "occasional" work for (low-intensity work for relatively few weeks per academic year). Perhaps, then, individual agency also matters in Taiwan, and the previous analyses simply have not uncovered it because they focused on work as a dichotomous variable, rather than exploring the intensity and extensity of work. This section tests Hypothesis 4.5, namely that "steady" work, following Mortimer's work typology, is associated with lower delinquency for the respondents to the Taiwan Youth Project survey. This hypothesis test should shed light on the degree to which Taiwanese youth, not just their parents, can lower their propensities to offend by means of proactive behavior.

In the Taiwan Youth Project survey, questions about work intensity and extensity are not available until respondents are in college. The present analysis therefore is limited to Waves 8 and 9, when respondents are in their sophomore and junior years and such questions are available in the survey at the same time as the delinquency items. For these waves, respondents were asked to describe several jobs that they had held, the month and year in which they started and stopped each job, and how many hours per week they had worked. This approach enabled the creation of measures for work intensity and extensity during an entire 12-month period before respondents completed each survey, during the previous academic year, and during the previous summer and winter breaks.

Unfortunately, the way in which these survey questions were asked may have generated substantial measurement error. For many respondents, the Wave 8 questions did not ask about enough different jobs to truly capture how long and hard they had worked, and so Wave 9, administered two years after Wave 8, included additional questions asking respondents to provide more detail about jobs that they had held during the last three years. Accordingly, calculating work intensity and extensity was possible for the present research, but required the design and implementation of a complex data filter described in Appendix D. The appendix describes the filter in detail so that the present research can be replicated, and in the process quantifies the measurement error involved to the extent possible.

Given these caveats about the data quality, the next step in the present research is to determine an appropriate threshold to use for work intensity and extensity. To distinguish between “high” and “low” work extensity, Mortimer (2003) captures work typologies for those who work as a categorical variable with four levels, with work intensity and extensity each bifurcated into “high” and “low” categories. She chooses as thresholds delimiting each category

the median number of months that 15-17-year-olds worked during the academic year; for intensity, she and her colleagues (Mortimer 2003; Mortimer et al. 1996; Mortimer and Johnson 1998) select 20 hours of work per week as a threshold value based on convention among researchers in the U.S.<sup>119</sup> In Taiwan, few studies have looked at time-use patterns among students working part-time; one recent study (Huang 2006b) that draws on a cross-sectional sample of more than 30,000 Taiwanese college students suggests that the relationship between work hours and other types of time use such as homework, studying ahead, exercise, sleep, talking with friends, and dating qualitatively changes once students work more than either 10 or 15 hours per week, depending on the particular type of time use. Huang's study does not touch upon deviance, however, and as a time allocation study contains no information about the relationship between how many weeks students continuously work at the same job and various outcomes such as scholastic achievement.

Accordingly, it makes sense to determine threshold points corresponding to Mortimer's typology inductively, using the TYP data itself.<sup>120</sup> I therefore conduct a series of nonparametric locally-weighted robust least squares regressions of delinquency on weeks worked and hours worked per week during the academic year, then plot them in three dimensions (see Fox 2002). Nonparametric regression allows a researcher to view the shape of the data for hours worked per week and weeks worked per year, and its relationship to deviance. The procedure involves, for each data point, computing the 50% of *all* observations that have values for weeks worked per year and hours worked per week that are closest to the corresponding values for that data point (distances calculated to determine this set of observations are Euclidean). These subsets of the data are then used in robust regressions of the IRT measures for crime on hours worked per week

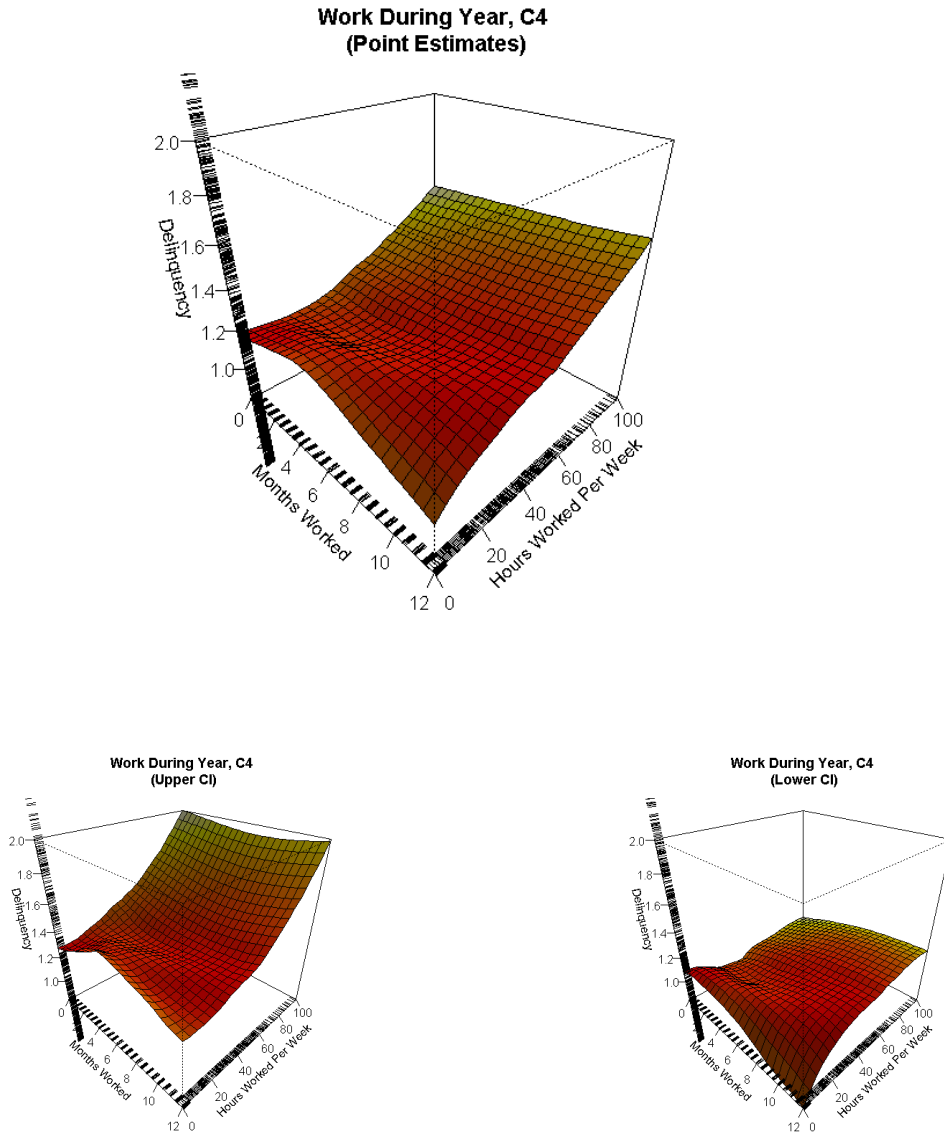
---

<sup>119</sup> Mortimer states that "these break points are somewhat arbitrary, but they offer convenient and intuitively meaningful ways of characterizing youth's cumulative work experience," (2003: pp. 71-73).

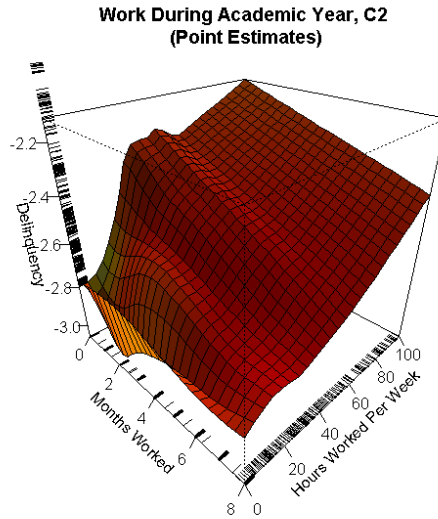
<sup>120</sup> The methodological description in this paragraph is drawn from Callahan (2010:35).

and weeks worked per year, one for each data point, using a Huber biweight weighting function. A smoothing algorithm then generates a surface that shows the slopes of the regression at each point, and a heat schematic colors the surface such that darker areas indicate narrow confidence intervals for predicted values of deviance and lighter areas indicate wide confidence intervals. The benefit of the nonparametric model is that peaks and valleys in the resulting surfaces yield insight into nonlinear relationships and places where threshold values might reasonably be drawn. Further, because educational tracking has such a strong relationship to deviance in the present study, after examining models for work during the full year, the academic year, and the summer and winter breaks, I then further stratify the models for work during the academic year by educational tracking to investigate whether tracking may affect the thresholds (it appears not to).

Figure 4.2 illustrates the nonlinear relationship between intensity, duration, and deviance for respondents who worked at any time during the year during Wave 9, along with surfaces corresponding to the corresponding upper and lower confidence intervals for the nonparametric regressions. The surface corresponding to the point estimates provides graphical evidence supporting Mortimer's finding: respondents with low-intensity, high-duration jobs appear to express the least deviant behavior among those respondents who have jobs, so perhaps Taiwanese youth who express agency are indeed less deviant.



**Figure 4.2.** Nonparametric locally-weighted robust least squares regressions of deviance on work intensity and extensity among respondents to Wave 9 of the Taiwan Youth Project survey. Responses for Wave 9 deviance are on the J3 Wave 6 scale as per Chapter 2, but have not yet been further transformed to produce a scale with all positive values. The densities of the rug plots on each axis correspond to the number of respondents providing the corresponding answers for each question.



**Figure 4.3.** Nonparametric locally-weighted robust least squares regressions of deviance on work intensity and extensity during the academic year among respondents to Wave 8 of the Taiwan Youth Project survey. Responses for Wave 8 deviance are on the J3 Wave 6 scale as per Chapter 2, but have not yet been further transformed to produce a scale with all positive values. The densities of the rug plots on each axis correspond to the number of respondents providing the corresponding answers for each question. The scale for deviance has been compressed (compared to Figure 4.2) to emphasize change in the slope of the surface.

Because of the imprecision of the TYP responses concerning work duration, most of the respondents who worked steadily during the academic year were coded as working for all eight months. Based on both the shape of the lowess graphs and the fact that so many respondents worked for all eight months, then, the present research uses six months as the threshold for work duration and 20 hours per week as the threshold for work intensity.<sup>121</sup> Table 4.18 then shows the results of negative binomial regressions for each wave of deviance on the Mortimer work categories (using deviance measures that have been transformed to be positive and, for the negative binomial models, transformed into count variables), controlling for gender, educational

<sup>121</sup> To be precise, respondents working 20 or fewer hours are coded as having low-intensity work, and respondents working for six or more months during the academic year are coded as having high-extensity work.

tracking, and lagged parental monitoring.<sup>122</sup> As in the previous analysis, a BIC statistic compares the two models, and is calculated based on both models run using the same dependent variables, the IRT measures after they have been converted to counts.

**Table 4.18.** Negative binomial regressions of delinquency on Mortimer’s work typologies during the year for Waves 8 and 9, with robust standard errors. Coefficients and standard errors are estimates from 33 separate regressions, one for each multiply-imputed data set combined using Rubin’s rules.

	Wave 8 Deviance		Wave 9 Deviance			
	coef.	SE	Coef.	SE		
<b>No work<sup>a</sup></b>	<b>-.29</b>	<b>*</b>	<b>.15</b>	<b>-.18</b>	<b>***</b>	<b>.066</b>
<b>Occasional work<sup>a</sup></b>	<b>-.24</b>		<b>.22</b>	<b>.081</b>		<b>.095</b>
<b>Sporadic work<sup>a</sup></b>	<b>.11</b>		<b>.20</b>	<b>-.0052</b>		<b>.11</b>
<b>Steady work<sup>a</sup></b>	<b>-.027</b>		<b>.20</b>	<b>-.098</b>		<b>.074</b>
Female	-1.05	***	.11	-.70	***	.048
Vocational track <sup>b</sup>	-.35	***	.12	-.045		.050
Lagged parental	-.23	***	.048	-.13	***	.031
Intercept	8.17	***	.20	7.92	***	.11
Multiple Imputations	33		33			
N	1,080		1,152			

\*\*\* p<.01 \*\*p<.05 \* p < .10

<sup>a</sup> Referent category: Most invested

<sup>b</sup> Referent category: general track

<sup>c</sup> Using Wave 6 parental monitoring for the Wave 8 regression, and Wave 8 monitoring for the Wave 9 regression.

Figure 4.2 provides no evidence that steady work is associated with less crime than other forms of work. Indeed, it simply reproduces the findings that youth who work part-time are more delinquent than those who do not. The findings for this section suggest that perhaps findings from the West that youth who are playful are less delinquent do not hold for Taiwanese youth, at least when it comes to part-time work. Additional measurement error associated with the Wave 8

<sup>122</sup> As was the case for the previous analyses in this chapter, both negative binomial and inverse Gaussian models were fit, and the findings were substantively the same. The negative binomial models were a better fit to the data as indicated by BIC statistics calculated using the same procedure as in the previous analyses, so those models are presented here.

observations, however, likely explains why the standard errors are larger in the model for Wave 8 than in the Wave 9 model. A more conclusive test in Taiwan of Mortimer's rather nuanced hypothesis requires future research with better measures of work intensity and extensity.

### **Discussion and Conclusion**

This chapter tested the causal effect of youth employment on delinquency in Taiwan. Three major hypotheses in the Western literature characterize the relationship between employment and crime as a “zero-sum” relationship in which work leads to lower grades in school and a greater proclivity to take the easy way out by offending, an “agentic” model in which those who plan ahead receive good jobs that reinforce norms for conventional behavior and lower their propensity to offend, and a “syndrome of precocious maturation” in which youth who are too eager to become adults take on work roles at an early age and also commit crimes. The research presented in this chapter finds a positive relationship between work and deviance among Taiwanese youth of high school and college age, and some support for a causal effect of work on deviance. The most interesting findings indicate a strong causal effect of work on deviance after it is interacted with parental monitoring, in which case work causes deviance for youth whose parents do not keep track of their whereabouts, yet does not lead to delinquent behavior for youth who are closely monitored by their parents. The statistical significance of this finding for work carries with it the benefit of demonstrating that the Bayesian Item Response Theory model and associated vertical scaling methods presented in Chapter 2 introduce relatively little measurement error. Substantively, these findings also suggest that deviance does not cause crime via a “zero-sum” relationship for Taiwanese youth, but instead can lead unsupervised youth to crime, perhaps by exposing them to settings in which they have the opportunity to offend. Hence,

work for the unsupervised can be part of a syndrome of precocious maturation. The message for Taiwanese parents appears to be that work is not inherently problematic for their children, as long as they keep track of where their children are when they are away from home.

This chapter finds no support for a hypothesis that Taiwanese youth following work patterns associated in the United States with agency (Mortimer 2003) – that is, so-called “steady workers” who work only a few hours per week but for long periods of time – are less likely to be involved in deviant behavior than youth with other work patterns. Graphical models generated by nonparametric regressions suggest a slight trend in the data indicating a relationship between “steady work” and lower deviance, but the relationship is not strong enough to remain significant in negative binomial regression models of deviance on work patterns. These results differ from findings in the West, in which youth with no job and youth with steady jobs commit crimes at approximately the same rate, which is lower than that for other use (Callahan 2010). Due to problems with the quality of some retrospective survey questions in the TYP concerning work intensity and extensity for jobs that took place up to three years in the past, these findings should be replicated in the future with better measures of work intensity and duration.

In summary, this chapter provides evidence that working, especially in the absence of parental monitoring, is criminogenic for Taiwanese youth, and no evidence that steady part-time work is related to desistence in Taiwan. This research is designed primarily to help people to understand what would happen to deviance if a youth’s involvement in part-time work were to be changed by a factor, such as a new policy, that is unrelated to the youth’s personal characteristics and social context. Accordingly, the research does not constitute a test of a particular theory, and perhaps for this reason raises more theoretical questions than it answers.

The difference between the criminogenic effect of work on deviance in Taiwan and the mild protective effect in the United States, for example, could be explained by several competing theories. First, consider social learning theory, which draws on differential association theory's premise (Sutherland 1947) that people internalize a set of informal "definitions" either favorable or unfavorable to crime that they consult when presented with opportunities to offend, and which change as they interact with others. For youth of high school and college age, work constitutes a new social environment in which they may learn new "definitions" from their co-workers. Taiwanese youth, by virtue of being raised in a system that emphasizes conformity and subordination of one's own will to those with authority including parents and teachers, simply may have learned a set of baseline definitions that are less favorable to crime than their Western counterparts, especially those Westerners in urban areas with concentrated poverty and a disproportionate number of broken families. A part-time job waiting tables may therefore indicate a step towards conventional behavior and away from crime for many Western youth, but the same job for the Taiwanese student may indicate a step towards less conventional behavior and more short-term benefit than staying focused on academic success. In gaining a new peer network at the workplace with views focused more on immediate gratification, that student may end up internalizing a new set of definitions more favorable to crime. Thus, the same job that lowers deviance in the West may increase it in Taiwan, even if social learning theory explains deviance equally well in both places.

Alternatively, control theory (Hirschi 1969) could drive the differences between Taiwan and the West in the effect of work on deviance if "baseline" social control is higher in Taiwan than in the West. In particular, if spending more time at work is associated with a weakening of social bonds to school and the family, but work itself produces new social bonds to co-workers that are

less important than school and family ties to Taiwanese youth but more important to American youth, then the same job in each society could potentially lower the total effect of social control on the individual in Taiwan and raise it in the West. Such a scenario is plausible given that in Taiwan, social ties to family members, classmates, and teachers are reinforced by traditional Confucian views about the importance of teachers and of filial piety and, for school, by long hours spent in conventional behavior (studying and attending class) with a small group of peers. In a similar vein, following Institutional Anomie Theory (IAT), the fact that youth employment is protective in the U.S. but harmful in Taiwan could be explained by the fact that Americans are motivated by a cultural pressure to get a good job and less influenced by a sense of obligation to their families, communities, and religion (Messner and Rosenfeld 2007), while Taiwanese are relatively more influenced by these social obligations and less by cultural norms rewarding success in the workplace.

Finally, following Sampson and Laub's age-graded theory of informal social control (1993), the precise timing of turning points during the life course may matter as much in changing behavior as what the turning points actually are. Following this line of reasoning, the positive effects of work in lowering deviance could very much exist in Taiwan, but take place later than they do in the West. Recent years have seen Taiwanese youth, especially those living in Taipei, living with their parents well into their late twenties and even thirties due to the high cost of housing, and this trend could partially explain why early work at a time when most youth are still going to school may lead to low-level deviance for unsupervised youth.

Untangling these causal mechanisms and the theories that may drive them is beyond the scope of the present research. Furthermore, while the above discussion is intended simply to highlight the fact that the findings do not help to prove or disprove any criminological theory, it

also brings to light the danger of undirected positivism: it is always possible to reconcile differing findings related to two variables by arguing that they are associated with differences in some third variable. Such an approach can lead to repeated bifurcations of a researcher's analytical framework *ad absurdum*. This research is particularly useful for those who would need to make counterfactual inferences concerning work and deviance for Taiwanese youth, and is not intended to pixelate the way sociologists approach tests of criminological theory. At the same time, it does underscore the utility of further work to investigate the quality of the jobs that Taiwanese youth are pursuing part-time – above and beyond simple measures related to work intensity and extensity – as well as the behavior of co-workers and social norms about deviance encountered at the workplace that could shed light on the criminological theories that best explain the increase in deviance associated with part-time work among unsupervised youth.

All told, this chapter finds that in Taiwan's social context comprised of relatively low deviance and close-knit social ties in school and the family, the most positive outcomes for youth appear to be associated with conformity. In doing so it sheds light on the importance of the high level of social control in Chinese culture, a context that appears connected to the way that turning points influence deviance. Perhaps a look at a second turning point, marriage, may shed further light on how turning points influence delinquency within Taiwan's cultural context. The next chapter therefore investigates the effect of marriage and steady romantic relationships on deviance among Taiwanese youth.

## **Chapter 5: Modeling the Effect of Steady Romantic Relationships and Marriage on Deviance in Taiwan<sup>123</sup>**

The objective of this chapter is to further our understanding of the effect on deviant behavior in Taiwan of another turning point in the life course, marriage. According to Sampson and Laub's age-graded theory of social control (1993), marriage can introduce a person to a new set of social bonds that steer that person away from deviance. In an application of marginal structural models to the study of deviance, Sampson, Laub, and Wimer (2006) found that marriage lowers crime by about 40% for men in the United States, basing their conclusions on an application of marginal structural models using IPTW weighting methods to the life histories of 500 boys born between 1924 and 1932 and re-surveyed in 1994 between the ages of 61 and 70. The present chapter seeks, to the extent possible, to replicate their methods and findings using data from Taiwan.

As often happens with empirical research, limitations of the current survey design, along with recent demographic changes in Taiwan, prevent a direct replication of Sampson, Laub, and Wimer's analyses. The present analysis is instead restricted to an investigation of the effect of early marriage on smoking, and of the effect on deviance of the intent to marry among those who are still in their early twenties. In recent years, Taiwan has experienced a significant increase in the age of first marriage, for both men and women. For Taiwanese born between 1961 and 1965, 33% of men and 63% of women were married by age 25. That number dropped to about 14% and 22%, respectively, for Taiwanese born between 1976 and 1980 – with about half the decrease occurring during the five years accounted for by the difference in age at first marriage

---

<sup>123</sup> The author is grateful to Dr. Lin Wen-Hsu, a postdoctoral student in the Institute of Sociology, Academia Sinica, for assistance with data preparation for the causal models presented in this chapter.

for the 1971-76 cohorts and the 1976-1980 cohorts, indicating a recent acceleration in the trend towards later marriage (Chen and Chen 2014). As a result of these changes, by Wave 7 of the J3 cohort and Wave 9 of the J1 cohort (the last survey years for which respondents were interviewed about delinquency), only 59 out of 3,745 survey respondents, or 1.6% of the sample, were married – not enough people to produce meaningful results for the present research. The average age for these respondents was 22.

In 2011, however, the first wave of the adult version of the Taiwan Youth Project survey was administered to both cohorts, with both groups responding to the same survey questions. The 2011 survey took place three years after the J3 Wave 7 survey, two years after the J3 Wave 8 survey (a telephone survey that did not include delinquency questions), and one year after the J1 Wave 9 survey. By 2011 more than 200 respondents were married, or about 7% of the sample. The models predicting survey attrition in Chapter 4 indicated that high grades are associated with a lower probability of attrition in the TYP survey – so it may be that respondents with low grades, who are more likely to marry early, are precisely those who may be lost to follow-up. Hence, the low prevalence of marriage among TYP respondents may be suggestive both of a continuing trend towards delayed marriage in Taiwan and of the possible loss to follow-up of some married couples.

An analysis of the effect of marriage on crime therefore appears to be feasible, although survey reweighting to account for unit nonresponse appears essential.<sup>124</sup> Unfortunately, the only types of deviance questions included in the 2011 TYP survey were smoking and drinking, and the exclusion of other deviance items from the survey was likely due to the very difficulty, addressed in Chapter 2, of modeling deviance with the survey items changing each year. That

---

<sup>124</sup> As a reminder, unit nonresponse refers to a respondent not responding to an entire wave of a survey, while item nonresponse refers to survey questions to which a respondent does not respond despite answering other questions in the same survey year.

said, for smoking only, respondents to the 2011 survey were asked to recall up to ten years in the past, and to state for each of these years whether they had smoked and if so, if they had smoked five or fewer packs or more than five packs during the year. Accordingly, for the present chapter models of the causal effect of marriage on smoking are possible using the retrospective survey questions about smoking, but the effect of marriage on general delinquency cannot be studied.

The effect of *intent* to marry on general deviance, however, is possible to model for survey waves prior to the 2011 survey. TYP respondents were asked both questions about their involvement in deviant behavior and if they intended to marry their current partner for Waves 6 and 7 for the J3 cohort and Waves 8 and 9 for the J1 cohort. If the strength of the social bond between a Taiwanese person and his or her partner is the mechanism that leads to desistence, then it may be the case that desistence both from smoking and from general deviance begins even prior to marriage as a person aspires to become a more suitable spouse for his or her current partner. Hence, a question about intent to marry may also predict desistence from crime.

Does intent to marry, however, capture as strong a social bond as marriage? The answer would be “yes” if intent to marry predicts actual marriage, so an important question for the present research is whether intentions accurately predict behavior, a topic that has been extensively researched. In a widely cited review, for a wide variety of studies Ajzen and Fishbein (1977) find high correlations between intentions and subsequent observed behavior, provided that the measure of the attitude or intention closely matches the measure of the behavior along four dimensions: the specific action described (e.g. marriage in the present case), the target of the action (e.g. marriage to someone in general vs. marriage to a specific person), the context of the action (such as the marriage market where it occurs), and the time frame in which the action

takes place.<sup>125</sup> When the wordings of measures of intentions and behavior match along these four dimensions, the authors find correlations between the two that are generally 0.5 or higher.<sup>126</sup> Later research (e.g. Fischhoff, Parker, de Bruin, Downs, Palmgren, Dawes, and Manski 2000) also shows that youth, when asked to assess the probability of various life events occurring in the future, tend to give responses that reasonably match up to the overall prevalence of such occurrences in the general population. Given that only one party to each romantic relationship is interviewed in the TYP, the implication for the present research is that a statement of intent to marry one's current partner may be considered a valid indicator of a social bond that is strong enough for a male respondent to later propose marriage or for a female respondent to later express to her partner a desire to marry.

This chapter therefore investigates the effects of intent to marry on desistance from general deviance and of marriage on smoking, beginning with fixed effects models and moving into marginal structural models for marriage. The two main research questions to be tested for Taiwanese youth are that intent to marry causes a reduction in general deviance, and that marriage causes desistance from smoking. Accordingly, the formal hypotheses tested in this chapter are presented as pairs, one counterfactual and one relational, corresponding to these research questions. Because the counterfactual models require a great deal of effort to extract

---

<sup>125</sup> Ajzen and Fishbein also find in their review that when the targets of intention and subsequent behavior differ, the two tend to be uncorrelated. These findings also imply that self-reported marital status in the TYP survey should not be used to assess the validity of self-reported intent to marry one's partner in previous waves, as the targets of the two actions differ (the questions about intent to marry refer to a specific person, while subsequent marriage does not have to be to the person the respondent had previously been dating when they answered the question about intent).

<sup>126</sup> Manski (1990) also advocates for the measurement of intentions using probabilities instead of questions with "yes," "no," and "undecided" categories. He notes that at best, a respondent stating an intention to carry out an action by some time  $t$  suggests that his or her perceived probability of performing the action by time  $t$  is greater than some probability threshold  $\pi$ , and the researcher might even know the value of  $\pi$ . In an extreme example, if  $\pi = .5$  and all of the respondents to a survey question believe they are exactly 51% likely to perform the action by time  $t$ , then 100% of the respondents would indicate an intention to perform the action by time  $t$ . If a follow-up survey at time  $t$  asks if they actually did what they said they intended to do and no unforeseen factors influenced the respondents' probability of taking the action in the meantime, then only 51% of the respondents would answer "yes." The example is extreme, but drives home Manski's point that intent should be measured using probabilities.

covariates from the TYP survey to address selection into treatment, the relational hypotheses are tested first. If they yield null findings, the causal models are assumed to yield null findings as well and are not further tested.<sup>127</sup> Additionally, intent to marry is measured in two ways in the TYP survey. First, respondents who are currently dating are asked whether they intend to marry their partners. Second, to capture intent to marry among all youth regardless of whether they have a partner, all respondents are asked if they intend to marry in the future and if so, at what age. The hypotheses related to intent to marry and deviance are tested using both of these measures. These research questions and the available data yield the following six hypotheses to be tested in this chapter:

*H5.1.* Unmarried Taiwanese youth who state that they intend to never get married in the future are more deviant than those who either have not thought about it or intend to get married at some point.

*H5.2.* Were some outside factor to cause a young, unmarried Taiwanese person to go from either not having thought about marriage or intending to marry in the future to deciding to never marry, he or she would then become more deviant.

*H5.3.* Taiwanese youth who are dating and intend to marry their current partners are less deviant than Taiwanese youth who are dating and either do not intend to marry their current partners or are undecided.

*H5.4.* Were a young Taiwanese person with a current partner to hypothetically go from either not intending to marry his or her current partner or being undecided to intending to marry that person, that change in mindset would cause him or her to become less deviant.

*H5.5.* Marriage is associated with a reduction in smoking for Taiwanese youth.

*H5.6.* Marriage causes a reduction in smoking for Taiwanese youth.

---

<sup>127</sup> This approach of course is prone to suppressor effects from unobserved covariates masking a significant relationship between intent to marry and deviance or to bias due to survey attrition or the absence of a selection model for intent to marry producing the same results, but suppressor effects are fairly rare in sociological research, and usually models that properly handle selection into treatment reveal spurious causal relationships rather than demonstrating that spurious relationships observed through simple correlation analyses are in fact significant causal relationships.

Limitations of the data related to both the number of married respondents and the content of surveys administered to older TYP respondents, however, prevent the direct interaction of social control measures with marriage as was done with parental monitoring and work in Chapter 4.<sup>128</sup>

As was the case for the counterfactual models in Chapter 4, the marginal structural models in this chapter require the identification of a set of covariates that predict selection into marriage. This research does so by reviewing the literature concerning the cultural factors influencing marriage in China and Taiwan, and through regression models designed to explain intent to marry among the TYP respondents. Probably because sociologists spend more time studying behavior than intentions, very little sociological research investigates why people intend to marry – only why they do marry. Given the findings in the literature that intentions are generally predictive of actual behavior, however, it is reasonable to use the same determinants of marriage as a behavior to predict a persons' intention to marry. This chapter therefore uses the factors that affect selection into marriage as a proxy for selection into intent to marry, and treats both social processes as the same during the literature review and modeling. Also, note that all respondents to the TYP survey who are married by 2011 are married by age 26, so the generality of the findings should be restricted to this age group.<sup>129</sup> The selection models for marriage will also provide some insight into how social control, as measured by respondents' attachment and commitment to their parents and school, influences selection into marriage. The analyses in this chapter are thus restricted to an investigation of the causal effect of early marriage on smoking,

---

<sup>128</sup> Also important to keep in mind is that the selection models presented here, as was the case in Chapter 4, are designed to maximize the explained variance in marriage and not to test any specific theory, including social control. They nonetheless contain useful information for those who would conduct explicit tests of such theories in the future.

<sup>129</sup> I would argue, however, that because of the age-crime curve, this age group is exactly the subset of the population in which criminologists would be interested – and it is also the group that China's gender imbalance, discussed in the introduction to this dissertation, is affecting right now.

and of the relationship between the intent to marry while still fairly young and general deviance.<sup>130</sup>

The chapter continues with a short, focused literature review to frame the selection models for marriage and intent to marry. An examination of correlations between deviance and the survey items for intent to marry for each wave follows, followed by negative binomial regressions that control for a few simple variables (age, gender, and cohort). Should at least one negative binomial regression prove significant, then fixed effects models for that survey question and deviance are also reported. Significant findings in the fixed effects framework then justify a causal model.<sup>131</sup> These statistical tests find no support for Hypotheses 5.1 and 5.3 and indicate a spurious relationship between intent to marry and general deviance instead, but fixed and random effects models for marriage and smoking do find that marriage is associated with desistence from smoking. Thus, following the analytical design for this chapter, Hypotheses 5.2 and 5.4 are considered highly likely to be unsupported, but counterfactual models are fit to test Hypothesis 5.6. Given that these empirical results indicate the need for selection models for marriage but not for intent to marry, the analysis then predicts intent to marry with other background covariates in order to supplement the literature review and provide additional empirical information that can inform the choice of variables for selection into marriage. These additional models provide additional insight into the role of parental social control in affecting the strength of romantic

---

<sup>130</sup> This chapter continues the strategy used in Chapter 4 of first fitting fixed and random effects models of the relationship between deviance and measures of work, then fitting counterfactual models for those results that are significant. Because the fixed and random effects models do not show a significant relationship between intent to marry and deviance, the analysis in this chapter does not include counterfactual models for the effect of intent to marry on deviance.

<sup>131</sup> This approach of course is prone to suppressor effects from unobserved covariates masking a significant relationship between intent to marry and deviance or to bias due to survey attrition or the absence of a selection model for intent to marry producing the same results, but such effects are fairly rare in sociological research, and usually models that properly handle selection into treatment reveal spurious causal relationships rather than demonstrating that spurious relationships observed through simple correlation analyses are in fact significant causal relationships.

relationships for Taiwanese youth. Selection models for marriage then follow, along with tests of Hypothesis 5.6.

*The Role of Parents and Higher Education in Chinese Marriage Decisions*

In his classical work *Xiangtu Zhongguo* (“Earth-Bound China”), the Chinese sociologist Fei Xiaotong writes that individual Chinese have a mental model of the relative importance of social relationships and obligations best understood as a set of concentric circles centered on the individual; closest to the individual is his or her family, followed by extended family, friends and fellow villagers, co-workers, other Chinese, and then everyone else (Hamilton and Zheng 1992). As families were the primary unit of economic organization in agrarian China (including Taiwan), familial ties have historically constituted the most important type of social relationship. In Chinese culture, women permanently leave their own families for their husbands’ upon marriage, so over time families that produced many sons gradually grew in size. Members of such families remained close to ancestral graves and shrines, both to express filial piety to their ancestors and to leverage kinship ties into local economic and political power. China thus evolved into a system of clans with entire villages or townships sharing a common surname (Fricke, Chang, and Tang 1994).

Fan and Huang (1998), summarizing other research into the functions of marriage in traditional Chinese society, note that for the parents of the groom, among other things marriage was important for preserving the family line, brought labor resources into the family, formed social networks, ensured care in old age for parents who had sons, and involved the transfer of economic resources through dowries. Most of these factors have little to do directly with the bride and groom, so most marriages were arranged by their parents (Chu and Yu 2010). With the

onset of mechanization and urbanization in Taiwan after 1950, the rapid growth of higher education, and the emergence of the welfare state, family ties have become less important for getting ahead and the timing of marriage has changed. The average age of marriage in Taiwan increased from 20.61 during the decade from 1955-1964 to 25.06 during the decade from 1995-2004 among those with low education and from 20.72 to 27.61 within the same time for those with high education (Chu and Yu 2010:94),<sup>132</sup> while the fraction of marriages that were arranged dropped from 39.12% before 1970 to 6.44% after 1990 (Chu and Yu 2010:92), a trend also accompanied by a dramatic increase in higher education.<sup>133</sup>

While modernization might expose modern Chinese youth to new ideas, it can also reinforce traditional values. China's civil service examinations, in place from 1371 to 1905, long constituted a family's primary avenue for upward socioeconomic mobility. Educational achievement, especially testing, has therefore taken a place in Chinese thinking as a virtue in and of itself (Chu and Yu 2010). Entrance examinations exist in China and Taiwan for both high school and college, and parents often encourage their children to put off romance in order to better study for the tests. Evidence also suggests that modernization has not supplanted traditional Chinese cultural values concerning filial piety: in a study administering questionnaires to 110 Chinese and 132 American college students, Zhang and Kline (2009) found that Chinese students were also more likely than their American counterparts to state that they expected their married partners to make their parents happy and respect their parents' opinions, and that whether a student's current romantic partner met this requirement constituted an important

---

<sup>132</sup> The authors measure education in years, but they do not state where they draw the threshold between low and high education in their section that discusses the timing of marriage.

<sup>133</sup> According to an island-wide survey of married women in 1986, 62% of women born between 1945 and 1949 who had no education had their marriages arranged by parents, compared to 46% of those with a primary school education, 22% of those with a senior high school education, and 11% of women with a college education. As the fraction of the population who had graduated from senior high school increased from 38.8% in 1950 to 80.05% in 2004 (Chu and Yu 2010:34), the reason for the drop in arranged marriages is clear.

criterion affecting his or her decision to marry the person or not. These results suggest that both parental attachment and being in school should predict selection into marriage, and both of these variables are included in the selection models that follow.

### *Early Romance as a Type of Precocious Maturation*

Romance and part-time work share a common characteristic in Chinese culture: they are both often seen as permissible only with parental approval, and otherwise are themselves a type of deviant behavior. The previous chapter has demonstrated that early work can be criminogenic, but only in the absence of social control, such that parents do not know that their children are working. It may be, then, that for Taiwanese youth, dating at an early age may also be associated with deviance. Table 5.1, which presents a random effects negative binomial regression of delinquency on dating for TYP survey respondents as a preliminary step in the analyses for this chapter, provides empirical evidence to support this notion. The table shows that dating and deviance are positively associated for TYP survey respondents in both cohorts who were interviewed during their senior year of high school, sophomore year of college, or senior year of college. In particular, a respondent stating that he or she currently has a boyfriend or girlfriend is associated with a multiplicative increase in the odds of general deviance of about 4.9% – an effect size similar in magnitude to that seen for part-time work in the previous chapter.<sup>134</sup> These

---

<sup>134</sup> A few details about the models are worth mentioning: as these models pool the J1 and J3 cohorts and an artifact of the IRT modeling approach requires the inclusion of an indicator variable for each wave of the survey, the models in Table 5.1 control for cohort and wave (“J1 cohort”, “Wave C2” for the year in which most respondents were sophomores in college, and “Wave C4” for the wave in which most respondents were college seniors), with two additional interactions between cohort and wave. The model parameters when all five of these terms are zero correspond to members of the J3 cohort during their senior year in high school. The effect of dating on delinquency for the fixed effects model is also positive, but the effect size is lower and the standard errors are also much larger. As these models omit time-varying covariates (most notably any social control variables such as parental monitoring), and as I use a negative binomial model for convenience in interpreting model parameters without testing its goodness of fit compared to the Inverse Gaussian model, these models are not intended to be the final

findings justify controlling for predictors of deviance in modeling selection into marriage, so this chapter includes in those selection models the same predictors of delinquency in Taiwan derived from the literature review in Chapter 3 that were used to model selection into work in Chapter 4 (see Table 3.1).

**Table 5.1.** Random effects negative binomial regression of delinquency on dating for respondents to Waves 5, 8, and 9. The outcome variable used here is the Expected A Posteriori (EAP) score for delinquency as generated by the IRT models in Chapter 2, which is then transformed to be positive for all observations, multiplied by 1,000, and rounded to the nearest integer to be suitable for modeling with a negative binomial regression.

	Random Effects Negative Binomial Model		
	Coef.		SE
<b>Having a Partner</b>	<b>.048</b>	<b>***</b>	<b>.0066</b>
Female	<b>-.134</b>	<b>***</b>	<b>.0063</b>
J1 cohort	.042	<b>***</b>	.0089
Wave C2 (college sophomore)	-.471	<b>***</b>	.018
Wave C4 (college senior)	-.0061		.027
J1 respondent in wave C2	-2.15	<b>***</b>	.023
J1 respondent in Wave C4	-.077	<b>***</b>	.015
Age	.077		.108
Age <sup>2</sup>	-.079		.0070
Intercept	9.04	<b>***</b>	.17
N	4,712		
Person-years	11,076		
*** p<.01 **p<.05 * p < .10			

The previous discussion has indicated the importance of social control via family bonds on impacting the timing of romantic relationships in Taiwan, but social control factors also influence romantic relationships in Chinese culture through their effect on peer relationships. In a

---

word on the relationship between dating and deviance, but rather demonstrate that deviance is associated with dating for the TYP respondents.

structural equation model of 496 high school sophomores and juniors in Xi'an, China, Li Zhihong and her colleagues (2010) find that increased attachment to parents is *directly* related to decreased attachment to romantic partners, but also that increased parental attachment may also *indirectly* raise attachment to romantic partners because it raises attachment to peers, which in turn is positively associated with attachment to romantic partners. Their findings suggest that those in healthy relationships with their parents go on to develop strong relationships with their friends and then partners. The role of peers therefore merits further discussion.

### *Peers and Romantic Relationships in Taiwan*

Because traditional Chinese marriages required the approval of members of the kinship networks of both the husband and wife, marriages tended to take place within the same neighborhood. Fan and Huang (1998) cite a state newspaper in Beijing (People's Daily 1989) reporting results of a survey in the late 1980s showing that the brides and grooms of most rural marriages lived within 25 km of each other.<sup>135</sup> And in pre-industrial China, a Chinese youth's friends were likely to be from the same clan, so peer opinions were almost as influential as those of family (Hamilton and Zheng 1992).<sup>136</sup> Through in-depth interviews of 40 Chinese students and a questionnaire survey of 231 students at four prestigious universities in Beijing, Moore (1998) finds that Chinese students avoid casual romantic relationships in order to avoid the derogatory and stigmatizing label of "*qingfu*" (a compound word associated with being light or

---

<sup>135</sup> This dissertation generally uses only primary references, but in this case the original People's Daily article was difficult to obtain.

<sup>136</sup> One result of these historically interconnected peer and family networks is that the Chinese are cautious about disclosing secrets to friends, family, and dating partners. A study comparing 32 Chinese Canadian college students to 32 European Canadian college students found that Chinese Canadian males who express "gender role traditionalism" (agreeing to statements such as "the first duty of a woman with young children is to home and family") are less likely to express satisfaction in intimate relationships than European Canadian counterparts. The mechanism driving this decreased satisfaction is the fact that they do not disclose as much to their romantic partner (Marshall 2008).

floating that implies a lack of grounding in a network of relationships with family and friends). Peer values therefore often reinforce parental control by means of these cultural norms. Zhang and Kline (2009) also found that Chinese male students were more likely than their female counterparts to state that they would go along with their peers, parents, and extended family when deciding who to marry.<sup>137</sup> Additionally, they set up a model for intent to marry, predicted by respondents' assessment of the importance of parental and peer views about their own relationships combined as a "network effect" covariate,<sup>138</sup> and controlling for factors derived from respondents' statements of what qualities they sought in a potential spouse. They found that only the network effects explained significant variation in the intent to marry for the Chinese sample, while other personal views remained important for Americans. Attachment to peers is unfortunately not available in later waves of the TYP survey for the J3 cohort, but the survey does contain questions about the fraction of respondents' peers who have had sex that are important for the present research, as those peers who are having sex early may be more likely to encourage a respondent to ignore parents' advice to delay both sex and marriage until after school.

Even as the importance of parents to Taiwanese youth may not have changed much during modernization, a few social factors have affected the power dynamics between parent and child that influence the timing and nature of marriage. First, Taiwan's shift from an agricultural to a wage-based economy means that parents have experienced a relative loss of power compared to a time in which they controlled the means of production (Thornton, Fricke, Yang, and Chang 1994). Combined with higher education, this shift in the nature of labor might lead an observer to

---

<sup>137</sup> In contrast, for the American students in the sample, female students were more likely than the male students to be guided by social network factors.

<sup>138</sup> An example of a question used in this scale is "generally speaking, I want to do what my parents think I should do regarding my relationship decisions, especially marriage decisions" (Zhang and Kline 2009:13)

expect that Taiwanese youth would make more independent decisions about marriage now than in the past. Additionally, because college-educated Taiwanese women feel indebted to their parents for covering tuition, they may delay entry into a husband's family while supporting their biological parents with wages from a job, or else may use their degree as a bargaining chip to persuade in-laws to let them support their biological parents as well (Thornton, Fricke, Yang, and Chang 1994:93). The models for selection into marriage in this chapter therefore include as predictors measures of whether students have taken out loans; those who do should marry later.<sup>139</sup>

### *Self-Esteem as a Predictor of Marriage in Taiwan*

The rising social status of women in Taiwan may impact both deviance and the timing of marriage via self-esteem. Giordano et al. (2007) argue that emotions are social, and that crime and other deviant activities are a way of expressing feelings such as anger and of expressing identity through a process of symbolic interaction. For Giordano and her colleagues, youth with high self-esteem would be less likely to offend because they are more likely to be in relationships involving love that lower crime through positive reflective appraisals that reduce negative emotions such as anger. Self-esteem also affects selection into marriage, moreover, by giving young people in Asia a sense of freedom to choose a mate. A study comparing university students in Indonesia and Canada (MacDonald, Marshall, Gere, Shimotomai, and Lies 2012) models the effect of self-esteem on "relationship value" (a composite of satisfaction, trust, and

---

<sup>139</sup> Unfortunately, this question is only asked of students currently in school, but similarly to the techniques in Chapter 4 used to model selection into work, the variable can be included in regressions predicting marriage without loss of sample size due to missing values by assigning an arbitrary integer to the student loan variable for respondents not in school, interacting the student loan variable with an indicator variable for being in school, and including the interaction effect without a main effect for student loans. A similar approach is used in this chapter to include variables measuring school attachment and commitment (asked only of respondents in school), variables only meaningful for respondents who are not already married, and variables measuring characteristics of ongoing romantic relationships (which require respondents to be dating someone).

intimacy). The authors hypothesize that those with high self-esteem are more likely to believe that their friends would agree that they have made a good decision in finding a partner suitable for marriage – a prerequisite for marriage that is important in collectivist societies. Comparing Indonesian students to Canadian students and then Japanese students to Australian students, in all cases self-esteem has a positive indirect effect on “relationship value” via reflected appraisals: respondents with high self-esteem believe their peers would say their partner is of high value, and respondents reporting these high reflected appraisals in turn are more likely to state that their relationship is indeed of high value. But only in the Asian samples does self-esteem also indirectly affect relationship value via family approval, showing that self-esteem in Asia allows a person to more proactively manage their social relationships, thus gaining some freedom in a system normally seen as constraining.<sup>140</sup> Self-esteem is therefore included in regression models for marriage when the survey variables are available, and all else being equal, positive self-esteem should predict early marriage.

#### *Characterizations of Romance as Predictors of Marriage in Taiwan*

Cho and Cross (1995) also find that self-esteem predicts relationship quality, but their most important finding with respect to the present research is that the outcomes of Taiwanese students’ relationships are related to their values regarding relationships in general. They measure the responses of 99 Taiwanese students at the University of Texas to the Love Attitudes Scale (Hendrick and Hendrick 1990), designed to capture six different types of love: *eros* (love predicated on a person’s physique), *agape* (selfless, giving love), *storge* (love predicated on

---

<sup>140</sup> A related study of relationship attachment among Chinese college students found that attachment to one’s romantic partner, net of attachment to mother, father, friends, and oneself, predicted greater self-esteem (Wang and Wang 2012). For this study, “attachment to self” differed from self-esteem in that the former measured a respondent’s view of whether he or she was worthy of being loved.

friendship), *ludos* (a non-committal love based on the concept of playing a game), *pragma* (love based on pragmatic concerns such as income or social position), and *mania* (obsessive love). Via a principal components analysis, they find that respondents are most likely to perceive love as a blend of eros and agape that the authors describe as “romantic and considerate” (explaining 11.0% of the total variance in their analysis), followed by a combination of storge and eros items (8.7%), ludos (7.1%), pragma (6.7%), and mania (6.5%). The last factor, explaining 6.4% of the variance, is a blend of pragma and agape reflecting traditional values prioritizing the extended family (such as “a main consideration in choosing a partner is how he/she reflects on my family.”) Correlations among the factors suggest that Taiwanese who predicate love on pragmatic considerations often do so out of a sense of obligation to other family members such as their parents. The authors also find a positive correlation between self-esteem and romantic / considerate love, pragmatic love, and love based on friendship. Additionally, respondents predisposed towards romantic / considerate love are more satisfied with their relationships, and those involved in friendship-based love are less likely to break up with their partners.<sup>141</sup> The implication of these findings is that love typologies may affect selection into marriage.

The studies mentioned so far in this literature review share a common methodological flaw, however: they draw on cross-sectional samples of college students, limiting the generalizability of their findings. One longitudinal survey (Cheng and Landale 2011), however, draws on Waves 1 and 3 of the Adolescent Health survey to investigate early marriage among Asian Americans in comparison to their counterparts who are white, black, or Hispanic. The authors posit that parental values have a greater impact on the timing of marriage for Asian respondents in their

---

<sup>141</sup> Additionally, respondents who described their love types as “romantic and considerate” or based on friendship also had a lower self-reported probability of breaking up with their current partner, and respondents in “romantic and considerate” relationships also reported higher satisfaction with their relationship while those in obligatory relationships were slightly less satisfied. The number of conflicts per month in a relationship was not associated with relationship type.

sample because of the Asian incorporation of the family views towards dating partners into their own identity mentioned previously. In discrete time event history models predicting transition to first marriage, they find that parental religiosity predicts earlier marriage, while parents' disapproval of their child having sex at the time of the survey predicts later marriage;<sup>142</sup> these values remain consistent even after accounting for the child's own beliefs, romantic experiences,<sup>143</sup> and educational attainment.<sup>144</sup>

Additional empirical evidence from the Taiwan Youth Project data itself also helps to explain the role of social control in explaining whether a young Taiwanese person would intend to marry his or her romantic partner at an early age, but as indicated in the earlier description of the organization of this chapter's analyses, in this chapter empirical models of intent to marry are more important for selecting a set of variables to model selection into marriage than for setting up counterfactual models for the effect of intent to marry on deviance. In fact, the relationship between intent to marry and general deviance turns out to be spurious for the TYP respondents while those who marry are less likely to smoke. The next set of analyses demonstrate this fact through tests of Hypotheses 5.1 and 5.3, and the social control factors described in this review will be returned to as part of investigating marriage itself.

### *Models for Intent to Marry and Deviance*

Next, this chapter presents generalized linear and fixed effects regression models for the effect of intent to marry on deviance, as a way to assess if the relationship between these two

---

<sup>142</sup> Both of these questions are asked directly of parent respondents, and are not the child's appraisals of the parent's beliefs.

<sup>143</sup> These variables include "ever had a romantic relationship, ever had a sexual experience, and ever had sex with a nonromantic partner" (Cheng and Landale 2011:219).

<sup>144</sup> Following Cheng and Landale, the models predicting intent to marry in this chapter also include parental religiosity, but since the predictor is associated with a large percentage of missing values due to its being drawn from a parent survey and preliminary analyses indicated that it does not predict intent to marry, it is excluded from the models predicting selection into marriage.

variables is strong enough to justify fitting a causal model. In contrast to Chapter 4, which used the J1 cohort only to model the causal effect of part-time work on deviance, the models testing Hypotheses 5.1 and 5.3 pool the J1 and J3 cohorts to gain statistical power. Prior to conducting any analyses, it is important to use measures of deviance that minimize measurement error. Chapter 4 used Wave 4 of the J3 cohort as the root node in the graph representing the sequential common item linking process for vertical scaling, such that the vertical scaling procedure placed deviance measures for all other waves onto the scale for that wave. The present analyses, however, only use Waves 8 and 9 of the J1 cohort and Waves 6 and 7 of the J3 cohort (corresponding to respondents who are of the same age as college sophomores and seniors), and therefore use a different vertical scaling procedure entirely to avoid the problematic link between Wave 4 for the J3 cohort and Wave 9 for the J1 cohort (see Figure 2.5), and thus keep measurement error low. A second repeated application of Dijkstra's algorithm to each node of this subset of the graph in Figure 2.5 identifies Wave 9 for the J1 cohort as the root node.

As mentioned previously, two types of variables for intent to marry are included in these four survey waves, and examined in the present chapter. The first question asks all respondents regardless of whether they are currently dating someone, "At what age do you intend to marry?" Respondents who are not already married can write down an age in a blank, or they can circle "I do not want to get married" (不想結婚) or "I haven't thought about it." "Intent to not marry," created by encoding "I do not want to get married" as 1 and any other response as 0, then constitutes the dependent variable for testing Hypothesis 5.1. The second, used to test Hypothesis 5.3, asks respondents who currently are dating someone if they intend to marry that person, with "yes," "no," and "undecided," as possible responses and respondents with no partners coded as

missing.<sup>145</sup> The present analysis collapses the “no” and “undecided” categories together to denote a person who does not currently intend to get married.

Tables 5.2 and 5.3 present biserial correlations between deviance and intent to never marry for each survey wave under investigation. Intent to never marry is more or less uncorrelated with deviance, with the lone exception that respondents to J1 Wave 8 who state that they do not wish to get married in the future are less deviant ( $p = .045$ ).<sup>146</sup> Correlation analyses and negative binomial regressions of deviance on intent to marry one’s current partner and controlling for cohort, age, and gender also resulted in null findings (not shown), so the chances of causal or fixed effects models finding any significant effect of intent to marry one’s partner on deviance are quite small, and thus this analysis finds no support for Hypotheses 5.3 or 5.4.

**Table 5.2.** Correlations between deviance and intent to marry for respondents to the Taiwan Youth Project survey who are currently dating someone. Responses to “intent to marry” are coded as “no” (1), “undecided” (2), or “yes” (3).

	<b>Correlation with Deviance (IRT Measure)</b>			
	J1 Wave 8	J1 Wave 9	J3 Wave 6	J3 Wave 7
Intent to marry, for respondents with partners	.059	-.015	.013	.011
N	604	690	571	708

\*\*\*  $p < .001$  \*\*  $p < .01$  \*  $p < .05$  +  $p < .10$

<sup>145</sup> In fact the TYP survey also asks respondents who were previously in a relationship (but are no longer in the relationship) if they intend to marry their previous partner. These responses to the “intent to marry” question are unanalyzed in the present analysis, as the intent to marry someone who likely broke up with the survey respondent is likely not a good predictor of marriage.

<sup>146</sup> Because collapsing “no” and “undecided” into one category may have eliminated useful information related to the intent to marry one’s current partner, these correlation analyses were followed with ANOVA tests for whether the mean IRT deviance value was different for varying levels of intent to marry, treating “no,” “yes,” and “undecided” as separate categories. None of these tests yielded significant results.

**Table 5.3.** Correlations between deviance and intent to never marry for respondents to the Taiwan Youth Project survey. Respondents not intending to ever marry are coded as 1, with respondents intending to marry in the future coded as 0.

	<b>Correlation with Deviance (IRT Measure)</b>			
	J1 Wave 8	J1 Wave 9	J3 Wave 6	J3 Wave 7
Intent to never marry, for all respondents	-.048*	.007	-.033	-.029
N	1717	1844	1831	1846
Number intending to never marry	199	189	189	155

\*\*\* p < .001 \*\*p < .01 \* p < .05 + p < .10

The findings in Table 5.3 are similarly null, except for the negative correlation between intent to never marry and deviance for the J1 Wave 8 responses (which could also be significant due to random chance alone, as a result of the multiple tests for more than one wave). As a follow-up, Table 5.4 thus presents negative binomial regressions for deviance on intent to never marry, combining the J1 and J3 cohorts to maximize statistical power and controlling for age, gender, and cohort.<sup>147</sup>

<sup>147</sup> In particular, the analysis pools J1 Wave 8 and J3 Wave 6 to capture respondents at the age at which most are college sophomores, and similarly pools J1 Wave 9 and J3 Wave 7 for respondents who are two years older. To fit the negative binomial regression, as usual the deviance measure is transformed by adding a scalar value to all IRT scores to make them positive, multiplying the latent variable by 1,000, and rounding to the nearest integer.

**Table 5.4.** Negative binomial regressions of deviance on intent to never marry for respondents to the Taiwan Youth Project survey. Respondents not intending to ever marry are coded as 1, with respondents intending to marry in the future coded as 0.

	Negative Binomial Regressions for			
	J1 Wave 8 / J3 Wave		J1 Wave 9 / J3 Wave 7	
	Coef.	Std. Error	Coef.	Std. Error
Intent to never marry	-.049	.056	.027 *	.011
Age	-.076 *	.036	-.0033	.0065
Gender (female = 1)	-.80 ***	.035	-.018 ***	.0063
J3 cohort	1.43 ***	.049	.024 **	.0087
Intercept	9.98 ***	.74	10.1 ***	.14
N	3547		3689	

\*\*\* p < .001 \*\*p < .01 \* p < .05 + p < .10

Table 5.4 indicates that for the J1 Wave 9 / J3 Wave 7 survey (administered at a time when most respondents are college seniors), those who report intent to never marry are slightly *more* deviant than those who do intend to marry or are undecided (note the sign flip compared to the corresponding correlation in Table 5.3).<sup>148</sup> Taken at face value, Table 5.4 suggests that respondents who have decided not to marry may be more deviant, perhaps because to some degree their very stated intention to never marry runs counter to social norms and hence reflects some crude measure of deviance, or because they do not consider themselves a good catch due to their deviant behavior or the personal characteristics that have led to it. Still, the results are quite weak due to the sign flip when comparing the regressions to the biserial correlations; controlling for gender quite likely caused the sign change, and so as a check, controlling for other forms of time-invariant heterogeneity with fixed effects models makes sense. Table 5.5 presents unconditional fixed effects negative binomial regressions of deviance on intent to never marry.

<sup>148</sup> A similar pooling of cohorts and regression analyses of intent to marry one's current partner yielded null findings.

The unconditional regressions are less vulnerable to incidental parameters bias than negative binomial models in which fixed effects are conditioned out of the estimation (Allison and Waterman 2002).

**Table 5.5.** Negative binomial regressions of deviance on intent to never marry for respondents to the Taiwan Youth Project survey. Respondents not intending to ever marry are coded as 1, with respondents intending to marry in the future coded as 0.

	Fixed Effects Negative Binomial Model		Random Effects Negative Binomial Model	
	coef.	SE	Coef.	
<b>Intent to Never Marry</b>	<b>-.024</b>	<b>.050</b>	<b>.019</b>	
Female			-.19	*** .010
J1 cohort	-2.90	3.98	-2.00	***
C4 (J1W9 / J3W6)	.49	*** .064	.46	***
J1 cohort * C4	2.66	*** .066	1.95	***
Intercept	9.29	** 3.96	2.00	***
N	2,982			
Person-years	5,964			

\*\*\* p<.01 \*\*p<.05 \* p < .10 † Outer Product of the Gradient.

<sup>a</sup> Referent category: general track

<sup>b</sup> Referent category: Taipei City

Table 5.5 shows that both the fixed and random effects negative binomial regressions indicate no relationship between intent to never marry and deviance, and an equivalent inverse Gaussian random effects model (not shown) also finds no relationship.<sup>149</sup> Hypothesis 5.1 is therefore not supported, and so Hypothesis 5.2 also has little to no support and is not explicitly tested with additional counterfactual models. The analysis therefore finds that intent to marry is unrelated to general deviance, regardless of whether the instrument measuring intent to marry

<sup>149</sup> Since both models produce the same null findings, a test of the random effects null hypothesis that the fixed effects are uncorrelated with the substantive parameters is unnecessary here.

refers to a specific person (Hypothesis 5.3) or a more abstract intention to get married (Hypothesis 5.1). Perhaps the expression of intent does not carry with it the same behavior-changing commitment to the bond, then, as actually following through and getting married. The focus of the present analysis thus turns at this point to an examination of marriage and a test of Hypothesis 5.5.

### Fixed Effects Analyses of Marriage

Table 5.6 lists the number of TYP respondents who are married by wave and cohort, along with the approximate age of the respondents.<sup>150</sup> With the IRT deviance variables only available for J3 Waves 6 and 7 and for J1 Waves 8 and 9 and given how few respondents to the TYP survey have married by these waves, it is no surprise that preliminary correlational analyses indicate no relationship between marriage and general deviance. Accordingly, tests of Hypotheses 5.5 and 5.6 draw on the 2011 TYP survey, which interviewed J3 respondents when they were between 26 and 27 years old and J1 respondents when they were between 24 and 25 years old, and which includes more than 200 married respondents, most of whom are members of the J3 cohort. During this wave, respondents were asked how much they smoked each year for every year of their life between ages 12 and 25, with an additional single question asking, for a handful of slightly older respondents, how much they smoked during an average year when they were 26 or older. In an apparent attempt to reduce measurement error associated with the imprecision of the retrospective survey questions, the survey design simply asks each respondent

---

<sup>150</sup> Comparing the J1 Wave 8 and J3 Wave 6 surveys, as well as the J1 Wave 9 and J3 Wave 7 surveys, respondents are on average a year older for the J1 cohort despite being interviewed during the same year of school. This is because the J3 surveys were typically administered in the fall and the J1 surveys in the spring, and due to limitations of the data age here is approximate, derived from the respondent's birth year and the year of the interview rather than from the respondent's birthday and the interview date.

to describe his or her smoking behavior for each year in terms of three categories: “did not smoke,” “smoked five or fewer packs [that year]”, or “smoked more than five packs.”

**Table 5.6.** Number of married respondents for each cohort-wave in the TYP Survey.

<b>Wave</b>	<b>Approximate Age of Respondents</b>	<b>Number of Married Respondents</b>	<b>Total Number of Respondents</b>	<b>Percent Married</b>
J3W6	19.3	9	1,844	0.5%
J3W7	21.3	27	1,873	1.4%
J3W8	23.3	58	1,797	3.2%
J3W2011	26.4	141	1,536	9.2%
J1W8	20.3	10	1,739	0.6%
J1W9	22.3	32	1,874	1.7%
J1W2011	24.3	64	1,591	4.0%

Perhaps not surprisingly given the retrospective nature of the survey questions, respondents to this survey are more likely to report the same smoking behavior for consecutive ages than is usually the case for prospective longitudinal surveys of delinquency. This fact poses a significant problem for fixed effects models, for which within-individual change in the dependent variable must be observed in order for a respondent’s person-years to be included in the regression. Projection bias associated with retrospective self-reported delinquency, then, can lead to both biased regression coefficients and a reduction in sample size for fixed effects regressions.

One way to gain a sense of whether projection bias could be at play is to examine the correlations of the retrospective responses with each other and with contemporaneous measures of smoking. Respondents to the Waves 6 and 7 surveys for the J3 cohort stated how many cigarettes they had smoked in the past week. These variables cannot be directly compared to the retrospective survey questions, which ask about smoking during the entire year and contain categories that are more coarse-grained than the prospective questions. They can, however, be coded into binary variables corresponding to whether respondents smoked during the past week,

and correlated with retrospective binary variables for whether respondents smoked at any time during that year. As smoking is an addictive behavior, the correlation between the two might be expected to be rather high despite the different time intervals used to measure smoking.

Regardless, if the correlation between the retrospective binary smoking variables for Waves 6 and 7 is higher than the correlation between the corresponding prospective survey questions, projection bias might be at play.

**Table 5.7.** Tetrachoric correlations for prospective and retrospective smoking questions for the TYP survey, J3 cohort.

	Wave 6 prospective	Wave 7 prospective	Wave 6 retrospective	Wave 7 retrospective	Wave 8 retrospective
Wave 7 prospective	.938				
Wave 6 retrospective	.925	.939			
Wave 7 retrospective	.919	.962	.993		
Wave 8 retrospective	.907	.957	.970	.989	
Wave 10 retrospective	.881	.939	.934	.964	.993

Table 5.7 gives the tetrachoric correlations between all pairs of prospective and retrospective survey questions about smoking for Waves 6, 7, 8, and the first wave of the adult survey (“2011”) for the J3 cohort. The prospective survey questions for Waves 6 and 7 are 93.8% correlated, while the retrospective questions for the same waves are 99.3% correlated. The higher correlation between the retrospective responses could be due to projection bias, but it could also be due to the fact that the prospective survey questions only ask about smoking during the

previous week, so differences between the survey instruments could also explain the discrepancy.<sup>151</sup>

The fact that all the correlations in Table 5.7 are so high, however, is perhaps not surprising given that smoking is an addictive behavior. One possible concern, then, is that perhaps not enough within-individual variation in smoking is available to observe an effect of marriage on smoking, particularly given that marriage is a somewhat rare event for the TYP respondents. One way to be certain is to calculate an intraclass correlation coefficient (ICC) for an unconditional hierarchical linear model in which smoking is modeled with only level-one and level-two intercepts and no explanatory variables at either level (Raudenbush and Bryk 2002). Doing so using responses for Waves 7, 8, and the 2011 survey wave for the J3 cohort indicates that 82.3% of the variance in smoking is explained by between-individual variation, while 17.7% is due to within-individual variation.<sup>152</sup> Therefore, despite the high correlation of smoking across survey years, enough within-individual variation in smoking should be available to test Hypotheses 5.5 and 5.6.

The previous correlational analyses in Table 5.7 treat smoking as a binary variable, but the variability in smoking would no doubt be greater were its original coding of three categories for light, moderate to heavy, or no smoking preserved. Multinomial logistic regression models would therefore be preferable to logistic regression models if possible, given the importance of

---

<sup>151</sup> An example may be helpful here: an occasional smoker (perhaps smoking once a week on average) who does not change his or her behavior between Waves 6 and 7 might not report smoking during Wave 6 and report it during Wave 7, whereas that person would certainly report smoking during the year for both retrospective questions. In such cases the higher correlation for the retrospective questions would not be due to projection bias.

<sup>152</sup> This ICC coefficient treats smoking as a binary variable (for smoking at all during the year or not), and is calculated based on a sample of 1,008 respondents observed for a total of 2,745 person-years of data – the same sample used for the causal models for marriage and smoking later in this chapter. Reweighting the sample using the inverse probability weights calculated using the models presented later in this chapter that compensate both for survey attrition and selection into marriage (although what matters for this empty model is only survey attrition) only has modest effects on the resulting ICC coefficient, such that 83.0% of the variance in smoking is at the between-individual level.

within-individual variation in the smoking items when modeling marriage and smoking using fixed effects models. A shortcoming of the responses to these smoking questions, however, is that no married respondents in the J3 cohort reported smoking between one cigarette and five packs a year during Waves 7, 8, and in 2011 when answering the retrospective questions, but instead reported smoking either not at all or more than five packs per day. Relying completely on the J1 cohort for observations about light smoking for married responses is feasible in a model pooling the J1 and J3 cohorts, but may also demand too much of the data. Accordingly, for a primary model testing Hypothesis 5.5, the present analysis recodes smoking variable into a binary indicator variable for whether or not respondents smoked at all in a given year, and models smoking using fixed and random effects logistic regression models. Hypothesis 5.5 is also tested using multinomial logistic regression models as a sensitivity check.

The left panel in Table 5.8 presents the findings from a conditional fixed effects logistic regression model for smoking on marriage, for J1 and J3 respondents of marriageable age (ages 18 and older). Requiring at least two person-years of data and change in the binary dependent variable restricts the number of eligible respondents to just 333 respondents.<sup>153</sup> Marriage is associated with lower smoking for the fixed effects model, but not for the random effects model.

---

<sup>153</sup> Because the statistical power of the test for marriage is weak and the changing nature of the Taiwan Youth Project survey questions also makes the inclusion of additional time-varying covariates in the fixed effects models difficult, these fixed effects models exclude a number of important time-varying predictors, such as parental attachment and parental monitoring. These potential confounders will be controlled for in the selection models for the causal models to follow. Also note that despite the fact that respondents who are only observed for a single person-year can be included in the random effects model (but not the fixed effects model), the same sample is used for both models in order to avoid the possibility that bias associated with survey attrition leads to different findings for each model.

**Table 5.8.** Fixed and random effects logistic regression models for the effect of marriage on smoking for respondents to the Taiwan Youth Project survey who are at least 18 years old. Standard errors for parameter estimates are in parentheses.

	<b>Smoking</b>	
	Fixed effects logit model	Random effects logit model
Age <sup>a</sup>	4.23*** (.96)	3.60*** (.88)
Squared Age <sup>a</sup>	-3.68*** (.93)	-3.07*** (.85)
Married	-1.68*** (.48)	-.35 (.35)
Intercept		-.34*** (.061)
N	333	333
Person-years	1,342	1,342
Log likelihood	-449.5	-880.6

\*\*\* p < .001 \*\*p < .01 \* p < .05 + p < .10

<sup>a</sup> These variables are standardized.

A test of whether the random effects assumption that the person-specific error terms are uncorrelated with the time-varying covariates holds for this model can be conducted with a Hausman test (Hausman 1978). The test indicates that the coefficients for the two models in Table 5.8 are significantly different, suggesting that the fixed effects model estimates, which are inefficient but consistent, are preferable to the efficient but potentially inconsistent random effects model estimates. The results from the fixed effects model indicate that marriage is associated with a dramatic 81.3% reduction in the odds of smoking – or between a 51.3% reduction and a 92.9% reduction in the odds of smoking when taking a 95% confidence interval into account. The model, in controlling for age with a linear and quadratic term, also captures the age-crime curve. In particular, the positive linear and negative quadratic terms reflect the fact that respondents tend to smoke more as they get older, but the rate of this increase decreases over

time and involvement in smoking may also reach a maximum and then begin to decrease. A problem with the analysis, however, is that limiting the sample size to 333 for the fixed effects regression reduces the total number of married respondents included in the analysis to just 39 people, so the confidence interval for the percent reduction in the odds of smoking is quite large – and it is hard to imagine that the protective effect of marriage would amount to a 90% reduction in the odds of smoking.

It is possible to bring more respondents into the analysis, however, at the risk of potentially underestimating the protective effect of marriage. Including in the regressions person-years for which respondents were not yet 18 years old increases the number of respondents for whom change in smoking is observed, such that the subsequent regressions could include more married respondents and thus decrease the standard errors of the parameter estimates for marriage.<sup>154</sup>

---

<sup>154</sup> A disadvantage of this approach is that the model adds person-years for which the amount of smoking can vary but the marriage variable is necessarily always coded as zero. Because of the nature of the age-crime curve, with delinquent behavior increasing for teenagers and decreasing afterwards, we would expect most person-years added to the fixed effects regression in this manner to reflect respondents who age into smoking and then fail to desist (as those who would start as smokers and then quit smoking before age 18 would be relatively rare). Consequently, as the model includes person-years for which smoking gradually increases over time and that are associated with years when respondents are not yet legally able to marry, and as observed smoking for these respondents does not decrease when they are of marriageable age, any bias in the parameter estimates would be in the direction of underestimating the magnitude of the protective effect of marriage. That said, controlling for age and squared age likely mitigates most of this bias.

**Table 5.9.** Fixed and random effects logistic regression models for the effect of marriage on smoking for respondents to the Taiwan Youth Project survey who are at least 12 years old. Standard errors for parameter estimates are in parentheses.

	<b>Smoking</b>	
	Fixed effects logit model	Random effects logit model
Age <sup>a</sup>	8.50*** (.44)	8.21*** (.43)
Squared Age <sup>a</sup>	-6.81*** (.42)	-6.54*** (.41)
Married	-.75* (.31)	-.28 (.30)
Intercept		-.84*** (.062)
N	847	847
Person-years	6,711	6,711
Log likelihood	-1,564.2	-3,203.3

\*\*\* p < .001 \*\*p < .01 \* p < .05 + p < .10

<sup>a</sup> These variables are standardized.

As a sensitivity check for Table 5.8, then, Table 5.9 presents the same conditional fixed effects logistic regression of marriage on a binary indicator variable for smoking, except that it brings in all person-years since age 12. An accompanying random effects model uses the same person-years that are included in the fixed effects model. The resulting gain in statistical power amounts to the inclusion of 847 respondents and a total of 119 person-years of marriage – more than tripling the number of observations of marriage. A Hausman test that the consistent but inefficient fixed effects estimates are different from the efficient but potentially biased random effects estimates cannot be performed in the present case,<sup>155</sup> but the test is usually conducted

<sup>155</sup> A Hausman test of a null hypothesis that the parameter estimates in the fixed and random effects models in Table 5.9 are the same cannot be performed, because the matrix corresponding to the difference between the two variance-covariance matrices associated with the parameter estimates for the two models is not positive definite. In such cases the Hausman test statistic can be an invalid negative value and therefore not meaningful for the purposes of conducting a chi-squared test, and the present case is one such example. Moreover, standard workarounds for

when the fixed effects model yields insignificant parameter estimates due to higher standard errors, but the potentially biased random effects models result in significant findings. In Table 5.9, the unbiased, inefficient estimators for the fixed effects model yield significant findings due to the greater magnitude of the parameter estimates for marriage, whereas the potentially biased random effects estimators do not. Thus, the results favor the inefficient but unbiased fixed effects model, for which marriage retains its significance, with marriage associated with a reduction in the odds of smoking of approximately 45.7% (or between 8.6% and 67.7% to use the 95% confidence interval). Marriage is therefore associated with a reduction in smoking after controlling for within-individual time-invariant heterogeneity, and Hypothesis 5.5 is supported.

An analysis of whether marriage actually causes this reduction in smoking is definitely warranted, but first it may be worthwhile to model all three levels of the smoking variable using fixed and random effects multinomial logistic regressions.<sup>156</sup> With the dependent variable reformulated, Hypothesis 5.4 (that marriage is associated with lower smoking) would hold if married respondents are less likely to smoke one or more packs per year than not to smoke, or less likely to smoke five or more packs than between one and four. Because the additional category in the dependent variable and the additional estimated regressions are more demanding of the limited variability in marriage than the logistic regressions, these models also include person-years since age 12.

---

negative Hausman tests cannot be performed in the present case, because the current implementation of the random effects logistic regression in Stata via `-xtlogit-` does not allow the user to predict equation-level score variables (necessary to implement Stata's generalized Hausman test with the `-suest-` command).

<sup>156</sup> At first glance, ordered logistic regressions might appear preferable to multinomial logistic regressions to model smoking, as the three categories for the smoking variable follow an obvious ordering in terms of the degree of smoking involved. Because smoking is an addictive behavior, however, there is likely to be a qualitative difference between going from smoking zero cigarettes to one cigarette and going from one cigarette to more than five packs per year. Perhaps better would be capturing smoking using nested logistic regression models following Maddala (1983), except that fixed effects models are not currently available for this regression. This section therefore uses multinomial logistic regressions.

**Table 5.10.** Fixed and random effects multinomial logistic regressions for the effect of marriage on smoking for Waves 1-10 of the Taiwan Youth Project survey. Standard errors for parameter estimates are in parentheses.

	<b>Smoking</b>			
	Model 1: Fixed effects model		Model 2: Random effects model	
	No reported smoking <sup>a</sup>	More than five packs per year <sup>a</sup>	No reported smoking <sup>a</sup>	More than five packs per year <sup>a</sup>
Age	-4.53*** (.53)	8.75*** (.84)	-3.87*** (.50)	11.0*** (.81)
Age <sup>2</sup>	3.81*** (.51)	-6.62*** (.79)	3.15*** (.49)	-8.68*** (.76)
Married	.46 (.53)	-1.56** (.59)	.58 (.50)	-.67 (.57)
Intercept			2.10*** (.088)	-.38* (.17)
N	878		878	
Person-years	6,942		6,942	
Pseudo-R <sup>2</sup>	.483			
Log likelihood	-1,539.2		-4,208.5	

\*\*\* p < .001 \*\*p < .01 \* p < .05 + p < .10

<sup>a</sup> Referent category: smoking five or fewer packs per year.

Model 1 in Table 5.10 therefore presents a fixed effects multinomial logistic regression model of smoking on marriage, controlling for age and squared age to capture the age-crime curve (Gottfredson and Hirschi 1990). The fixed effects model is an implementation in Stata of Chamberlain’s fixed effects multinomial logit model (Chamberlain 1980:231) by Klaus Pforr (2014), and its random effects counterpart is fit using the GLAMM package for Stata (Rabe-Hesketh and Skrondal 2012). Age has also been standardized in these two models. According to the fixed effects model, respondents who are married are between 32% and 94% less likely to smoke more than five packs per year compared to five or fewer packs (using the 95% confidence interval), but no more or less likely to not smoke at all compared to smoking five or fewer packs.

The implication is that either heavy smokers are less likely to get married – while light smokers and non-smokers are equally likely to marry – or else the bonds of marriage may lead heavy smokers to cut back on their tobacco intake, but cannot completely break the addiction.<sup>157</sup>

These findings, however, are not consistent with the findings from the previous fixed effects logistic regressions in Tables 5.8 and 5.9, which would seem to imply that marriage is associated with quitting smoking entirely. Especially given that not one single married respondent to a J3 cohort survey reported smoking between one cigarette and five packs during any year, it is possible that the results for the multinomial logit model are driven by just a handful of respondents and are therefore not very reliable. The causal models that test Hypothesis 5.6 therefore make use of smoking as a binary dependent variable, and test a counterfactual hypothesis that marriage causes Taiwanese youth to desist from smoking.

Prior to modeling selection into marriage, and as a way to test the robustness of concepts from the literature review such as love typologies as predictors of marriage in Taiwan prior to their inclusion in counterfactual models, the next section seeks to predict intent to marry using Wave 6 of the TYP survey for the J3 cohort and following the models used by Cheng and

---

<sup>157</sup> The random effects model, in contrast to the fixed effects model, yields point estimates for the effect of marriage on heavy smoking compared to light smoking that are smaller in magnitude than those for the fixed effects model, and which are not significant. Similar to the analyses in Table 5.9, a Hausman test of a null hypothesis that the parameter estimates in the fixed and random effects models are the same cannot be performed, because the chi-squared statistic the test returns is negative and therefore not meaningful. Standard workarounds for negative Hausman tests also cannot be performed in the present here, because the current implementation of the fixed effects multinomial logit regression does not return a root mean squared error for model goodness of fit (necessary in order to base the variance-covariance matrices for both tests on a common disturbance variance), and the implementation of the random effects multinomial logit model using GLLAMM does not allow the user to predict equation-level score variables (necessary to implement Stata's generalized Hausman test with the `-suest-` command).

As stated previously, a reasonable course of action is therefore to favor the results from the unbiased but inefficient fixed effects model. Additionally, a Hausman test performed for OLS models with fixed and random effects rejected a null hypothesis that the parameter estimates were the same, thereby favoring the fixed effects model. And when examining the ratio of the marriage parameter coefficients for the fixed effects multinomial logit regression equation for heavy smoking to its counterpart for the random effects model, then comparing this ratio to the same ratio for the marriage parameter estimates for the linear fixed and random effects models, the two ratios are nearly the same. Given that a Hausman test cannot be conducted for the models in Table 5.10, applying the same test for the linear model is therefore a useful reference.

Landale (2011). Cheng and Landale, however, predict marriage with lagged covariates, while these models predict intent to marry with contemporaneous covariates, in order to err on the side of generating significant effects and then controlling for those covariates in a selection model to test Hypothesis 5.6. This analysis also presents an opportunity to examine how factors related to social control, social learning, and respondents' own beliefs about love may influence the intent to marry and thus selection into marriage.

### *Predicting Intent to Marry for Wave 6 of the J3 Cohort*

This section models intent to marry one's current partner among respondents to Wave 6 of the J3 Taiwan Youth Project with contemporaneous predictors,<sup>158</sup> in order to identify covariates to model selection into marriage for a test of Hypothesis 5.6. In keeping with Cheng and Landale (2011), these models control for family income, family structure,<sup>159</sup> gender, and mother's education (less than high school, high school, some college, or a college degree or more). They also include the father's education, previously shown to better predict educational attainment in Taiwan (e.g. Chu and Yu 2010:169). In particular, this analysis uses multinomial logistic regression, treating respondents as choosing between three options: not dating at all, dating with

---

<sup>158</sup> Lagged predictors would establish a clearer causal order, but at this point in the analysis the purpose of the model is to identify a set of covariates that could plausibly predict marriage, so contemporaneous predictors are used in order to observe a stronger relationship between the predictors and the outcome, maximize the available statistical power by avoiding issues concerning survey attrition, and thus err on the side of including a covariate in the later regressions. Should the later fixed effects regressions for intent to marry indicate a strong enough relationship between intent to marry and delinquency, then subsequent regressions predicting intent to marry to calculate IPTW weights would lag the covariates.

<sup>159</sup> Cheng and Landale code family structure as a categorical variable with categories that include a nuclear family, a single-parent family, two-parent families involving stepparents, and other family types. Due to data limitations, this analysis uses a binary variable for whether the child is still living with both biological parents.

the intent to marry, and dating either with no intention to marry or having not yet decided to marry.<sup>160</sup>

A number of factors concerning the relationship between the youth respondent and his or her parents predict dating with intent to marry in these models. Parental attachment is a scale comprised of two questions concerning a respondent's level of satisfaction with his or her relationship to each parent (see Appendix E). "Commitment to family activities" is a scale from the Wave 4 parent survey capturing respondents' involvement in household chores during high school, while parental religiosity is a single question asking the parent completing the survey about his or her involvement in religious activities. "Parental monitoring of sex" is a survey item concerning the extent to which respondents' parents monitor and control their ability to spend the night at a friend's house. "Parental monitoring of relationship status" is a scale based on a similar question concerning whether parents control and monitor respondents' dating behavior, and a question asked of the parent in Wave 4, "How worried are you that your child will have a boyfriend or girlfriend too soon?"

Concerning peer variables, "peer delinquency" is a variety score concerning whether any of respondents' three close friends had engaged in vandalism, theft, fraud, gambling, racing motorcycles, viewing pornographic material, drug use, and chewing betel nut (see Appendix E). "Peer sex" involves a question asking respondents to estimate what fraction of their friends had had sex at the time of the survey, and two additional questions capture respondents' beliefs about whether love is for life and whether they consider "going steady" and marriage to be different

---

<sup>160</sup> Perhaps better still would be nested logit models in which a youth is modeled as choosing to enter a relationship, and also whether he or she would like to marry that person (with the two decisions not required to take place in any particular chronological order). In practice, the nested logit model is quite complicated to fit and failed to converge in Stata for the present data, so as the purpose of the present analysis is simply to identify variables for selection into marriage, these analyses use a multinomial logistic regression and test for violations of the Independence of Irrelevant Alternatives (IIA) assumption (e.g. Heiss 2002).

concepts. Covariates concerning respondents' romantic experience include the total number of romantic partners respondents have ever had, the greatest number of romantic relationships they have ever had simultaneously, and whether they have ever cohabited with a partner.

To capture attachment and commitment to the relationship, Table 5.2 includes a direct question about how satisfied each respondent is with his or her relationship (to measure attachment) and the length of the relationship in months (to measure commitment).<sup>161</sup> As these questions are only asked of respondents in relationships at the time of the survey, they are coded as zero for respondents not in relationships. In the multinomial logistic regressions predicting intent to marry, two linear constraints then fix the coefficients for these variables to zero for the “no partner” category, so that they play no role in predicting whether a respondent has a partner.

Finally, the regressions predicting intent to marry include a categorical variable representing attitudes towards love as described by Cho and Cross (1995). Western measures of attitudes towards romantic relationships are generally comprised of several dozen Likert-scale questions about respondents' attitudes toward romantic relationships, either in general (Hendrick and Hendrick 1986) or with respect to a particular romantic partner (Hendrick and Hendrick 1990). Taiwanese clinical psychologists have adapted them to a Chinese context and tested them for internal reliability and external validity (Wang and Wang 2003), and others such as Qiu (2013) have used them to study romantic relationships among Taiwanese college students.<sup>162</sup> The TYP survey, however, only asks respondents, “Which of the following do you consider important criteria for deciding who to marry?” followed by a number of items describing characteristics of romantic partners (such as a good education or having compatible interests). Regardless of

---

<sup>161</sup> The frequency with which respondents go out on dates might be a better measure of commitment, but this information was not captured by the survey.

<sup>162</sup> Qiu's motive for using Hendrick and Hendrick's Love Attitudes Scale (1986; 1990) is to provide relationship counselors with empirical evidence about how the collectivist nature of Chinese culture shapes how participants in romantic relationships discern themselves as separate individuals from their partners.

whether they are currently dating someone, unmarried respondents are then asked to indicate which characteristics they consider important when thinking about whom they would potentially marry, checking as many boxes as appropriate. Such a design is not likely to be as reliable an instrument for measuring love attitudes as Hendrick and Hendrick's (1986; 1990), but as the purpose of the selection models for causal inference is simply to identify covariates that may affect both marriage and deviance, this chapter nevertheless makes an attempt to explain some variance in intent to marry using these survey items. I first conduct an exploratory factor analysis, described in Appendix H, to determine a set of latent factors associated with these attitudes for Wave 6 of the J3 survey, and then create a categorical variable from these scores such that each respondent is associated with the love typology on which his or her individual responses load the highest.<sup>163</sup> This categorical variable is then dummied and included in a model predicting intent to marry for that year (see Table 5.11 below). As the love attitude typologies turn out to not explain much variance in intent to marry at all, they are excluded from the later selection models for marriage.

---

<sup>163</sup> Reification of latent factors is always a concern when such procedures are used. Each respondent's answers to the survey items load onto multiple love attitude factors, and the single factor for which they are coded as 1 indicates the love typology onto which the responses load the highest. Using latent factor scores associated with these typologies would constitute a more accurate means of representing individual responses to these questions, and had these factors turned out to explain a significant amount of variability in intent to marry, factor scores would be used here to more accurately capture this measure. The current approach is simply to a fast way to estimate the variance in intent to marry that these factors are able to explain, and since in the end they explain little variability in intent to marry, they are not replaced with factor scores.

**Table 5.11.** Multinomial logistic regressions for selection into intent to marry for Wave 6 of the TYP J3 cohort (models 1-4), with robust standard errors.

	Model 1		Model 2		Model 3		Model 4	
	No partner	Intend to marry	No partner	Intend to marry	No partner	Intend to marry	No partner	Intend to marry
Female	-.53***	-.48*	-.40**	-.62*	-.43**	-.59**	-.56***	-.50 <sup>+</sup>
Family income (x 10 <sup>-6</sup> )	-1.47 <sup>+</sup>	.27	-3.10**	-.59	-1.84*	-.0041	-1.30 <sup>+</sup>	-.13
Mother's education	.19**	-.17	.087	-.21	.14*	-.16	.011	-.21 <sup>+</sup>
Father's education	-.085	-.025						
Family structure	.22	-.26	.12	-.33	.085	-.32	.052	-.18
Parental attachment			-.45***	-.41 <sup>+</sup>	-.21	-.70**	-.38**	-.42*
Parental commitment (chores)			.13	.17	.11	.0036		
Parent's religiosity			-.057	.047	.26	-.35		
Parental monitoring of			.094	.10	.17	.076	-.089	-.17
Parental monitoring of relationship status			-.36***	-.16	-.38***	-.11	-.23*	-.13
Parent's educational aspirations for child			.15*	.18				
Parental attachment x parental religiosity					-.090	.12		
Peer delinquency							-.055	-.16
Peer sex							-.49***	.024
Intercept	1.11***	-.26	2.24***	-.062	1.75*	1.89 <sup>+</sup>	3.74***	1.80*
N	1739		1298		1478		1094	
Pseudo-R <sup>2</sup>	.0165		.0355		.0322		.0572	
Log pseudolikelihood	-1,364.6		-965.1		-1,129.9		-846.4	

\*\*\* p < .001 \*\*p < .01 \* p < .05 <sup>+</sup> p < .10

(cont.)

**Table 5.11 (cont.)** Multinomial logistic regressions for intent to marry for the J3 cohort, Wave 6 (models 5-8), with robust standard

	Model 5		Model 6		Model 7		Model 8	
	No partner	Intent to marry	No partner	Intent to marry	No partner	Intent to marry	No partner	Intent to marry
Female	-.53**	-.25	-.63**	-.26	-.62***	-.89**	-.55**	-.66 <sup>+</sup>
Family income (x 10 <sup>-6</sup> )	-1.21 <sup>+</sup>	.33	-1.93**	.560	-1.83**	.40	-1.99**	.47
Mother's education	.012	-.19	-.13	-.24 <sup>+</sup>	.0028	-.026	-.0087	-.079
Family structure	.054	-.14	-.11	-.11	.033	-.30	.059	-.31
Parental attachment	-.38**	-.50*	-.46**	-.53**	-.45***	-1.14***	-.43**	-1.16***
Parental monitoring of sex	-.089	-.15	-.18	-.00022	-.15	.16	-.12	-.098
Parental monitoring of relationship status	-.24*	-.16	-.13	-.25	-.11	-.29	-.12	-.28
Peer delinquency	-.046	-.094	.043	-.094	.028	-.10	.029	-.13
Peer sex	-.49***	.014	.080	.077				
"Dating and marriage are the same concept"	-.058	-.38***	.0067	-.39***	.0042	-.46***	.0022	-.49***
"If I love someone, I will love them my whole"	.036	.41***	-.000022	.39***	.017	.46***	.033	.46***
Total number of partners			-.24*	-.026	-.21*	-.082	-.22*	-.075
Ever cohabited			-.84*	.27	-.70*	.50	-.70*	.60
Ever had sex			-.86***	.54	-1.67***	-.28	-1.67***	-.26
Max. no. of concurrent partners			-1.52***	.0092	-.61**	.30	-.59**	.23
Attachment to partner					0 <sup>a</sup>	1.95***	0 <sup>a</sup>	2.00***
Length of relationship					0 <sup>a</sup>	.028***	0 <sup>a</sup>	.027***
Eros							.040	.24 <sup>+</sup>
Storge							-.11	.13
Pragma (education)							.19**	-.22
Pragma (resources)							-.044	-.27
Intercept	3.85***	1.91 <sup>+</sup>	5.70***	1.85	5.14***	1.45	4.95***	-2.29
N		1094		1080		1720		1720
Pseudo-R <sup>2</sup>		.0840		.271		N/A <sup>b</sup>		N/A <sup>b</sup>
Log pseudolikelihood		-822.3		-643.2		-846.4		-836.0

\*\*\* p < .001 \*\*p < .01 \* p < .05 <sup>+</sup> p < .10

<sup>a</sup> Coefficient constrained to zero.

<sup>b</sup> pseudo-R<sup>2</sup> not computed due to imposition of linear constraints in the model.

Table 5.11 presents the model results, reporting robust standard errors using Huber-White sandwich estimators to correct for heteroskedasticity (Huber 1967; White 1982). The table is based on 1,833 respondents to J3 Wave 6, 1,262 of whom have no current partner, 144 of whom have a current partner and intend to marry him or her, and 427 of whom have a current partner and either do not intend to marry or are undecided. The referent category for the multinomial logit model refers to those in a dating relationship who either do not intend to marry their partner or have not decided whether they intend to marry. Actual sample sizes used for each regression in the table are smaller than 1,833 due to item non-response.

Model 1, a baseline model, contains the control variables. The fact that the mother's education is associated with less intent to marry is consistent with Cheng and Landale (2011). Cheng and Landale find that mothers of Asian Americans who have a college degree or higher have children who marry later, and Model 1 suggests that the trend may be similar in Taiwan, because children with highly-educated mothers are less likely to enter into dating relationships at a young age.<sup>164</sup> Respondents from families with low incomes are also slightly more likely to date, but not more likely to intend to marry their partners given that they are dating. Model 2 introduces variables associated with parental monitoring and respondents' attachment and commitment to parents, and, consistent with Cheng and Landale (2011), shows that a youth's attachment to parents is associated with a lower likelihood both of dating and of intending to marry for those who are dating. The coefficient for "parental monitoring of relationship status" in Model 2 also indicates that respondents whose parents want to know about their child's dating behavior have children who are more likely to be dating in Wave 6. Although half the scale is a lagged measure from Wave 4, the result may simply indicate that parents monitor their child's

---

<sup>164</sup> This fact raises an important question for future research: are the effects of parental social control on desistance from delinquency in Taiwan stronger for highly-educated parents?

dating more carefully because they are dating. Additionally, respondents with parents who expect them to go far in school tend to be less likely to date, conforming to a hypothesis that parents who emphasize education discourage dating in Taiwan. Cheng and Landale (2011) also find that parents who are more religious have children who marry earlier; accordingly, Model 2 includes parental religiosity as measured in the J3 Wave 4 parent survey, and Model 3 includes an interaction term between parental religiosity and parental attachment (testing a hypothesis that parental religiosity affects the timing of a child's decision to marry only if the child has a strong bond to the parent). In contrast to Cheng and Landale's findings, Models 2 and 3 find no effect of parental religiosity on intent to marry in Taiwan.<sup>165</sup>

Model 4 introduces the peer variables, which mediate the effect of parental monitoring of dating but not parental attachment. The dependent variable is unrelated to peer deviance, while respondents with peers having sex are more likely to date at age 20 but no more or less likely to intend to marry given that they are dating. Model 5 adds respondents' own beliefs about romance ("Dating and marriage are the same concept [to me]" and "If I love someone it will be for my whole life,") variables that explain whether respondents are likely to intend to marry their partner given they are dating, but not the decision to date in the first place. The effect of peer sex on dating still holds even given respondents' own beliefs about romance. Model 6 adds respondents' own sexual and romantic experience, including the number of partners respondents have ever had, the maximum number they have ever dated concurrently, whether they have ever had sex,

---

<sup>165</sup> The coefficients, however, are in the expected directions: the interaction term in Model 3 (which ranges from 1 for low parental attachment to 5 for high parental attachment), combined with the main effect of religiosity, implies that respondents with religious parents who are not close to them are less likely to date and less likely to marry if they date, while those with religious parents and who are close to them are more likely to date and more likely to marry if they date. A likelihood ratio test comparing Model 3 with a model without the interaction term or the main effect of religiosity finds that the additional two covariates do not significantly explain more variation in the dependent variable for dating and intending to marry – but the p value for the test is 0.12, suggesting that measurement error or perhaps biases due to survey attrition could also explain the null findings. Cheng and Landale's findings concerning the effect of parental religiosity on the timing of marriage in a Chinese context therefore probably warrant further study with another Chinese dataset in the future.

and whether they have ever lived with a partner. These personal experiences mediate the effect of peer sex, so Models 7 and 8 omit that variable as nearly half of all respondents to the survey year simply do not know what proportion of their friends have had sex, leaving the response for that covariate missing and significantly reducing the sample size. Model 7 adds characteristics about the respondent's bond to his or her partner (attachment and commitment), which explain a great deal about intent to marry.<sup>166</sup> Respondents who are satisfied with their relationships and who have been dating for some time are also more likely to marry.

Finally, Model 8 adds the latent factors for love attitudes, which turn out to explain only a little more variation in intent to marry. In particular, a chi-squared test of twice the difference in the likelihood ratios for Models 7 and 8 with eight degrees of freedom indicates that Model 8 does explain slightly more variation in the dependent variable than Model 7, but the improvement in model fit is still quite small. That said, most of the information associated with this variable has to do with the fact that respondents who are of college age are more likely to report that they are in a relationship if they value a good education as a characteristic of someone they seek to marry. It would make sense that respondents who are in college and who wish to marry a partner with a college degree would hasten to enter dating relationships while they are still in college, but this fact does not seem likely to predict selection into marriage during later waves of the study, once respondents have already graduated from college. The statistical models that follow therefore do not include love attitudes as predictors.

The strongest finding from the models in Table 5.11 is that youth who are satisfied with their relationships with their parents are less likely to intend to marry at age 20, even after controlling

---

<sup>166</sup> The linear constraints added to Model 7 make the calculation of pseudo-R<sup>2</sup> values for Models 6 and 7 impossible, but a model using the same sample as in Model 7 but removing the covariates for relationship attachment and relationship length yields a likelihood ratio of -969.0. A likelihood ratio test comparing this model to Model 6 indicates a clear improvement in model fit when the two relationship covariates are added.

for sex among peers, respondents' own sexual experiences and views about relationships, and characteristics of respondents' own romantic relationships. This finding underscores the significance to Taiwanese youth of parents' traditional views about the timing of family formation, as reflected in the literature review. Social control also proves more important than social learning for predicting intention to marry in these models, as attachment and commitment to one's romantic partner mediates the effects of peer sexuality. (Perhaps this fact is not surprising, given that the dependent variable is a measure of a person's intent to form a new social bond.) The survey, however, does not contain information about peer appraisals of the suitability of a respondent's partner, so this chapter cannot directly replicate Zhang and Kline's findings that such appraisals determine intent to marry more than personal values among Chinese (2009). Finally, the respondents' views about relationships introduced in Model 5 (e.g., "if I love someone, it will be for my whole life") support Giordano and her colleagues' argument that emotion supplies individuals with the agency through which they forge and break social bonds. Thus, despite living in an environment filled with structural factors that impact their personal decisions, Taiwanese youth still make important decisions based on their own convictions.<sup>167</sup>

---

<sup>167</sup> These findings require a technical caveat: to avoid biased parameter estimates, multinomial logit models require the "independence of irrelevant alternatives" (IIA) assumption, which states that given two choices *A* and *B* and an unavailable choice *C*, inclusion of *C* as a new available choice should not affect the likelihood of *A* to be chosen over *B* (e.g. Maddala 1983). In the case of the multinomial logit models in Table 5.2, the IIA assumption implies that the probability that a respondent states that he or she is not currently dating should always be the same regardless of the number of categories included in the model as options for respondents who are dating. A nested logit model relaxes this assumption, but is difficult to fit with the present data. In the present case, I use a Hausman-McFadden test (Hausman and McFadden 1984; Long and Freese 2014) to test the IIA assumption; for Models 7 and 8, the test fails to converge, apparently due to the linear constraints. For Model 6, the test fails to reject a null hypothesis that the assumptions are violated. The implication is that a nested logit model would be preferable to the multinomial logit model in the present case, although as the multinomial logit models in Table 5.11 are intended only for variable selection, this analysis uses the less ideal approach, trading a small amount of precision for efficiency in allocation of research time.

### *Causal Models for Marriage and Smoking Using the J3 Cohort*

The analytical methodology for the causal models in this section mirrors that presented in Chapter 4. Due to the number of covariates included in these selection models and the fact that respondents to the J1 and J3 cohorts are administered separate surveys through the 2011 survey, the test of Hypothesis 5.6 is restricted to a single cohort, and the J3 cohort is therefore used given its higher percentage of married respondents. As was the case in the previous chapter, it is also necessary to run a series of logistic regressions predicting survey attrition in the J3 cohort that is equivalent in form and substance to those in Chapter 4. This section briefly presents these models first, followed by a description of the covariates used to model selection into marriage, the selection models themselves, and the marginal structural models for causal inference.

Table 5.12 presents the logistic regressions predicting unit nonresponse for the J3 cohort. The results closely match those for the J1 cohort (Table 4.11), with school commitment and living with parents, in Wave 1 predicting a lower probability of loss to follow-up in later survey waves for both cohorts. Importantly, as was the case for the J1 cohort, deviant respondents are more likely to be lost to follow-up, meaning that not reweighting the sample to account for survey attrition would likely result in biased regression estimates.

**Table 5.12.** Logistic regressions predicting unit nonresponse in the Taiwan Youth Project survey (J3 cohort).

Variable	Wave 2		Wave 3		Wave 4		Wave 5	
	Coef	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Wave 1 <sup>a</sup>	1.53 *	9.6	-2.22 ***	9.4	-5.77	6.9	-3.45	7.2
Wave 2 <sup>a</sup>			2.65 ***	.15	.55 ***	.15	.60 ***	.17
Wave 3 <sup>a</sup>					1.35 ***	.13	1.75 ***	.16
Wave 4 <sup>a</sup>							1.43 ***	.11
Female <sup>a</sup> * J3W1	-.36 **	.13	-.10	.13	-.087	.098	-.052	.10
Taipei County <sup>b</sup> * J3W1	.13	.15	-.35 *	.16	-.31 **	.11	.40 ***	.11
Yilan County <sup>b</sup> * J3W1	-.11	.18	-.057	.19	.26 +	.14	.25 +	.14
Birth year * J3W1	-.020	.13	.0015	.13	.079	.094	.038	.097
Mother's years of education * J3W1	-.023	.027	-.0066	.028	.016	.020	-.025	.021
Father's years of education * J3W1	.015	.026	.034	.026	-.026	.019	.038 *	.019
Wave 1 family cohesion * J3W1	.27	.24	.64 **	.24	.37 *	.18	-.0073	.19
Wave 1 deviance (logged) * J3W1	-.050 **	.017	-.031 +	.018	-.0053	.013	-.018	.013
Wave 1 Live with parents * J3W1	.78 ***	.16	.73 ***	.17	.48 ***	.14	-.16	.16
Wave 1 school involvement * J3W1	-.040	.078	.087	.079	.090	.059	.081	.060
Wave 1 school commitment * J3W1	.15 ***	.035	.11 **	.035	.11 ***	.025	.085 ***	.025
Wave 1 educational aspirations * J3W1	.020	.023	.030	.022	-.0078	.016	.041 *	.017
Intercept	.92	.59	-.20	.76	-1.16 ***	.58	-2.70 ***	.63

+ p < .10 \* p < .05 \*\* p < .01 \*\*\* p < .001. Sample size for all models: 2,642.

<sup>a</sup> Denotes an indicator variable.

<sup>b</sup> Referent category: Taipei City.

(cont.)

**Table 5.12 (cont.)** Logistic regressions predicting unit nonresponse in the Taiwan Youth Project survey (J3 cohort).

Variable	Wave 6		Wave 7		Wave 8		Wave 10 (2011)	
	Coef.	S.E.	Coef	S.E.	Coef.	S.E.	Coef.	S.E.
Wave 1 <sup>a</sup>	-15.2 *	6.98	3.14	7.2	-4.34	7.30	4.91	7.0
Wave 3 <sup>a</sup>	.41 **	.14						
Wave 4 <sup>a</sup>	1.28 ***	.11	.89 ***	.11				
Wave 5 <sup>a</sup>	1.34 ***	.10	.82 ***	.11	.80 ***	.11		
Wave 6 <sup>a</sup>			1.52 ***	.10	.77 ***	.11	.69 ***	.11
Wave 7 <sup>a</sup>					1.86 ***	.11	.89 ***	.11
Wave 8 <sup>a</sup>							1.72 ***	.11
Female <sup>a</sup> * J3W1	.13	.10	-.21 *	.10	-.27 **	.10	-.27 **	.098
Taipei County <sup>b</sup> * J3W1	-.23 *	.11	.10	.12	.11	.12	-.19 +	.11
Yilan County <sup>b</sup> * J3W1	.23 +	.14	.046	.14	.18	.14	.66 ***	.14
Birth year * J3W1	.20 *	.094	-.041	.097	.057	.099	-.069	.095
Mother's years of education * J3W1	.010	.020	.025	.021	-.028	.021	.020	.020
Father's years of education * J3W1	-.018	.019	-.011	.020	.0031	.020	-.00040	.019
Wave 1 family cohesion * J3W1	.47 *	.19	.35 +	.19	.11	.20	-.15	.19
Wave 1 deviance (logged) * J3W1	-.020	.013	-.027 *	.013	-.00081	.014	.00090	.013
Wave 1 Live with parents * J3W1	.50 ***	.15	.18	.15	.16	.16	.24	.16
Wave 1 school involvement * J3W1	.082	.060	.018	.061	.043	.062	.034	.060
Wave 1 school commitment * J3W1	.0013	.024	.082 ***	.025	.067 **	.025	.027	.024
Wave 1 educational aspirations * J3W1	.031 +	.016	-.027	.017	-.0011	.017	.0092	.016
Intercept	-2.89 ***	.67	-1.69 **	.63	-1.39 *	.65	-2.21 ***	.65

+ p < .10 \* p < .05 \*\* p < .01 \*\*\* p < .001. Sample size for all models: 2,642.

<sup>a</sup> Denotes an indicator variable.

<sup>b</sup> Referent category: Taipei City.

Next, Table 5.13 lists the covariates used for selection into marriage, separating them into time-invariant covariates (the same as in Chapter 4),<sup>168</sup> predictors of delinquency (as described in the literature review in Chapter 3), and predictors of marriage (as covered in this chapter's literature review and the regression models predicting intent to marry).<sup>169</sup> These analyses include a more comprehensive measure for parental social control labeled as *guan* (literally, "control" or "management"). Respondents to the J3 Wave 7 survey were asked:

"At present, do your mother and father or male / female guardians keep track of or ask you about the following things? How strictly do they monitor you?"

- 1) Having a boyfriend or girlfriend
- 2) Your dress and overall appearance
- 3) Free time at home
- 4) Time online
- 5) Use of money
- 6) Looking for a job
- 7) Attitude towards elders
- 8) Whether you will continue your education
- 9) [For respondents in school] Studying and examinations."

---

<sup>168</sup> See Appendix E for a detailed description of the wording of the covariates used in both the Chapter 4 analysis (using the J1 cohort) and in this chapter (using the J3 cohort). The wording of the questions measuring the same concepts in the two surveys, one for each cohort, is almost identical, and the alpha reliabilities are similarly close.

<sup>169</sup> Most of these covariates have already been described either in Chapter 4 or previously in this chapter along with the models predicting intent to marry, but a few additional details are important to note. First, the J3 Wave 8 survey is a telephone interview, so a number of detailed questions about respondents' sexual behavior, attachment and commitment to school, and so on are not in that survey, and as a consequence about half the predictors for marriage in 2011 are lagged from Wave 7. This gap amounts to four years, a stretch of time that is not ideal but is also necessary in the present case. Second, for the Chapter 4 analyses, household income is interpolated from other years for survey years in which it is not available. Wave 8 does not contain questions about household income, so lagged income from Wave 7 is used instead of an interpolation based on the 2011 responses. This approach is used because the 2011 survey questions about income ask about respondents' own household's income, meaning that married respondents may include a spouse's income in their response, thereby contaminating family income as a predictor of selection into marriage were an interpolated measure based on the 2011 survey item to be used.

Third, the variable for cram school attendance is somewhat different from that used in Chapter 4. This variable denotes the number of entrance examinations for higher education, such as the GRE or Test of English as a Foreign Language (TOEFL), for which a respondent is preparing and also attends cram school. The number of such examinations is top-coded at 3, and this new variable is labeled "Cram school (for test preparation)" instead of "cram school" in the tables to follow. Respondents were also asked if they were studying for the exams but not attending cram school, and cram school attendance may be better because it represents a commitment of time and money to the test preparation and therefore is a better measure of commitment to future education. It is used to predict marriage in 2011. As it stands, however, the item may be *too* good of an indicator for delayed marriage, as in Wave 8 not a single married respondent was both preparing for an entrance examination and also attending cram school. Using that variable would therefore violate the experimental treatment assignment assumption, so the Wave 8 predictor used here is the number of entrance examinations for which a respondent is preparing regardless of whether he or she is attending a cram school, with the measure also top-coded at 3.

Responses to the item are on a four-point Likert scale with levels that include “not at all,” “they talk about it but do not monitor me,” “they monitor me but not strictly,” and “they monitor me strictly.” Together these items comprise a one-dimensional scale with an alpha reliability of 0.86.

**Table 5.13.** Covariates used for the selection models into marriage, for each survey wave used for the marriage analysis. In the wave labels, “J3” refers to the J3 cohort, and “W” followed by an integer denotes the wave number.

<b>Variable</b>	<b>J3W7</b>	<b>J3W8</b>	<b>J32011</b>
<b>Time-Invariant Covariates</b>			
Respondent’s gender	J3W1	J3W1	J3W1
Area (Taipei city, Taipei county, Yilan county)	J3W1	J3W1	J3W1
Respondent’s birth year	J3W1	J3W1	J3W1
Mother’s education	J3W1	J3W1	J3W1
Father’s education	J3W1	J3W1	J3W1
<b>Time-Varying Covariates (Delinquency Predictors)</b>			
Household income	J3W6	J3W7	J3W7
Educational tracking	J3W6	J3W7	J3W8
School / teacher attachment	J3W6	J3W7	J3W7
School commitment: time spent on homework	J3W6		
School commitment: grades	J3W6	J3W7	J3W7
Parental attachment	J3W6	J3W7	J3W8
Parental monitoring	J3W6		J3W6
<i>Guan</i>	J3W6	J3W7	J3W7
Live with parents	J3W6	J3W7	J3W8 <sup>b</sup>
Peer deviance	J3W6	J3W6	J3W6
Prior deviance	J3W6	J3W7	J3W7
Stressful life events	J3W6	J3W7	J3W7
If have student loans	J3W6	J3W7	J3W7
Cram school		J3W7 <sup>a</sup>	J3W8
Parental aspirations for child’s educational attainment	J3W6	J3W7	J3W7
Depression	J3W6	J3W7	J3W8
How often online	J3W6	J3W6	J3W6
Materialism	J3W6	J3W6	J3W6
<b>Time-Varying Covariates (Marriage Predictors)</b>			
Peer sex	J3W6	J3W6	J3W6
“Going steady and marriage are the same”	J3W6	J3W6	J3W6

(Continued on next page)

**Table 5.13 (cont.)** Covariates used for the selection models into marriage, for each survey wave used for the marriage analysis. In the wave labels, “J3” refers to the J3 cohort, and “W” followed by an integer denotes the wave number.

<b>Variable</b>	<b>J3W7</b>	<b>J3W8</b>	<b>J32011</b>
<b>Time-Varying Covariates (Marriage Predictors)</b>			
Self-esteem		J3W7	J3W7
Autonomy	J3W6	J3W7	J3W8
“If I love someone, it will be for my whole life”	J3W6	J3W6	J3W6
Cohabitation	J3W6	J3W7	J3W7
Number of concurrent sexual partners ever had	J3W6	J3W7	J3W7
Currently in a relationship	J3W6	J3W7	J3W8
Attachment to current relationship	J3W6	J3W7	J3W8
Length of current relationship	J3W6	J3W7	J3W8
Ever had sex	J3W6	J3W7	J3W7
Anticipated age of first marriage	J3W6	J3W7	J3W8
Intention to marry current partner	J3W6	J3W7	J3W8
Currently working	J3W6	J3W7	J3W8

<sup>a</sup> No married respondents to Wave 8 attended cram school in Wave 7, so the Wave 7 “cram school” predictor is the number of graduate school entrance examinations for which the respondent is studying that year and may or may not attend cram school as part of the preparation process.

<sup>b</sup> “Live with both parents” is actually “live at home” for J3 Wave 8.

One disadvantage of using the J3 cohort is the three-year lag between Wave 8 and the 2011 survey. The lag between Wave 9 for the J1 cohort and the 2011 survey is a single year, so the lagged J1 covariates should better predict selection into marriage in 2011. Also, a few J3 cohort survey items involving attitudes towards dating, marriage, and peer sexual behavior are only asked in Wave 6 and also significantly predict marriage in 2011, so they must be included in the model with a five-year lag. Needing to include predictors from Waves 6, 7, and 8 in modeling marriage in 2011 considerably reduces the sample size for the causal model for that year and makes the inverse probability weights to correct for survey attrition more important. While unit nonresponse due to attrition remains a problem that must be addressed for the causal models in

this chapter, item nonresponse is less serious than in Chapter 4, with a response rate of more than 93% for all waves (see Appendix I for a discussion of item non-response for the selection models for marriage). The multiple imputation methods used there are therefore omitted here.

In Table 5.14, lagged deviance uses the Expected a Posteriori (EAP) scores for the continuous IRT measures derived in Chapter 2 and used to test Hypotheses 5.1 and 5.3. As the EAP scores for respondents in some transformed scales may be negative, a scalar value is added to the scores for all waves to make them all positive, and then they are logged. Additionally, the predictors on the second page of the table, which apply to romantic relationships, are nearly all interaction effects. Most of these predictors only make sense in the context of predicting marriage for unmarried respondents or for respondents who were dating when interviewed. To avoid loss of sample size for respondents who legitimately skip the questions, these predictors are coded to an arbitrary value (9999) for those who are not qualified to respond, and the variables in the regressions are then interactions between each question and indicator variables for whether the respondent is qualified to respond. Occasionally these interactions are three-way interaction effects (such as in the case of cohabitation, for which respondents must both be unmarried and have a current partner). Interpretation of the interaction term for cohabitation, for example, would be the effect of cohabitation for unmarried respondents with a current partner.<sup>170</sup>

---

<sup>170</sup> A few covariates that are theoretically relevant cannot be included as predictors because they perfectly predict marriage. Tracking is not included for Wave 6 because every respondent in the academic track in Wave 6 was still unmarried in Wave 7, so an indicator variable for being in school is used instead. Similarly, as every respondent who was married in Wave 7 was still married in Wave 8, marriage is lagged by two waves in predicting marriage in Wave 8. Additionally, the inclusion of lagged predictors from three waves reduces the available sample size to just 912, so to recover some cases the model leaves out several covariates that are not significant predictors of marriage (with p values greater than .3) and which collectively do not explain significantly more variation in the dependent variable. For Wave 7, these covariates include school attachment, school commitment, stressful life events, school loans, cohabitation, maximum number of simultaneous partners, and whether the respondent had ever had sex; for Wave 6, they include materialism, internet use, and attitudes towards marriage (“dating and marriage are the same” and “if I love someone, it will be for my whole life”); and for Wave 8, they include depression, autonomy, and currently having a job (net of the educational tracking variable corresponding to not being in school). Excluding these covariates recovers 19 respondents.

**Table 5.14.** Logistic regressions predicting marriage in the Taiwan Youth Project survey (J3 cohort).

Variable <sup>1</sup>	Wave 7			Wave 8			Adult Wave (2011)		
	Coef.		S.E. <sup>a</sup>	Coef.		S.E. <sup>a</sup>	Coef.		S.E. <sup>a</sup>
Lagged marriage	7.69	***	1.23	5.13 <sup>h</sup>	**	1.98	5.51	***	.93
Female <sup>b</sup>	2.18	*	1.08	1.65	*	.66	.61		.38
Taipei County <sup>b,d</sup>	-1.30		1.11	.62		.92	.88	*	.42
Yilan County <sup>b,d</sup>	-.038		.89	.11		1.02	.91	*	.42
Age	1.19	+	.70	.52		.63	.43		.28
Mother's years of education	-.24	*	.12	-.095		.12	.074		.066
Father's years of education	-.025		.12	-.12		.12	-.074		.068
Household income <sup>c</sup>	-1.34	+	.70	.11		.41	-.020 <sup>h</sup>		.26
Vocational track <sup>b,e</sup>				-1.01		1.73	.89		.89
"Not in school" track <sup>b,e</sup>				9.73		9.17	1.2	+	.65
In school <sup>b</sup>	-3.37	***	.82						
School attachment * in school				2.2		1.5			
School commitment * in school				-.47		.32			
Have a student loan * in school				.97		2.0	-.41 <sup>h</sup>		.42
Test preparation				-1.3	*	.59			
Cram school (for test preparation)							-.79	*	.39
Parental attachment	1.91	*	.76	1.25	+	.65	.30		.30
Parental monitoring	-1.43	**	.47				.31 <sup>i</sup>		.21
<i>Guan</i>	-1.04		1.1	-1.14	*	.48			
Living with parents <sup>b</sup>	-.21		.93	-.60		.66	-.36 <sup>j</sup>		.36
Peer deviance <sup>c</sup>	-.11		.084	-.072 <sup>h</sup>		.080	-.032 <sup>i</sup>		.050
Prior deviance <sup>c</sup>	.62	+	.37	-.033		.25	.10		.10
Number of stressful life events <sup>c</sup>	-.094		.11	.22	+	.12			
Self-esteem				-.28		.71	.39		.38
Autonomy	.26		.41	.56		.51			
Depression <sup>c</sup>	.14		.18	-1.90	+	1.10	.020		.052
Materialism	.095		.65	-.42 <sup>h</sup>		.40			
N	1,438			1,209			931		

+ p < .10 \* p < .05 \*\* p < .01 \*\*\* p < .001.

(continued on next page)

**Table 5.14 (continued).** Logistic regressions predicting marriage in the Taiwan Youth Project survey (J3 cohort).

Variable	Wave 7		Wave 8		Adult Wave (2011)	
	Coef. <sup>a</sup>	S.E.	Coef.	S.E.	Coef.	S.E.
Know if peers are having sex <sup>b</sup>			-.57 <sup>h</sup>	.75	-.26	.40
Fraction of peers having sex * know that fraction <sup>b</sup>			.82 <sup>h</sup> **	.31	.23	.18
“Dating and marriage are the same [to me]”	.033	.26	-.012 <sup>h</sup>	.19		
“If I love someone, I will love them all my life”	.19	.23	.10 <sup>h</sup>	.28		
Unmarried <sup>b</sup> * currently dating <sup>b</sup>	.56	3.5	-7.6	4.4	2.5	1.4
Cohabitation <sup>b</sup> * unmarried <sup>b</sup> * currently dating <sup>b</sup>	.27	.83	1.8	.81		
Unmarried <sup>b</sup> * maximum number of concurrent partners respondent has ever had	-.69	.64	-1.3	.51	.30 <sup>h</sup>	.33
Attachment to partner * single <sup>b</sup> * currently dating <sup>b</sup>	-1.1	.70	1.7	1.2	-.24	.38
Length of relationship * single <sup>b</sup> * currently dating <sup>b,c</sup>	.35	.43	-.28	.39	-.091	.20
Ever had sex <sup>b</sup> * unmarried <sup>b</sup>	.63	.80	-.47	.97		
Want to marry in future <sup>b,g</sup> * desired age of marriage coded as number of years in future <sup>k</sup>	-.099	.12	-.32	.14	-.026	.046
Want to marry partner <sup>f</sup> * single <sup>b</sup> * currently dating <sup>b</sup>	1.7	.91	1.4	.83	.56	.24
Currently working <sup>b</sup>	.16	.94	-.49	.66		
Time spent online <sup>c</sup>			-.048 <sup>h</sup>	.072		
Intercept	-22.3	15.1	-23.9	18.7	-18.0	7.1
N	1,438		1,209		931	
Pseudo R <sup>2</sup>	0.663		0.581		0.367	

\* p < .10 \*\* p < .05 \*\*\* p < .01.

<sup>a</sup> Standard errors are robust Huber-White sandwich estimators.

<sup>b</sup> Denotes an indicator variable.

<sup>c</sup> Denotes a variable that has been logged.

<sup>d</sup> Referent category: Taipei city.

<sup>e</sup> Referent category: Academic track.

<sup>f</sup> Three-level ordinal variable with levels for “no” (0), “undecided” (1) and “yes” (2).

<sup>g</sup> This indicator variable is coded as 0 for respondents who are already married.

<sup>h</sup> Denotes a variable that has been lagged by two waves.

<sup>i</sup> Denotes a variable that has been lagged by three waves.

<sup>j</sup> “Living with both parents” is really “living at home” for Wave 8.

<sup>k</sup> Recoded to zero years in the future for a handful of unmarried respondents who state ages corresponding to times in their past.

<sup>l</sup> All covariates are lagged by one survey wave unless otherwise indicated.

Consistent with the literature, Table 5.14 indicates that education is associated with delayed marriage.<sup>171</sup> First, respondents who are in school are much less likely to be married during the next survey year. Comparing a respondent in the academic track in Wave 8 to a respondent serving in the military or already in the work force (“not in school” track), for example, the respondent not in school is 232% likelier to marry. Secondly, those who intend to pursue a graduate education are also much less likely to marry. The panel for Wave 8 indicates that each additional standardized test for which a respondent is preparing in Wave 7 is associated with a reduction in the odds of marriage of 72.7% on average. Similarly, the rightmost panel indicates that for each additional entrance examination for which a respondent attends cram school, the odds of being married three years later decrease by 54.6%. Parental monitoring, as measured either through Hirschi’s traditional measure, “Do your parents know where you are when you are away from home?” or through the Chinese concept of *guan*, predicts a decreased likelihood of marriage in Waves 7 and 8, but has no effect for marriage in 2011.<sup>172</sup> Prior deviance predicts marriage in Wave 7 but not in the later waves, suggesting that perhaps marriage is becoming less of a deviant behavior as respondents age.

Not surprisingly, in all three regressions the desire to marry one’s current partner is associated with an increase in the likelihood of marriage in the next wave. For Wave 8, peer sex and cohabitation, but not whether the respondent had ever had sex, predict a greater propensity to

---

<sup>171</sup> In Table 5.14, due to the relatively small number of married respondents compared to marriage in 2011, some of the large effect sizes could be driven by a few cases and might be taken with a grain of salt.

<sup>172</sup> The fact that the regression for marriage in 2011 generally contains fewer significant predictors could indicate that the concepts reflected by those variables are not predictive of marriage at that time, but could also indicate that the three-year lag between the measurement of these covariates and the observation of marriage is attenuating the magnitude of their effects on marriage. Additionally, the fact that several variables concerning sex and romantic relationships in Wave 7 are significant in predicting marriage in Wave 8 and not significant in other waves could be due to the two-year lag for marriage as a predictor for the middle panel, thereby potentially leaving more variance in marriage for Wave 8 to be explained by the other predictors. (When lagged marriage is removed as a predictor from the rightmost panel, other predictors including peer sex and the desired age of marriage in the future become significant.)

marry. It may be that absent a strong measure of how often a person has sex, peer sex explains additional variance in the propensity to marry because it predicts a respondent's own sexual behavior due to social learning factors, and those who have frequent sex may marry sooner. Yet despite the association of peer sex with increased odds of marriage, having had multiple concurrent partners in Wave 7 is associated with a *decreased* propensity to be married by the following wave. If the peer sex variable is associated with the respondent's own, unobserved frequency of sex, then it could be that it is *monogamous* sexual behavior that predicts marriage. Additionally, for unmarried Wave 7 respondents who state the age at which they would like to marry, each additional year in the future in which they would like to marry is associated with a reduction of 27.4% in the odds of being married by Wave 8 (two years in the future). This finding further supports findings that intentions generally predict behavior in life course surveys of youth (Fischhoff et al. 2000). Interestingly, even though contemporaneous parental attachment was associated with a decreased propensity to intend to marry one's partner in the regressions for Wave 6 presented previously, here lagged parental attachment, net of parental monitoring, predicts a *greater* propensity to marry.<sup>173</sup>

With propensity to marry and survey attrition modeled, the stage is now set for the causal models. Following Sampson, Laub, and Wimer (2006), this chapter uses a hierarchical generalized linear model for the causal regression, such that level 1 of the hierarchical model

---

<sup>173</sup> Note that the parental attachment variables used in Tables 5.11 and 5.14 are slightly different; Table 5.11 uses questions concerning whether respondents are satisfied with their relationships with their mother and father figure. For parental attachment in Waves 6 and 7, Table 5.14 in contrast uses questions about whether the respondent's parent cares about him or her, provides them with emotional and financial assistance, helps them to solve problems, and understands them – with separate questions for the mother and father that are combined into a single scale. Hirschi (1969) makes the case that parental attachment should be measured from the perspective of the child, using questions about what the child thinks about the parents rather than vice versa, in order to capture whether the parent is psychologically present when the child is presented with an opportunity to offend. Perhaps the measure used in Table 5.11, then, adheres more strictly to the notion of parental attachment. That said, replacing the measures used in Table 5.14 with variables asking respondents how satisfied they are with their relationships with their parents does not change the significance and direction of the parental attachment variables, and does not affect the findings in the causal models either.

captures within-individual change and level 2 captures random differences between individuals in the level of smoking. The functional form for the hierarchical model is a logistic regression, with marriage as a level-1 predictor and no predictors at the second level (only a constant term, with random disturbances to account for differences between individuals in the level of smoking). The IPTW weights added to the model at level 1 then account for bias due to attrition, age and squared age, and the observed confounders, including lagged marriage, included in the selection models for marriage.<sup>174</sup> Table 5.15 presents the unit-specific and population-average causal effects of marriage on smoking for respondents to Waves 7, 8, and the 2011 survey for the J3 TYP cohort. Only the model with stabilized weights is presented, as the unstabilized weights vary quite dramatically, with the largest weight corresponding to more than 1,000 individual respondents.<sup>175</sup>

---

<sup>174</sup> Calculation of the weights uses Equation 4.3, with an important twist. Regardless of the particular year for which the causal effect of marriage on smoking is being modeled, every person-year of data used in the model is constructed in part from data obtained from the 2011 (“Wave 10”) survey that contained the retrospective smoking questions. The set of survey respondents used to model the causal effect of marriage in Wave 7, for example, is actually those respondents who responded to three waves: Wave 7 (to provide information about marriage), Wave 6 (from which lagged predictors are drawn for the IPTW weights), and Wave 10 (in which respondents are asked about smoking during Wave 7). The inverse probability weight used in the causal model for marriage and smoking for Wave 7, then, is the product of the IPTW weight for Wave 7 and the inverse of the missingness weight for Wave 10. Similarly, the Wave 8 treatment and Wave 10 missingness weights are inverted and multiplied to compute the weights for the Wave 8 causal model, and the Wave 10 treatment and missingness weights are multiplied and inverted for the Wave 10 causal model weights.

<sup>175</sup> These large values are observed even after truncating the 1% of weights with the highest values to the 99<sup>th</sup> percentile, and similarly recoding the 1% of weights with the lowest values to the first percentile following Cole and Hernán (2008).

**Table 5.15.** Counterfactual random effects logistic regressions for the effect of marriage on smoking for Waves 7, 8, and 2011 of the TYP survey, J3 cohort, using inverse probability weighting for unit nonresponse and selection into marriage. Standard errors for parameters are in parentheses.

	<b>Smoking</b>	
	Unit-Specific Model	Population-Average Model
Married	-.465 (.38)	-.256 (.23)
Intercept	-1.68*** (.10)	-1.22*** (.074)
N	<i>1008</i>	<i>1008</i>
Person-years	2,745	2,745
Likelihood	-3,418.1	-3,583.2

\*\*\* p < .001 \*\*p < .01 \* p < .05 + p < .10

Table 5.15 indicates no causal effect of marriage on smoking. The “unit-specific” model coefficients represent the change in the propensity to smoke associated with the counterfactual of taking an unmarried person and causing him or her to be married instead, holding the random effect for that person fixed to zero for that wave. The coefficient thus corresponds to the effect of marriage on smoking for an “average” person (one with a random effect equal to zero). “Population-average” coefficients are estimated without the random effect included, and model coefficients therefore represent average effects of the model covariates at the population level. The first model therefore answers the question, “what happens to the average single Taiwanese youth’s smoking behavior if an exogenous force were to somehow cause him or her to be married?” The second model answers the question of what would happen to the proportion of smokers in a group of young single Taiwanese people if an outside force such as a government

policy were to cause them all to become married. In both cases, marriage does not change smoking behavior.<sup>176</sup>

### Conclusion

Through an analysis of the effects of marriage and intent to marry on smoking and deviance, this chapter has helped to shed light into the nature of turning points in the life course for Taiwanese. The main conclusion is that bonds with romantic partners may not as important as those to parents or to school during early adulthood in Taiwan: the simple intent to marry one's partner is unrelated to deviance, and while marriage is associated with a reduction in smoking, getting married does not cause a person either to stop smoking or to start. This finding, however, requires a caveat. As an addictive behavior demonstrated by the correlation analyses in Table 5.7, smoking may be more difficult for a turning point in the life course to influence than other forms of deviance, and so future researchers would do well to examine the causal effect of marriage on other types of delinquency.

The chapter also presents two smaller findings. First, youth who date tend to be more delinquent (although these findings were only preliminary and designed only to explain why the models for selection into marriage make use of covariates that predict delinquency). Second and more importantly, in describing intent to marry in order to set up selection models for marriage, Chapter 5 also provides support in Taiwan for Cheng and Landale's findings (2011) that Asian-American youth who have a strong attachment to their parents are less likely to marry at an early age, even controlling for a wide variety of other factors including individual beliefs about the

---

<sup>176</sup> It is possible that the null findings could be produced by the relatively low within-individual variability in smoking and the modest number of married respondents, such that bringing the J1 cohort into the analysis could produce different results. Fitting those models, however, is reserved for future work.

importance of marriage, individual sexual behavior and peer sexual behavior.<sup>177</sup> Consistent with the literature, these findings suggest a role for family social control in influencing the timing of marriage. Together with the findings in Chapter 4, these results have a number of implications concerning the importance of social control in Taiwan and how social norms in Chinese culture affect the timing of turning points in the life course and their subsequent impact on deviance. These implications are covered in the conclusion for this dissertation, which immediately follows.

---

<sup>177</sup> Note that the findings here are somewhat weaker than those of Cheng and Landale (2011), as they predict marriage using lagged covariates, while the models predicting intent to marry in this chapter were intended as a means for choosing explanatory variables for selection models for marriage, so the covariates used were intentionally kept as contemporaneous predictors. Doing so likely increased the strength of their relationship with marriage while creating the possibility that the predictors are actually caused by marriage and not the other way around.

## Chapter 6: Conclusion

This dissertation contributes to the literature on turning points in the life course by showing how work and stable romantic relationships affect deviance for Taiwanese youth. The findings underscore the importance of social control in influencing individual behavior in Chinese culture. Using causal models for counterfactual inference, the research also sheds light on the extent to which social policies that influence the timing of work and marriage can influence deviant behavior. This final chapter discusses the substantive and methodological contributions of this work, as well as its limitations and opportunities for future research on which it sheds light. Following the convention of the Chinese publications cited herein, this dissertation also concludes with a brief discussion of the policy implications of the research.

### Contributions

Driven by characteristics of the Taiwan Youth Project data set introduced in Chapter 1, Chapter 2 of the dissertation addressed the problem of measuring crime in a longitudinal survey with changing items. In so doing, the chapter makes several methodological and substantive contributions to the field of criminology. First, Chapter 2 quantifies the relative seriousness of various types of delinquency in Taiwan. Second, it builds upon Osgood, McMorris, and Potenza's initial application of Item Response Theory (IRT) models to the study of deviance, demonstrating how vertical scaling, a technique usually used to compare test scores, can allow criminologists to model changes in individual deviance in longitudinal surveys in which the survey items change. Third, because so many respondents in Taiwan report committing no crimes at all, maximum likelihood estimation of IRT model parameters was infeasible due to violations of model assumptions, and Chapter 2 solves this class of problems for criminology

research through an application of Bayesian IRT models to criminological research for the first time. Fourth, it presents an algorithm for improving the precision of vertical scaling techniques in Bayesian research settings, a procedure that has wider utility outside of sociology.

Chapter 3 reviewed the literature on the delinquency in Taiwan, showing strong support for social learning and social control theory and weaker support for strain theory, social disorganization theory, and Gottfredson and Hirschi's general theory of crime. The chapter also emphasized the importance of educational tracking to explaining life course outcomes for youth. Chapter 4, the second empirical chapter, then investigated the relationship between part-time work and deviance for Taiwanese youth of high school and college age, and makes three contributions – one methodological and two substantive. First, in showing a significant causal relationship between part-time work and deviance, it demonstrates the utility of the IRT model in Chapter 2 for producing measures of crime with a low enough measurement error for viable empirical research. Second, the chapter demonstrates a positive relationship between work and deviance among Taiwanese youth of high school and college age, but finds no evidence of a causal effect of work on deviance – that is, until it is interacted with parental monitoring, in which case work is shown to cause deviance for youth whose parents do not keep track of their whereabouts, yet does not lead to delinquent behavior for youth who are closely monitored by their parents. In other words, part-time work in Taiwan raises crime, but only for the unsupervised. Third, Chapter 4 provides little to no evidence that Taiwanese youth who express agency – that is, those who work only a few hours per week but for long periods of time – are less deviant than youth with other work patterns. In Taiwan, youth with no jobs are the least deviant, a fact differs from findings in the West, in which youth with no job and youth with

steady jobs commit crimes at approximately the same rate. Collectively, the findings reinforce the importance of social control as opposed to agency in Taiwan.

Chapter 5 then examined the relationship between steady romantic relationships and deviance in Taiwan, covering dating (briefly), intent to marry one's romantic partner, and marriage, with four contributions. First, preliminary analyses of the TYP data touched upon during the literature review at the beginning of Chapter 5 suggest that Taiwanese who are of college age and date tend to also be more deviant. These findings are worth following up in future research to determine if dating causes an increase in deviance or if the relationship is spurious. To use examples of causal mechanisms from social learning theory and social control theory (which the literature review in Chapter 3 indicated explain the most about deviance in Taiwan), if exposure to deviant peers during the dating process explains the increase in deviance associated with dating, then the process of dating likely causes the increase. But if in an environment of strong social control and emphasis on schooling a prevailing cultural norm is for students to not date prior to completing their studies, then to some extent the very act of dating could be deviant, and hence dating and general deviance would be spuriously related.

Second, in describing intent to marry in order to set up selection models for marriage, Chapter 5 also replicates findings from Cheng and Landale (2011) that youth who have a strong attachment to their parents are less likely to marry at an early age, even controlling for a wide variety of other factors including individual beliefs about the importance of marriage, individual sexual behavior and peer sexual behavior. This finding echoes one of the strongest and most consistent findings in the literature on deviance in China and Taiwan, that youth who have strong bonds with their parents are less deviant. Third, fixed effects models indicate that the simple intent to marry one's partner is unrelated to deviance, while those who actually marry are less

likely to smoke. But this last finding is not borne out by the counterfactual models for causal inference, which find that marriage does not cause any change in smoking behavior.

Together, Chapters 4 and 5 also drive home the importance to the Taiwanese of going through life course transitions in the “right” sequence and at culturally-appropriate ages. In Taiwan, young people are expected to achieve a good education first before proceeding to marry and enter the work force. This general picture was reinforced by a variety of findings throughout the dissertation. Attachment and commitment to school not only is associated with lower crime, but it also is associated with a deferral of work and marriage to later ages. The more that respondents to the Taiwan Youth Project survey like school and the higher their grades, the less likely they are to work part-time while in high school and college, and youth who are still in school or preparing for graduate study are also much less likely to marry. Additionally, the reason that working does not appear to cause deviance for youth whose parents know where they are may have to be that the youth in these situations may have secured permission from their parents to “put the cart before the horse” in terms of the timing of these life-course transitions, and in so doing continue to perceive themselves as “good kids” and are less prone to pressure from deviant peers in their workplace environment.<sup>178</sup>

### Limitations and Future Research

This dissertation thus makes several substantial contributions to the literature on the measurement and causes of crime, especially with respect to Asian culture, but also has several limitations, some of which raise several important questions in the process. First, the marginal

---

<sup>178</sup> The literature review on the determinants and consequences of youth employment in Taiwan did not uncover any discussions of when and how parents express approval for their children working, aside from indicating that some youth work to earn income to support their families. That line of research would be valuable for teasing out some of the causal mechanisms producing the results in Chapter 4.

structural models used in this dissertation also have several technical limitations. They would benefit from empirical checks to verify that the experimental treatment assignment (ETA) assumption was not violated in any of the models (Petersen, Porter, Gruber, Wang, and van der Laan 2012; Wang, Bangsberg, Petersen, and van der Laan 2006). Second, additional outlier diagnostic plots would have made certain that none of the findings herein were a result of influential outliers. Third, estimates of treatment effects in marginal structural models can be sensitive to the specification of the selection model, so additional sensitivity checks such as the use of estimation techniques using covariate balancing propensity scores (Imai and Ratkovic 2014) would serve as additional useful robustness checks. Finally, this research implicitly assumes that the survey waves are spaced apart at equal intervals, when in fact they were administered at different times during each year and were not always spaced one year apart; statistical techniques exist to correct for this problem (Chow and Zhang 2008; Oud and Jansen 2000; Oud and Singer 2008) but were not used here.

The particular data set used here is a precious resource for researchers studying turning points in the life course in Taiwan, but nonetheless has its limitations as well. First, the models in this dissertation all involve implicit assumptions that respondents are telling the truth about their delinquent behavior. The literature review in Chapter 2 suggests that this assumption is not unreasonable, but the approaches used here would nonetheless have more support if future research in Taiwan were to assess the veracity of self-reported delinquency using record checks. An important question that this research addresses only addresses obliquely with the models of work intensity and extensity in Chapter 4 is the importance of individual agency in affecting deviance. The Taiwan Youth Project survey, like similar longitudinal surveys in the U.S., contains a plethora of questions about social bonds with respect to one's parents, family,

siblings, grandparents, classmates, teachers, neighbors, dating partners, and spouse. Aside from when respondents would like to marry and what they are looking for in a partner, the survey contains relatively few questions about individual agency that involve specific turning points in the life course. Additional survey items in this domain that would be helpful in future longitudinal studies include questions about whether respondents are considering changing educational tracks or their majors and if so, why; questions about their career plans and whether their part-time jobs are related to these goals; consistent and detailed questions about self-control, self-efficacy, and individuating and relational autonomy; and questions about other people in respondents' lives who affect their decisions to work and marry along with the weight that respondents assign to the opinions of those people when making their own decisions. Additional limitations of the data include the lack of questions concerning the neighborhoods in which respondents to the later survey waves live, and the inconsistency of the survey covariates and three-year interval between Wave 8 and the first adult wave for the J3 cohort that lead to long lags in the selection models for marriage.

The purpose of this dissertation is to establish causal inferences concerning how work and marriage affects deviant behavior for Taiwanese youth, and this goal has been achieved. Much still is unknown, of course, concerning some of the causal mechanisms at play. Using the example of the teenager in high school who has a job but whose parents know where they are when they are away from home, and who thus does not become more deviant as a result of working, it may be that parental monitoring directly protects the child from delinquent behavior – but it also may be that the child has been actively keeping the parent in the loop as part of a concerted effort to get as much help and advice as possible while making decisions about a

possible future career and balancing schoolwork at the same time. These two causal mechanisms of course have quite different policy implications.

Concerning the measurement of crime, the TYP survey regrettably omitted questions about general deviance from the 2011 survey, preventing an analysis of the causal effect of marriage on general deviance and restricting the present research to smoking. Should funding be secured for future waves of research, the survey team would do well to retain the general deviance questions, given the availability of the methods presented here for linking them to previous waves. They would also be free to include new measures of deviance that reflect crimes committed by adults (such as credit card fraud and so on), as long as some items used in previous waves are retained and measured using the same frequency scales. Such a design would allow for an empirical analysis of the causal effect of marriage on crime in a Chinese sample for the first time.

### *Policy Implications*

The causal analyses in this dissertation have some implications for policy makers, who have to make counterfactual inferences. The main implications for policymakers and parents alike in Chapter 4 seem to be that parental supervision may mitigate the criminogenic effects of part-time work for Taiwanese youth who are in their high school and college years. Alternatively, if parental monitoring must be low, then steering youth of these ages away from the workplace would lower their propensity to offend. Concerning marriage, the implications are less clear because the findings are weaker. The evidence available suggests that marriage neither raises nor lowers smoking, but the number of married respondents in the analyses presented in Chapter 5, along with only a modest amount of within-individual variation in smoking (an addictive

behavior), implies that studies involving a larger sample of married respondents or a different dependent variable for deviance may produce different findings.

A reader might wonder what the importance is of a counterfactual inference that uses marriage as a treatment variable. After all, no external policy could take an unmarried person and transform him or her into a married person, and vice versa. Yet social policies can and do influence marriage rates. As mentioned in the introduction, one purpose of this dissertation is to inform social science researchers in Mainland China of the costs and benefits of implementing longitudinal surveys of self-reported deviance there. For these researchers, the counterfactual questions concerning marriage and deviance raised in this dissertation are directly relevant to questions about the potential social impact of policies in China that restore the gender ratio to its natural value, thereby raising marriage rates for males. Moreover, a sample of ethnic Chinese respondents from Taiwan is culturally similar to a sample from Mainland China. As such, the findings from Chapter 5 are quite relevant for Chinese researchers.

This dissertation contributes to a body of social science research concerning how changes in the life course affect crime and deviance in Chinese culture. Despite the limitations of the data and methods, it sheds light on several important factors concerning the importance of social control and the measurement of crime in China. In so doing, it increases our knowledge about turning points in the life course in Taiwan. As Taiwan is a setting in which marriage and work are increasingly postponed in favor of higher education, the findings in this dissertation, while not directly generalizable to other contexts, do challenge our assumptions about the timing and effect on deviance of these turning points in the life course for young adults elsewhere. It is this author's hope that future researchers can further these lines of investigation, in order to make both Taiwan and Mainland China more prosperous and crime-free in the years ahead.

## BIBLIOGRAPHY

- Academia Sinica Institute of Sociology. 2011. "Taiwan Youth Project." Taipei: Academia Sinica Institute of Sociology. Retrieved from <http://www.typ.sinica.edu.tw/E/?q=node/8> on April 15<sup>th</sup>, 2015.
- Agnew, Robert. 1992. "Foundation for a General Strain Theory." *Criminology* 30:47-87.
- Agnew, Robert. 2006a. "General Strain Theory: Current Status and Directions for Further Research." Pp. 101-123 in *Taking Stock: The Status of Criminological Theory, Advances in Criminological Theory*, edited by F. T. Cullen, J. P. Wright, and K. R. Blevins. New Brunswick: Transaction Publishers.
- Agnew, Robert. 2006b. *Pressured Into Crime: An Overview of General Strain Theory*. Los Angeles: Roxbury.
- Ajzen, Icek and Martin Fishbein. 1977. "Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research." *Psychological Bulletin* 84:888-918.
- Akers, Ronald L. 1998. *Social Learning and Social Structure: A General Theory of Crime and Deviance*. Boston: Northeastern University Press.
- Akers, Ronald L. and Gary F. Jensen. 2006. "The Empirical Status of Social Learning Theory of Crime and Deviance: The Past, Present, and Future." Pp. 37-76 in *Taking Stock: The Status of Criminological Theory, Advances in Criminological Theory*, edited by F. T. Cullen, J. P. Wright, and K. R. Blevins. New Brunswick: Transaction Publishers.
- Allison, Paul D. 2009. *Fixed Effects Regression Models*, vol. 160. Thousand Oaks, CA: SAGE.
- Allison, Paul D and Richard Waterman. 2002. "Fixed Effects Negative Binomial Regression Models." *Sociological Methodology* 32:247-265.
- Anderson, Elijah. 1999. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. New York, NY: W.W. Norton & Company, Inc.
- Apel, Robert, Shawn Bushway, Robert Brame, Amelia M. Haviland, Daniel S. Nagin, and Ray Paternoster. 2007. "Unpacking the Relationship Between Adolescent Employment and Antisocial Behavior: A Matched Samples Comparison." *Criminology* 45:67-97.
- Auerbach, Randy P., Chad M. McWhinnie, Marc Goldfinger, John R. Z. Abela, Xiongzhao Zhu, and Shuqiao Yao. 2009. "The Cost of Materialism in a Collectivistic Culture: Predicting Risky Behavior Engagement in Chinese Adolescents." *Journal of Clinical Child & Adolescent Psychology* 39:117-127.
- Bachman, Jerald G. and John Schulenberg. 1993. "How Part-Time Work Intensity Relates to Drug Use, Problem Behavior, Time Use, and Satisfaction Among High School Seniors: Are These Consequences or Merely Correlates?" *Developmental Psychology* 29:220-235.
- Baker, Frank. 2001. *The Basics of Item Response Theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland.
- Bao, Wan-Ning, Ain Haas, Xiaojin Chen, and Yijun Pi. 2014. "Repeated Strains, Social Control, Social Learning, and Delinquency: Testing an Integrated Model of General Strain Theory in China." *Youth & Society* 46:402-424.
- Bao, Wan-Ning, Ain Haas, and Yijun Pi. 2007. "Life Strain, Coping, and Delinquency in the People's Republic of China: An Empirical Test of General Strain Theory from a Matching Perspective in Social Support." *International Journal of Offender Therapy and Comparative Criminology* 51:9-24.

- Becker, Gary. 1968. "Crime and Punishment: An Economic Approach." *The Journal of Political Economy* 76:169-217.
- Berg, Mark T. and Beth M. Huebner. 2011. "Reentry and the Ties that Bind: An Examination of Social Ties, Employment, and Recidivism." *Justice Quarterly* 28:382-410.
- Bjerk, David. 2009. "How Much Can We Trust Causal Interpretations of Fixed-Effects Estimators in the Context of Criminality?" Institute for the Study of Labor (IZA), Bonn.
- Bond, Trevor G. and Christine M. Fox. 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Bratton, William and Peter Knobler. 1998. *Turnaround: How America's Top Cop Reversed the Crime Epidemic*. New York: Random House.
- Byrne, Barbara M. 1998. *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*. Mahwah, N.J.: L. Erlbaum Associates.
- Callahan, Richard. 2010. "A Counterfactual Model for How Adolescent Work Patterns Affect Crime." M.A. Thesis, Department of Sociology, University of Washington, Seattle.
- Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *The Review of Economic Studies* 47:225-238.
- Chang, F. M. 張楓明. 2011. "學業層面之參與、抱負、自我效能及緊張因素對初次偏差行為影響之縱貫性研究[A Longitudinal Research on the Effects of Involvement, Commitment, Self-efficacy, and Strain from Academic Aspect on Onset of Delinquency]." *當代教育研究 [Contemporary Educational Research Quarterly]* 19:39-81.
- Chang, Hui-Chun 張惠君. 2002. "家庭系統、學校系統與國中生自我控制及偏差行為之研究——以台南地區為例 [Research into the Family and School Institutions and Middle School Students' Self-control and Deviance: a Case Study of Tainan]." M.A. Thesis, Department of Education, National Cheng Kung University, Tainan, Taiwan.
- Chang, Ming-Yi and Chin-Chun Yi. 2013. "Is it Better to Stay in the Academic Track for Taiwanese Youth? Change of Educational Status and Its Effect on the Psychological Well-Being." in *2013 年台灣青少年成長歷程研究第五次學術研討會 [The 5th Conference on Taiwan Youth Project, 2013]*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.
- Chang, Su. 2010. "The Influence of Social Bonds on Adolescent Delinquency: A Longitudinal Test of Social Control Theory." M.A. Thesis, Department of Social Psychology, Shih Hsin University, Taipei, Taiwan.
- Chang, Ying-Hwa. 2013. "Academic Competition and Cram Schooling." Pp. 131-153 in *The Psychological Well-Being of East Asian Youth, Quality of Life in Asia*, edited by C.-C. Yi. Dordrecht: Springer.
- Chen, Shyh-Wei. 2009. "Investigating Causality Among Unemployment, Income and Crime in Taiwan: Evidence From the Bounds Test Approach." *Journal of Chinese Economic and Business Studies* 7:115-125.
- Chen, Yi-Fu and Chyi-In Wu. 2004. "The Contingent Relationships Among Parenting, School Engagement, and Delinquent Behaviors: The Contextual Effect." in *2004 年台灣青少年成長歷程研究第一次學術研討會 [The First Youth Conference of Taiwan Youth Project, 2004]*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.

- Chen, Yu-Hua and Hsinmu Chen. 2014. "Continuity and Changes in the Timing and Formation of First Marriage Among Postwar Birth Cohorts in Taiwan." *Journal of Family Issues* 35:1584-1604.
- Cheng, Yen-hsin Alice and Nancy S. Landale. 2011. "Adolescent Precursors of Early Union Formation Among Asian American and White Young Adults." *Journal of Family Issues* 32:209-236.
- Cheung, Nicole Wai Ting and Yuet W. Cheung. 2010. "Strain, Self-Control, and Gender Differences in Delinquency Among Chinese Adolescents: Extending General Strain Theory." *Sociological Perspectives* 53:321-345.
- Chiao, Chi, Chin-Chun Yi, and Kate Ksobiech. 2014. "Adolescent Internet Use and its Relationship to Cigarette Smoking and Alcohol Use: A Prospective Cohort Study." *Addictive Behaviors* 39:7-12.
- Chib, Siddhartha and Edward Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician* 49:327-335.
- Chin, Ko-Lin. 2003. *Heijin: Organized Crime, Business, and Politics in Taiwan*, Edited by M. A. Rubinstein. Armonk, New York: M. E. Sharpe, Inc.
- Cho, W. and S.E. Cross. 1995. "Taiwanese Love Styles and Their Association with Self-Esteem and Relationship Quality." *Genetic, Social, and General Psychology Monographs* 121:283-310.
- Chow, Sy-Miin and Guangjian Zhang. 2008. "Continuous-Time Modelling of Irregularly Spaced Panel Data Using a Cubic Spline Model." *Statistica Neerlandica* 62:131-154.
- Chu, C. Y. Cyrus and Ruoh-Rong Yu. 2010. *Understanding Chinese Families: a Comparative Study of Taiwan and Southeast China*. Oxford: Oxford University Press.
- Cole, Stephen R. and Miguel A. Hernán. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology* 168:656-664.
- Committee on the Health and Safety Implications of Child Labor. 1998. "Protecting Youth at Work: Health, Safety, and Development of Working Children and Adolescents in the United States." Washington: National Academy Press.
- Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. 2001. *Introduction to Algorithms*. Cambridge: The MIT Press.
- Cronbach, L. J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16:297-334.
- Csardi, Gabor and Tamas Nepusz. 2006. "The igraph Software Package for Complex Network Research." *InterJournal (Complex Systems)*:1695.
- D'Amico, Ronald. 1984. "Does working in high school impair academic progress?" *Sociology of Education* 57:157-164.
- Deng, Shiyong and Mark W. Roosa. 2007. "Family Influences on Adolescent Delinquent Behaviors: Applying the Social Development Model to a Chinese Sample." *American Journal of Community Psychology* 40:333-344.
- Denq, Furjen, Lou-Jou Lee, and Michael S. Vaughn. 1996. "Drug Crime in Taiwan: A Time-Series Analysis from 1965 to 1994." *Deviant Behavior* 17:347-368.
- Ebenstein, Avraham. 2010. "The "Missing Girls" of China and the Unintended Consequences of the One Child Policy." *Journal of Human Resources* 45:87-115.
- Elliott, Delbert S. and David Huizinga. 1983. "Social Class and Delinquent Behavior in a National Youth Panel." *Criminology* 21:149-177.

- Elliott, Delbert S., David Huizinga, and Suzanne S. Ageton. 1985. *Explaining Delinquency and Drug Use*. Beverly Hills, CA: Sage Publications.
- Fan, C. Cindy and Youqin Huang. 1998. "Waves of Rural Brides: Female Marriage Migration in China." *Annals of the Association of American Geographers* 88:227-251.
- Fewell, Zoe, Miguel A. Hernán, Frederick Wolfe, Kate Tilling, Hyon Choi, and Jonathan A. C. Sterne. 2004. "Controlling for Time-Dependent Confounding Using Marginal Structural Models." *The Stata Journal* 4:402-420.
- Fischhoff, Baruch, Andrew M. Parker, Wändi de Bruin, Julie Downs, Claire Palmgren, Robyn Dawes, and Charles F. Manski. 2000. "Teen Expectations for Significant Life Events." *The Public Opinion Quarterly* 64:189-205.
- Foster, Chris. 2014. "A Laughter Turns Serious as UCLA Needs Two Overtimes to Beat Colorado." in *Los Angeles Times*. Los Angeles, CA: Los Angeles Times Company.
- Fox, Jean-Paul. 2010. *Bayesian Item Response Modeling: Theory and Applications*, Edited by S. E. Feinberg and W. van der Linden. New York, NY: Springer.
- Fox, John. 2002. "Nonparametric Regression: Appendix to An R and S-PLUS Companion to Applied Regression." CRAN-R Project online resource, retrieved April 2009 from <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-nonparametric-regression.pdf>.
- Fricke, T., J.S. Chang, and L.S. Tang. 1994. "Historical and Ethnographic Perspectives on the Chinese Family." Pp. 22-48 in *Social Change and the Family in Taiwan, Population and Development*, edited by A. Thornton and H.-S. Lin. Chicago: University of Chicago Press.
- Fukushima, Miyuki, Susan F. Sharp, and Emiko Kobayashi. 2009. "Bond to Society, Collectivism, and Conformity: A Comparative Study of Japanese and American College Students." *Deviant Behavior* 30:434-466.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*, Edited by C. Chatfield, M. Tanner, and J. Zidek. Boca Raton: Chapman & Hall / CRC.
- Giordano, Peggy C., Ryan D. Schroeder, and Stephen A. Cernkovich. 2007. "Emotions and Crime over the Life Course: A Neo-Meadian Perspective on Criminal Continuity and Change." *American Journal of Sociology* 112:1603-1661.
- Glueck, Sheldon and Eleanor Glueck. 1950. *Unraveling Juvenile Delinquency*. New York: The Commonwealth Fund.
- Goo, Ei-Ten 辜易天. 2007. "國中學生依附關係、活動參與、教育抱負與偏差行為：社會控制理論在台灣之適用性[The Attachment, Involvement of Activities, Educational Ambitions and Deviant Behaviors of Junior High School Students: The Adaptation of Social Control Theory in Taiwan]." M.A. Thesis, Department of Education, National Taidong University, Taidong, Taiwan.
- Goodman, Leo A. and William H. Kruskal. 1954. "Measures of Association for Cross Classifications." *Journal of the American Statistical Association* 49:732-764.
- Gottfredson, Michael and Travis Hirschi. 1990. *A General Theory of Crime*. Stanford: Stanford University Press.
- Gould, Leroy C. 1969. "Who Defines Delinquency: A Comparison of Self-Reported and Officially-Reported Indices of Delinquency for Three Racial Groups." *Social Problems* 16:325-336.

- Greenberger, Ellen and Laurence Steinberg. 1986. *When Teenagers Work: The Psychological and Social Costs of Adolescent Employment*. New York: Basic Books.
- Greenland, Sander. 2003. "Quantifying Biases in Causal Models: Classical Confounding vs Collider-Stratification Bias." *Epidemiology* 14:300-306.
- Haebara, T. 1980. "Equating Logistic Ability Scales by a Weighted Least Squares Method." *Japanese Psychological Research* 22:144-149.
- Hambleton, Ronald K., H. Swaminathan, and H. Jane Rogers. 1991. *Fundamentals of Item Response Theory*, Edited by R. M. Jaeger. Newbury Park: Sage Publications.
- Hamilton, Gary G. and Wang Zheng. 1992. "From the Soil, the Foundations of Chinese Society: a Translation of Fei Xiaotong's Xiangtu Zhongguo, with an Introduction and Epilogue." Berkeley: University of California Press.
- Harris, Deborah J. 2007. "Practical Issues in Vertical Scaling." Pp. 233-251 in *Linking and Aligning Scores and Scales, Statistics for Social and Behavioral Sciences*, edited by N. J. Dorans, M. Pommerich, and P. W. Holland. New York, NY: Springer.
- Hausman, J.A. 1978. "Specification Tests in Econometrics." *Econometrica* 46:1251-1271.
- Hausman, Jerry A. and Daniel McFadden. 1984. "Specification Tests for the Multinomial Logit Model." *Econometrica* 52:1219-40.
- Hawkins, J. David and Joseph Weis. 1985. "The Social Development Model: An Integrated Approach to Delinquency Prevention." *Journal of Primary Prevention* 6:73-97.
- Heiss, Florian. 2002. "Structural Choice Analysis with Nested Logit Models." *The Stata Journal* 2:227-252.
- Hendrick, C. and S. Hendrick. 1986. "A Theory and Method of Love." *Journal of Personality and Social Psychology* 50:392-402.
- Hendrick, C. and S. Hendrick. 1990. "A Relationship-Specific Version of the Love Attitudes Scale." *Journal of Social Behavior and Personality* 5:239-254.
- Hernán, Miguel A., Babette Brumback, and James M. Robins. 2000. "Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men." *Epidemiology* 11:561-570.
- Hindelang, Michael, Travis Hirschi, and Joseph Weis. 1981. *Measuring Delinquency*. Beverly Hills, CA: Sage.
- Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley: University of California Press.
- Hsu, Francis L. K. 1948. *Under the Ancestors' Shadow: Chinese Culture and Personality*. New York: Columbia University Press.
- Huang, Fung-Mey and Yu-Ning Chien. 2013. "Working, Schooling, and Psychological Well-Being: Evidence from Longitudinal Data for Taiwanese Youth." Pp. 185-207 in *The Psychological Well-Being of East Asian Youth, Quality of Life in Asia*, edited by C.-C. Yi. Dordrecht: Springer.
- Huang, Fung-Mey and Pin-Yu Lin. 2009. "同儕效果對於青少年抽菸、喝酒或吸食藥物的影響[The Influence of Peer Effects on Adolescent Smoking, Drinking and Drug Use]." in *2009年台灣青少年成長歷程研究第三次學術研討會[The 3rd Youth Conference of Taiwan Youth Project]*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.
- Huang, Fung-Mey, Chin-Chun Yi, and Jui-Che Yang. 2008a. "Educational Tracking, High School Employment and College Entry in Taiwan." in *The Second International Conference on Youth Studies, 2008 (台灣青少年成長歷程研究第二次學術研討會)*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.

- Huang, Fung-Mey 黃芳玫 and Chyi-Yin Wu 吳齊殷. 2010. "Taiwan Guozhong Xuesheng Geren Texing, Jiating Beijing yu Xueye Chengji – Zhuicong Diaocha Ziliao zhi Yanjiu 台灣國中學生個人特質、家庭背景與學業成績—追從調查資料之研究 [Individual Characteristics, Family Backgrounds, and Educational Outcomes of Taiwanese Junior High School Students – Panel Data Analysis]." *Jingji Lunwen Congkan 經濟論文叢刊 (Taiwan Economic Review)* 38:65-97.
- Huang, Fung-Mey 黃芳玫, Chin-Chun Yi 伊慶春, and Jui-Che Yang 楊睿哲. 2008b. "Educational Tracking, High School Employment and College Entry in Taiwan." in *The International Conference on Youth Studies, 2008*. Taipei, Taiwan (R.O.C.): Institute of Sociology, Academia Sinica.
- Huang, Hsin-Chang 黃信璋. 2006a. "苗栗縣國中學生偏差行為與師生衝突間相關性之研究 [Analysis of the relationship between junior high school student behavior and conflict with teachers in Miaoli County]." National Taiwan Normal University, Taipei, Taiwan.
- Huang, Jr-Tsung. 2007. "Labor Force Participation and Juvenile Delinquency in Taiwan: A Time Series Analysis." *Journal of Family and Economic Issues* 28:137-150.
- Huang, Ya-Rong 黃雅容. 2006b. "分析大學生打工對時間分配的影響 [Explaining the Effect of Part-Time Work on Time Allocation for College Students]." *學生事務 [Student Affairs]* 45:161-170.
- Huber, P.J. 1967. "The Behavior of Maximum Likelihood Estimation Under Nonstandard Conditions." Pp. 221-233 in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, edited by L. M. LeCam and J. Neyman: University of California Press.
- Huizinga, David and Delbert S. Elliott. 1986. "Reassessing the Reliability and Validity of Self-Report Delinquency Measures." *Journal of Quantitative Criminology* 2:293-327.
- Imai, Kosuke and Marc Ratkovic. 2014. "Covariate Balancing Propensity Score." *Journal of the Royal Statistical Society Series B - Statistical Methodology* 76:243-263.
- Jackson, Christopher. 2010. "msm: Multi-state Markov and hidden Markov models in continuous time." R package version 0.9.6. <http://CRAN.R-project.org/package=msm>.
- Jessor, R. and S.L. Jessor. 1977. *Problem Behavior and Psychosocial Development*. New York: Academic Press.
- Jessor, R. and S.L. Jessor. 1978. "Theory Testing in Longitudinal Reserch on Marijuana Use." Pp. 41-71 in *Longitudinal Research on Drug Use: Empirical Findings and Methodological Issues*, edited by D. B. Kandel. Washington, DC: Hemisphere.
- Jessor, Richard, Mark S. Turbin, Frances M. Costa, Qi Dong, Hongchuan Zhang, and Changhai Wang. 2003. "Adolescent Problem Behavior in China and the United States: A Cross-National Study of Psychosocial Protective Factors." *Journal of Research on Adolescence* 13:329-360.
- Jolliffe, Darrick, David P. Farrington, J. David Hawkins, Richard F. Catalano, Karl G. Hill, and Rick Kosterman. 2003. "Predictive, Concurrent, Prospective and Retrospective Validity of Self-reported Delinquency." *Criminal Behaviour and Mental Health* 13:179-197.
- Jöreskog, Karl and Dag Sörbom. 2002. *PRELIS 2: User's Reference Guide*. Lincolnwood, IL: Scientific Software International, Inc.
- Juang, Linda P. and Huong H. Nguyen. 2009. "Misconduct Among Chinese American Adolescents: The Role of Acculturation, Family Obligation, and Autonomy Expectations." *Journal of Cross-Cultural Psychology* 40:649-666.

- Kang, Joseph D. Y. and Joseph L. Schafer. 2007. "Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data." *Statistical Science* 22:523-539.
- Kim, Seonghoon and Michael J. Kolen. 2003. "POLYST: A Computer Program for Polytomous IRT Scale Transformation." Iowa City, IA: University of Iowa.
- Kim, Seonghoon and Michael J. Kolen. 2005. "Methods for Obtaining a Common Scale Under Unidimensional IRT Models: A Technical Review and Further Extensions (Occasional Paper)." Iowa Testing Programs, Iowa City, IA.
- Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi. 1983. "Optimization by Simulated Annealing." *Science New Series*, Vol. 220:671-680.
- Kolen, Michael J. and Robert L. Brennan. 2004. *Test Equating, Scaling, and Linking: Methods and Practices*, Edited by S. E. Fienberg and W. van der Linden. New York: Springer-Verlag.
- Laub, John H. and Robert J. Sampson. 2003. *Shared Beginnings, Divergent Lives: Delinquent Boys to Age 70*. Cambridge, MA: Harvard University Press.
- Lee, Chiao-Yen 李昭鑒. 2011. "影響台灣青少年偏差行為之貫時研究-以 TEPS 資料分析為例 [Panel study of deviance among Taiwanese youth: Evidence from the Taiwan Education Panel Survey]." M.A. Thesis, Dept. of Education, National Chengchi University, Taipei, Taiwan.
- Li, Haibin, Andrew J. Martin, Derrick Armstrong, and Richard Walker. 2011. "Risk, Protection, and Resilience in Chinese Adolescents: A Psycho-Social Study." *Asian Journal of Social Psychology* 14:269-282.
- Li, Jie-Zhi 李介至 and Shao-Yi Qiu 邱紹一. 2004. "技職學生打工動機與打工活動之相關研究—以苗栗地區為例 [Analysis of the Relationship Between Motivation for Part-Time Working and Part-Time Work for Vocational Students -- Case Study of Miaoli District]." *仁德學報 [Rende Xuebao]* 3:25-39.
- Li, K. H., T. E. Raghunathan, and Donald B. Rubin. 1991. "Large-Sample Significance Levels from Multiply Imputed Data Using Moment-Based Statistics and an F Reference Distribution." *Journal of the American Statistical Association* 86:1065-1073.
- Li, Zhi Hong, Jennifer Connolly, Depeng Jiang, Debra Pepler, and Wendy Craig. 2010. "Adolescent Romantic Relationships in China and Canada: A Cross-National Comparison." *International Journal of Behavioral Development* 34:113-120.
- Liao, Nian-Miao 廖年淼 and Meng-Shan Liu 劉孟珊. 2004. "高中職教育資源投入之比較分析 [Comparative Analysis of Resource Allocation for General and Vocational High School Education]." *教育政策論壇 [Education Policy Forum]* 7:41-58.
- Lin, Kuei-Hsiu 林桂綉 and Zheng-Wang 李政旺 Li. 2013. "青少年成長計劃資料流失軌跡分析 [Trajectory Analysis of Survey Attrition in the Youth Development Project]." in *2013 年台灣青少年成長歷程研究第五次學術研討會 [The 5th Conference on Taiwan Youth Project, 2013]*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.
- Lin, Wen-Hsu, Richard Dembo, Christine S. Sellers, John Cochran, and Thomas Mieczkowski. 2014. "Strain, Negative Emotions, and Juvenile Delinquency: The United States Versus Taiwan." *International Journal of Offender Therapy and Comparative Criminology* 58:412-434.

- Lin, Wen-Hsu and Chin-Chun Yi. forthcoming. "Educational Tracking and Juvenile Deviance in Taiwan: Direct Effect, Indirect Effect, or Both." *International Journal of Offender Therapy and Comparative Criminology*.
- Liu, Jianhong. 2008. "Data Sources in Chinese Crime and Criminal Justice Research." *Crime, Law, & Social Change* 50:131-147.
- Liu, Jun-Cheng 劉峻誠. 2003. "青少年家庭依附、就學經驗與偏差行為相關性之研究—以南投縣為例[The Relevant Research of Teenager's Family Attachment, School Experience, and Deviance -- Nantou County as an Example]." M.A. Thesis, Institute of Educational Sociology, Nanhua University, Chiayi, Taiwan.
- Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: Stata Press.
- Lord, F.M. 1980. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.
- Lord, F.M. 1983. "Maximum Likelihood Estimation of Item Response Parameters When Some Responses are Omitted." *Psychometrika* 48:477-482.
- Lord, F.M. and M.R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Lord, Frederic M. 1952. "A Theory of Test Scores." Psychometric Monograph No. 7. Psychometric Society.
- MacDonald, Geoff, Tara C. Marshall, Judith Gere, Atsushi Shimotomai, and July Lies. 2012. "Valuing Romantic Relationships: The Role of Family Approval Across Cultures." *Cross-Cultural Research* 46:366-393.
- Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.
- Mair, Patrick and Reinhold Hatzinger. 2007. "Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R." WU Vienna University of Economics and Business, Vienna.
- Manpower Demonstration Research Corporation. 1980. "Summary and Findings of the National Supported Work Demonstration." Cambridge, Massachusetts: Ballinger.
- Manski, Charles F. 1990. "The Use of Intentions Data to Model Behavior: A Best-Case Analysis." *Journal of the American Statistical Association* 85:934-940.
- Maple, Jack. 2000. *The Crime Fighter: How You Can Make Your Community Crime-Free*. New York: Doubleday.
- Marshall, Tara C. 2008. "Cultural Differences in Intimacy: The Influence of Gender Role Ideology and Individualism-Collectivism." *Journal of Social and Personal Relationships* 25:143-168.
- Masters, G.N. 1982. "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47:149-174.
- Matsueda, Ross L. 1989. "The Dynamics of Moral Beliefs and Minor Deviance." *Social Forces* 68:428-457.
- Matza, David. 1964. *Delinquency and Drift*. New York: Wiley.
- McCullagh, P. and J.A. Nelder. 1989. *Generalized Linear Models*, Edited by D. R. Cox, D. V. Hinkley, D. Rubin, and B. W. Silverman. London: Chapman and Hall.
- McGee, L. and Michael D. Newcomb. 1992. "General Deviance Syndrome: Expanded Hierarchical Evaluations at Four Ages from Early Adolescence to Adulthood." *Journal of Consulting and Clinical Psychology* 60:766-776.

- Merton, Robert K. 1938. "Social Structure and Anomie." *American Sociological Review* 3:672-682.
- Merton, Robert K. 1957. *Social Theory and Social Structure: Revised and Enlarged Edition*. Glencoe, Illinois: The Free Press.
- Messner, Steven F. and Richard Rosenfeld. 2007. *Crime and the American Dream*. Belmont, CA: Thomson / Wadsworth.
- Mislevy, Robert J. and Pao-Kuei Wu. 1996. "Missing Responses and IRT Ability Estimation: Omits, Choice, Time Limits, and Adaptive Testing." Educational Testing Service, Princeton, NJ.
- Moore, Robert L. 1998. "Love and Limerence with Chinese Characteristics: Student Romance in the PRC." Pp. 251-283 in *Romantic Love and Sexual Behavior: Perspectives from the Social Sciences*, edited by V. C. de Munck. Westport: Praeger.
- Mortimer, Jeylan T. 2003. *Working and Growing Up In America*, Edited by J. Brooks-Gunn. Cambridge: Harvard University Press.
- Mortimer, Jeylan T., Michael D. Finch, Seongryeol Ryu, Michael T. Shanahan, and Kathleen T. Call. 1996. "The Effect of Work Intensity on Adolescent Mental Health, Achievement, and Behavioral Adjustment: New Evidence From a Prospective Study." *Child Development* 67:1243-1261.
- Mortimer, Jeylan T. and Monica Kirkpatrick Johnson. 1998. "New perspectives on adolescent work and the transition to adulthood." Pp. 425-496 in *New perspectives on adolescent risk behavior*, edited by R. Jessor. Cambridge: Cambridge University Press.
- Muraki, Eiji. 1992. "A Generalized Partial Credit Model: Application of an EM Algorithm." *Applied Psychological Measurement* 16:159-76.
- Nagin, Daniel S. 2005. *Group-Based Modeling of Development*. Cambridge, MA: Harvard University Press.
- Newcomb, Michael D. 1995. "Adolescent Drug Abuse." in *Adolescent Drug Abuse: Clinical Assessment and Therapeutic Interventions. NIDA Research Monograph 156*, edited by E. Rahdert and D. Czechowicz. Rockville, MD: National Institute on Drug Abuse.
- Newcomb, Michael D. and Peter M. Bentler. 1988. *Consequences of Adolescent Drug Use: Impact on the Lives of Young Adults*. Newbury Park: Sage Publications.
- Osgood, D. Wayne, Barbara J. McMorris, and Maria T. Potenza. 2002. "Analyzing Multiple-Item Measures of Crime and Deviance I: Item Response Theory Scaling." *Journal of Quantitative Criminology* 18:267-296.
- Osgood, D. Wayne and Christopher J. Schreck. 2007. "A New Method for Studying the Extent, Stability, and Predictors of Individual Specialization in Violence." *Criminology* 45:273-312.
- Oud, Johan H.L. and Robert A.R.G. Jansen. 2000. "Continuous Time State Space Modeling of Panel Data by Means of SEM." *Psychometrika* 65:199-215.
- Oud, Johan H.L. and Hermann Singer. 2008. "Continuous Time Modeling of Panel Data: SEM Versus Filter Techniques." *Statistica Neerlandica* 62:4-28.
- Pan, En-Ling and Chin-Chun Yi. 2011. "Constructing Educational Resilience: The Developmental Trajectory of Vulnerable Taiwanese." *Journal of Comparative Family Studies* 42:369-383.
- Pan, Wei. 2001. "Akaike's Information Criterion in Generalized Estimating Equations." *Biometrics* 57:120-125.

- Patz, Richard J. 1996. "Markov Chain Monte Carlo Methods for Item Response Theory Models with Applications for the National Assessment of Educational Progress." Statistics, Carnegie Mellon University, Pittsburgh.
- Patz, Richard J. and Brian W. Junker. 1999a. "Applications and Extensions of MCMC in IRT: Multiple Item Types, Missing Data, and Rated Responses." *Journal of Educational and Behavioral Statistics* 24:342-366.
- Patz, Richard J. and Brian W. Junker. 1999b. "A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models." *Journal of Educational and Behavioral Statistics* 24:146-178.
- People's Daily. 1989. "Most Rural People's Marriage Boundaries are Less Than 25 km. [Duoshu nongmin tonghunchuan bu chaoguo 25 gongli]." P. 1 in *Renmin Ribao [People's Daily]*. Beijing.
- Petersen, Maya L., Kristen E. Porter, Susan Gruber, Yue Wang, and Mark J. van der Laan. 2012. "Diagnosing and Responding to Violations in the Positivity Assumption." *Statistical Methods in Medical Research* 21:31-54.
- Petersen, Nancy S. 2007. "Equating: Best Practices and Challenges to Best Practices." Pp. 59-72 in *Linking and Aligning Scores and Scales, Statistics for Social and Behavioral Sciences*, edited by N. J. Dorans, M. Pommerich, and P. W. Holland. New York, NY: Springer.
- Pfarr, Klaus. 2014. "Femlogit – Implementation of the Multinomial Logit Model with Fixed Effects." *The Stata Journal* 14:847-862.
- Piquero, Alex R., Carol A. Schubert, and Robert Brame. 2014. "Comparing Official and Self-report Records of Offending across Gender and Race/Ethnicity in a Longitudinal Study of Serious Youthful Offenders." *Journal of Research in Crime and Delinquency* 51:526-556.
- Pyrooz, David C. and Scott H. Decker. 2013. "Delinquent Behavior, Violence, and Gang Involvement in China." *Journal of Quantitative Criminology* 29:251-272.
- Qiu, Yi-Hsiang 邱羿翔. 2013. "大學生自我分化與愛情觀之相關研究[The Relationship Between Differentiation of Self and Love Attitude for College Students]." M.A. Thesis, Department of Psychological Counseling and Clinical Psychology, National Dong Hwa University, Hualien, Taiwan.
- Rabe-Hesketh, S. and A. Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata, Third Edition*, vol. 2. College Station, TX: Stata Press.
- Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111-163.
- Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: The Danish Institute of Educational Research.
- Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, Edited by J. de Leeuw. Thousand Oaks: Sage Publications.
- Robins, J. 2000. "Marginal Structural Models Versus Structural Nested Models as Tools For Causal Inference." Pp. 95-133 in *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, edited by M. E. Halloran and D. Berry. New York: Springer-Verlag.
- Robins, James M, Miguel Angel Hernán, and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11:550-560.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, New Jersey: John Wiley & Sons.

- Samejima, Fumiko. 1969. "Estimation of Latent Ability Using a Response Pattern of Graded Scores." in *Psychometric Monograph No. 17*. Richmond, VA: Psychometric Society.
- Sampson, Robert J. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago: University of Chicago Press.
- Sampson, Robert J. and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.
- Sampson, Robert J., John L. Laub, and Christopher Wimer. 2006. "Does Marriage Reduce Crime? A Counterfactual Approach to Within-Individual Causal Effects." *Criminology* 44:465-510.
- Sampson, Robert J. and Stephen W. Raudenbush. 1999. "Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods." *American Journal of Sociology* 105:603-651.
- Schmidheiny, Kurt. 2014. "Panel Data: Fixed and Random Effects." Retrieved Dec. 5th, 2014 from <http://kurt.schmidheiny.name/teaching/panel.pdf>. University of Basel, Basel, Switzerland.
- Schulenberg, John and Jerald G. Bachman. 1993. "Long Hours on the Job? Not So Bad for Some Types of Jobs: The Quality of Work and Substance Use, Affect, and Stress." in *Biennial Meeting of the Society for Research on Child Development*. New Orleans.
- Scientific Software International. 2003. *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*, Edited by M. du Toit. Lincolnwood, Ill: Scientific Software International.
- Seaman, Shaun R. and Ian R. White. 2011. "Review of Inverse Probability Weighting for Dealing with Missing Data." *Statistical Methods in Medical Research* 22:278-295.
- Shaw, C. R. and H.D. McKay. 1942. *Juvenile Delinquency and Urban Areas*. Chicago: University of Chicago Press.
- Sheu, Chuen-Jim. 1988. "Juvenile Delinquency in the Republic of China: A Chinese Empirical Study of Social Control Theory." *International Journal of Comparative and Applied Criminal Justice* 12:59-71.
- Sheu, Chuen-Jim 許春金 and Wei-Teh 孟維德 Mon. 1997. "家庭、學校、自我控制與偏差行為[Family, School, Self-Control, and Delinquency]." *中央警察大學學報[Journal of Central Police University]* 30:225-256.
- Short, James F. and F. Ivan Nye. 1957. "Reported Behavior as a Criterion of Deviant Behavior." *Social Problems* 5:207-213.
- Simons, Ronald L., Kuei-Hsiu Lin, Leslie C. Gordon, Gene H. Brody, Velma Murry, and Rand D. Conger. 2002. "Community Differences in the Association Between Parenting Practices and Child Conduct Problems." *Journal of Marriage and Family* 64:331-345.
- Snijders, Tom A. B., Gerhard G. van de Bunt, and Christian E. G. Steglich. 2010. "Introduction to Stochastic Actor-Based Models for Network Dynamics." *Social Networks* 32:44-60.
- South China Morning Post Staff Reporter. 2008. "Mainland Population Policy Needs Rethink." in *South China Morning Post*. Hong Kong.
- Stack, Carol B. and Linda M. Burton. 1994. "Kinscripts: Reflections on Family, Generation, and Culture." Pp. 33-44 in *Mothering: Ideology, Experience, and Agency*, edited by E. N. Glenn, G. Chang, and L. R. Forcey. New York, NY: Routledge.
- Staff, Jeremy, Jeylan T. Mortimer, and Christopher Uggen. 2004. "Work and Leisure in Adolescence." Pp. 429-450 in *Handbook of Adolescent Psychology*, edited by R. M. Lerner and L. Steinberg. Hoboken, New Jersey: John Wiley & Sons.
- StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.

- Steinberg, Laurence and Sanford M. Dornbusch. 1991. "Negative Correlates of Part-Time Employment During Adolescence: Replication and Elaboration." *Developmental Psychology* 27:304-313.
- Steinberg, Laurence, Suzanne Fegley, and Sanford M. Dornbusch. 1993. "Negative Impact of Part-Time Work on Adolescent Adjustment: Evidence From a Longitudinal Study." *Developmental Psychology* 29:171-180.
- Steinberg, Laurence, Ellen Greenberger, L. Garduque, and S. McAuliffe. 1982. "High school students in the labor force: Some costs and benefits to schooling and learning." *Educational Evaluation and Policy Analysis* 4:363-372.
- Stocking, M. and F.M. Lord. 1983. "Developing a Common Metric in Item Response Theory." *Applied Psychological Measurement* 7:207-210.
- Sutherland, Edwin H. 1947. *Principles of Criminology*. Chicago: J.P. Lippencott Company.
- Swygert, Kimberly A., Lori D. McLeod, and David Thissen. 2001. "Factor Analysis for Items or Testlets Scored in More Than Two Categories." Pp. 217-250 in *Test Scoring*, edited by D. Thissen and H. Wainer. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Thissen, David, Lauren Nelson, Kathleen Rosa, and Lori D. McLeod. 2001. "Item Response Theory for Items Scored in More Than Two Categories." in *Test Scoring*, edited by D. Thissen and H. Wainer. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Thornton, A., T. Fricke, L.S. Yang, and J.S. Chang. 1994. "Theoretical Mechanisms of Family Change." Pp. 88-115 in *Social Change and the Family in Taiwan, Population and Development*, edited by A. Thornton and H.-S. Lin. Chicago and London: The University of Chicago Press.
- Tsai, Shu-Ling. 2004. "高等教育的擴展對教育機會分配的影響[Effects of higher education expansion on inequality of educational opportunity]." *台灣社會學(Taiwanese Sociology)* 7:47-88.
- Tzeng, Shuping and Chin-Chun Yi. 2009. "The Nonlinear Effects of Self-Esteem on Adolescent Delinquency." in *2009 年台灣青少年成長歷程研究第三次學術研討會 [The 3rd Youth Conference of Taiwan Youth Project, 2009]*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.
- Uggen, Christopher. 1999. "Ex-Offenders and the Conformist Alternative: A Job Quality Model of Work and Crime." *Social Problems* 46:127-151.
- Uggen, Christopher. 2000. "Work as a Turning Point in the Life Course of Criminals: A Duration Model of Age, Employment, and Recidivism." *American Sociological Review* 65:529-546.
- van Buuren, Stef and Karin Groothuis-Oudshoorn. 2011. "mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45.
- Vanfossen, Beth E., James D. Jones, and Joan Z. Spade. 1987. "Curriculum Tracking and Status Maintenance." *Sociology of Education* 60:104-122.
- Wang, Fei and Dengfeng Wang. 2012. "Chinese College Students' General and Relationship-Specific Representations of Adult Attachment: Variable-Centred and Person-Centred Analysis." *Asian Journal of Social Psychology* 15:314-320.
- Wang, Qing-Fu 王慶福 and Yu-Ming Wang 王郁茗. 2003. "性別、性別角色取向、愛情觀與愛情關係的分析研究[Analysis of Gender, Sexual Orientation, Love Attitudes, and Romantic Relationships]" *中山醫學雜誌[The Chung Shan Medical Journal]* 14:71-82.

- Wang, Shu-Neu and Gary F. Jensen. 2003. "Explaining Delinquency in Taiwan: A Test of Social Learning Theory." Pp. 65-83 in *Social Learning Theory and the Explanation of Crime, Advances in Criminological Theory: A Guide for the New Century*, edited by R. L. Akers and G. F. Jensen. New Brunswick: Transaction Publishers.
- Wang, Yue, David Bangsberg, Maya L. Petersen, and Mark J. van der Laan. 2006. "Diagnosing Bias in the Inverse Probability of Treatment Weighted Estimator Resulting from Violation of Experimental Treatment Assignment." in *U.C. Berkeley Division of Biostatistics Working Paper Series*. Berkeley.
- Webb, Vincent J., Ling Ren, Jihong "Solomon" Zhao, Ni "Phil" He, and Ineke Haen Marshall. 2011. "A Comparative Study of Youth Gangs in China and the United States: Definition, Offending, and Victimization." *International Criminal Justice Review* 21:225-242.
- White, H. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50:1-25.
- White, Ian R., Rhian Daniel, and Patrick Royston. 2010. "Avoiding Bias Due to Perfect Prediction in Multiple Imputation of Incomplete Categorical Variables." *Computational Statistics & Data Analysis* 54:2267-2275.
- White, Ian R., Patrick Royston, and Angela M. Wood. 2010. "Multiple Imputation Using Chained Equations: Issues for Guidance and Practice." *Statistics in Medicine* 30:377-399.
- Wilson, M. 2005. *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Erlbaum.
- Wright, B.D. and G.N. Masters. 1982. *Rating Scale Analysis*. Chicago: MESA Press.
- Wu, Chyi-In and Man-Kit Lei. 2013. "Deviant Peers and Delinquency: The Influence of Classroom Context." Pp. 91-107 in *The Psychological Well-being of East Asian Youth, Quality of Life in Asia*, edited by C.-C. Yi. Dordrecht: Springer.
- Wu, Chyi-Wen and Kuang-Hui Yeh. 2013. "父母關心支持與工作自主經驗：雙元自主性促進因素的範疇優勢性檢証 [Work autonomy experience and parental care and support: empirical test of domain superiority for the factors that promote two dimensions of autonomy]." in *2013 年台灣青少年成長歷程研究第五次學術研討會 [The 5th Conference on Taiwan Youth Project]*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.
- Wu, Chyi-Yin and Wen-Chieh Lee. 2003. "青少年憂鬱症狀與偏差行為並發之關係機制 [Mechanisms driving the comorbidity relationship of youth depression and deviant behavior]." *台灣社會學 [Taiwanese Sociology]* 6:119-175.
- Wu, Chyi-Yin and M.-K. Lei. 2003. "Co-Occurring Depressive Symptoms and Delinquency in Adolescents." *台灣社會學 [Taiwanese Sociology]* 6:119-175.
- Wu, Chyi-Yin 吳齊殷. 2011. "青少年偏差與同儕網絡間的選擇與影響機制 [Mechanisms for Selection and Influence Between Adolescent Deviance and Peer Networks]." in *2011 年台灣青少年成長歷程研究第四次學術研討會 [The 4th Conference on Taiwan Youth Project, 2011]*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.
- Wu, Chyi-Yin 吳齊殷, Pei-Wen Chiang 江佩玟, and Zheng-Kwan Hwang 黃正寬. 2013. "重新檢視樣本流失問題：一個社會整合的觀點 [Re-investigating the Problem of Survey Attrition: A Social Integration Perspective]." in *2013 年台灣青少年成長歷程研究第五次學術研討會 [The 5th Conference on Taiwan Youth Project, 2013]*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.

- Wu, Ming-Yeh and Yuh-Huey Jou. 2009. "台灣青少年的道德信念：社會依附的影響 [Moral Beliefs of Adolescents in Taiwan: Impacts of Social Attachment]." *台灣社會學 (Taiwanese Sociology)* 17:61-100.
- Xie, Yu-Sheng 謝雨生 and Meng-Xian Zhou 周孟嫻. 2013. "台灣青少年及青年補習有效嗎？反事實因果推論之檢驗 [Is Attending Cram School Effective for Taiwanese Youth? A Test Using Counterfactual Causal Inference]." in *2013 年台灣青少年成長歷程研究第五次學術研討會 [The 5th Conference on Taiwan Youth Project, 2013]*. Taipei, Taiwan: Institute of Sociology, Academia Sinica.
- Yang, Shu-Lung and John P. Hoffmann. 1998. "A Multilevel Assessment of Social Disorganization Theory in Taipei, Taiwan." *Journal of Contemporary Criminal Justice* 14:222-247.
- Yeh, Kuang-Hui, Yih-Lan Liu, Hsiao-Shan Huang, and Yung-Jui Yang. 2007. "Individuating and Relating Autonomy in Culturally Chinese Adolescents." Pp. 123-146 in *Casting the Individual in Societal and Cultural Contexts*, edited by J. Liu, C. Ward, A. B. I. Bernardo, M. Karasawa, and R. Fischer. Korea: Kyoyook-Kwahak-Sa Publishing.
- Yeh, Kuang-Hui and Yung-Jui Yang. 2006. "Construct Validation of Individuating and Relating Autonomy Orientations in Culturally Chinese Adolescents." *Asian Journal of Social Psychology* 9:148-160.
- Yi, Chin-Chun. 2012. "Introduction to the Psychological Well-Being of East Asian Youth: The Transition from Early Adolescence to Young Adulthood." Pp. 1-28 in *The Psychological Well-Being of East Asian Youth*, edited by C.-C. Yi. Dordrecht: Springer.
- Yi, Chin-Chun, Gang-Hua Fan, and Ming-Yi Chang. 2013. "The Developmental Outcome of Taiwanese Youth: Effects of Educational Tracking During Adolescence." Pp. 157-184 in *The Psychological Well-being of East Asian Youth, Quality of Life in Asia*, edited by C.-C. Yi. Dordrecht: Springer.
- Yi, Chin-Chun, Chyi-Yin Wu, Ying-Hwa Chang, and Ming-Yi Chang. 2009. "The Psychological Well-Being of Taiwanese Youth: School versus Family Context from Early to Late Adolescence." *International Sociology* 24:397-429.
- Yogan, Lissa J. 2000. "School Tracking and Student Violence." *Annals of the American Academy of Political and Social Science* 567:108-122.
- Yu, Huang-Chieh 尤煌傑. 2003. "哲學大家談：你認為一切倫理德行中最重要的一項是什麼？ [Psychology Forum: What Do You Think Are the Most Important Moral Actions?]." *哲學與文化 [Psychology and Culture]* 30:195-196.
- Zhang, Lening and Steven F. Messner. 1995. "Family Deviance and Delinquency in China." *Criminology* 33:359-387.
- Zhang, Lening and Steven F. Messner. 1996. "School Attachment and Official Delinquency Status in the People's Republic of China." *Sociological Forum* 11:285-303.
- Zhang, Lening and Steven F. Messner. 1999. "Bonds to the Work Unit and Official Offense Status in Urban China." *International Journal of Offender Therapy and Comparative Criminology* 43:375-390.
- Zhang, Lening, Steven F. Messner, and Jianhong Liu. 2007. "A Multilevel Analysis of the Risk of Household Burglary in the City of Tianjin, China." *British Journal of Criminology* 47:918-937.

- Zhang, Shuangyue and Susan L. Kline. 2009. "Can I Make My Own Decision? A Cross-Cultural Study of Perceived Social Network Influence in Mate Selection." *Journal of Cross-Cultural Psychology* 40:3-23.
- Zimring, Franklin E. 2007. *The Great American Crime Decline*. Oxford: Oxford University Press.
- 國勢普查處[National Status Investigatory Division]. 2010. "98年青少年狀況調查報告[2009 Youth Status Survey Report]." Taipei: Directorate-General of Budget, Accounting, and Statistics, Executive Yuan, Republic of China, Taiwan.

## APPENDIX A

### Detailed Information about Items Measuring Deviance in the Taiwan Youth Project Survey

**Table A1.** Format of the delinquency question groups in the Taiwan Youth Project survey.

Wave	Question stem for wave	Question stem for wave (Chinese)	Dichotomous or Frequency Scale
J1W1	Thinking of the past year, how many of your good friends have done the following?...And how often have you done them? (Please circle a response according to the frequency with which each behavior occurred.)	回想去年一整年，你所有的好朋友中有多少人做了下列事情？…你自己呢？ (請請依照行為發生的頻率勾選)	Frequency
J1W4	In the past six months (since March / April), have the following happened?	過去半年（今年 3、4 月以來），有沒有發生過下列的事情：	Dichotomous
J1W5	In the past year (since last November), have the following happened?	過去一年（前年 11 月以來），有沒有發生過下列的事情	Dichotomous
J1W6	Thinking about the past 12 months, have these good friends [you mentioned] done the following? How about you? (please circle the appropriate response)	回想過去一整年，這些好朋友中有發生以下的行為嗎？你自己呢？(請圈選)	Frequency
J1W8	During the past year, how often have these good friends [you mentioned] done the following? How about you?	過去一年來，這些好朋友之間發生以下行為的頻率如何？你自己呢？	Frequency
J1W9	Thinking about the past 12 months, how often have you done the following?	回想過去一整年，你發生以下行為的頻率？	Frequency
J3W1	Thinking about the past twelve months, how many of your friends have done the following things? How about you? Please select a response according to the frequency with which each behavior occurred.	回想去年一整年，你所有的好朋友中有多少人做了下列事情？你自己呢？請依照行為發生的頻率勾選	Frequency

(continued on next page)

**Table A1 (cont.)** Format of the delinquency question groups in the Taiwan Youth Project survey.

Wave	Question stem for wave	Question stem for wave (Chinese)	Dichotomous or Frequency Scale
J3W2	In the past six months, have the following happened to you? [This question stem is for the dichotomous questions in this wave only]	過去半年，你有沒有發生過下列的事情？	Dichotomous
J3W3	In the past year (since last October), have you been involved in the following?	過去一年（去年10月以來），有沒有發生過下列的事情：	Dichotomous
J3W4	Thinking about the past 12 months, how frequently have these good friends [you mentioned] done the following? How about you?	回想過去一整年，這些好朋友中發生以下行為的頻率？你自己呢？	Frequency
J3W6	Thinking about the past 12 months, how frequently have these good friends [you mentioned] done the following? How about you?	回想過去一整年，這些好朋友中發生以下行為的頻率？你自己呢？	Frequency
J3W7	Thinking about the past 12 months, how often have you done the following?	回想過去一整年，你發生以下行為的頻率？	Frequency

**Table A2.** Response category wordings for the Taiwan Youth Project survey delinquency items.

Type of Scale	Category Wording (English)	Category Wording (Chinese)
Frequency	Never	沒有
	Once or twice	一兩次
	Sometimes	有時候
	Often	常常
	All the time	總是
Dichotomous	Yes	有
	No	沒有

**Table A3.** Wording of the Taiwan Youth Project survey delinquency items.

<b>Wave</b>	<b>Item</b>	<b>English wording</b>	<b>Chinese wording</b>
J1W1 <sup>a</sup>	Running away from home	Running away from home	逃家
	Skipping class	Skipping class	蹺課
	Vandalism	Intentionally damage something not belonging to you	故意破壞不屬於自己的東西
	Battery	Hurt others	打傷別人
	Bullying	Bully others	勒索別人
	Smoking or drinking	Smoking or drinking	抽菸或喝酒
	Betelnut	Chew betelnut	嚼檳榔
	Use drugs	Use medicine (such as [sniffing] strong glue, pain-killers, or amphetamines)	使用藥物 (如強力膠、速賜康、安非他命)
J1W4	Smoking	Did you smoke?	你有沒有抽煙?
	Drinking	Did you drink?	你有沒有喝酒?
	School problem behavior	Did you cause trouble at school?	你有沒有在學校惹麻煩?
	Skipping class	Did you skip class?	你有沒有蹺過課?
	Fighting	Did you get in fights?	你有沒有跟別人打過架?
J1W5	Smoking	Did you smoke?	有沒有抽煙?
	Drinking	Did you drink?	有沒有喝酒?
	School problem behavior	Did you cause trouble at school?	你有沒有在學校惹麻煩?
	Fighting	Did you get in fights?	你有沒有跟別人打過架?
J1W6	Smoking	Smoking	抽煙
	Drinking	Drinking	喝酒
	Running Away from Home	Running Away from Home	逃家
	Fighting	Fighting	打架
J1W8	Fraud	Cheating or fraud	作弊或詐欺
	Mahjong	Playing Mahjong	打麻將
	Drag racing	Drag racing	飆車
	Pornography	Viewing pornographic CD-ROMs or magazines	看色情光碟或書刊
	Betel nut	Chewing betel nut	嚼檳榔

(continued on next page)

**Table A3 (cont.)** Wording of the Taiwan Youth Project survey delinquency items.

<b>Wave</b>	<b>Item</b>	<b>English wording</b>	<b>Chinese wording</b>
J1W9	Vandalism	Intentionally damaged something not belonging to you	故意破壞不屬於自己的東西
	Theft	Theft	偷竊
	Fraud	Cheating or fraud	作弊或詐欺
	Mahjong	Playing Mahjong	打麻將
	Drag racing	Drag racing	飆車
	Pornography	Viewing pornographic CD-ROMs or magazines	看色情光碟或書刊
	Betel nut	Chewing betel nut	嚼檳榔
	Illegal downloading	Illegally downloading software or music	非法下載軟體或音樂
J3W1	Running away from home	Running away from home	逃家
	Skipping class	Skipping class	蹺課
	Vandalism	Purposefully damaging something not belonging to oneself	故意破壞不屬於自己的東西
	Theft	Stealing things	偷東西
	Battery	Hitting and hurting someone	打傷別人
	Smoking or Drinking	Smoking or drinking	抽菸或喝酒
	Betel nut	Chewing betel nut	嚼檳榔
J3W2	Smoking	Did you smoke?	你有沒有抽煙?
	Drinking	Did you drink alcohol?	你有沒有喝酒?
	School problem behavior	Did you cause trouble at school?	你有沒有在學校惹麻煩?
	Theft	Since graduating from middle school, have you stolen anything? [Frequency scale that omits the “all the time” category]	國中畢業後至今，你有沒有偷過東西?
	Battery	Since graduating from middle school, have you struck or beaten someone because they made you angry? [Frequency scale that omits the “all the time” category]	國中畢業後至今，你有沒有因為人家惹你生氣，所以你揍他或打他?

(continued on next page)

**Table A3 (cont.)** Wording of the Taiwan Youth Project survey delinquency items.

<b>Wave</b>	<b>Item</b>	<b>English wording</b>	<b>Chinese wording</b>
J3W3	Smoking	Did you smoke?	你有沒有抽煙？
	Drinking	Did you drink alcohol?	你有沒有喝酒？
	School problem behavior	Did you cause trouble at school?	你有沒有在學校惹麻煩？
	Running Away	Did you run away from home?	你有沒有逃家？
	Skipping class	Did you skip class?	你有沒有蹺過課？
	Fighting	Did you get in fights?	你有沒有跟別人打過架？
J3W4	Running Away from Home	Running Away from Home	逃家
	Theft	Stealing things	偷東西
	Fighting	Fighting	打架
	Smoking or Drinking	Smoking or Drinking	抽煙或喝酒
J3W6	Vandalism	Intentionally damaged something not belonging to you	故意破壞不屬於自己的東西
	Fraud	Cheating or fraud	作弊或詐欺
	Mahjong	Playing Mahjong	打麻將
	Drag racing	Drag racing	飆車
	Pornography	Viewing pornographic CD-ROMs or magazines	看色情光碟或書刊
	Drug use	Using illegal drugs (such as ecstasy or sniffing strong glue)	使用非法藥物（如強力膠、搖頭丸）
J3W7	Vandalism	Intentionally damaged something not belonging to you	故意破壞不屬於自己的東西
	Theft	Stealing things	偷竊
	Fraud	Cheating or fraud	作弊或詐欺
	Mahjong	Playing Mahjong	打麻將
	Drag racing	Drag racing	飆車
	Pornography	Viewing pornographic CD-ROMs or magazines	看色情光碟或書刊
	Betel nut	Chewing betel nut	嚼檳榔

<sup>a</sup> Delinquency in J1W1 is not part of the IRT chapter in this dissertation, but raw delinquency scores are later used to model survey attrition in the chapter on work.

**Table A4.** Fraction of respondents responding to each category of the Taiwan Youth Project survey delinquency items with categorical responses on a frequency scale.

Wave	Item	Number responding	% Not in last year	% Once or twice	% Sometimes	% Often	% All the Time
J1W6	Smoking	1817	1571 (86.5%)	53 (2.9%)	74 (4.1%)	64 (3.5%)	55 (3.0%)
	Drinking	1817	1317 (72.5%)	309 (17.0%)	166 (9.1%)	14 (0.8%)	11 (0.6%)
	Running Away from Home	1817	1780 (98.0%)	37 <sup>‡</sup> (2.0%)	--	--	--
	Fighting	1817	1859 (96.8%)	58 <sup>‡</sup> (3.2%)	--	--	--
J1W8	Fraud	1727	1550 (89.8%)	119 (6.9%)	58 <sup>†</sup> (3.4%)	--	--
	Mahjong	1727	1298 (75.2%)	222 (12.9%)	162 (9.4%)	37 (2.1%)	8 (0.5%)
	Drag racing	1727	1574 (91.1%)	79 (4.6%)	53 (3.1%)	21 <sup>*</sup> (1.2%)	--
	Pornography	1727	1429 (82.7%)	136 (7.9%)	140 (8.1%)	22 <sup>*</sup> (1.3%)	--
	Betel nut	1727	1675 (97.0%)	23 (1.3%)	29 <sup>†</sup> (1.7%)	--	--
J1W9	Vandalism	1875	1716 (91.5%)	131 (7.0%)	28 <sup>†</sup> (1.5%)	--	--
	Theft	1875	1836 (97.9%)	39 <sup>‡</sup> (2.1%)	--	--	--
	Fraud	1875	1486 (79.3%)	280 (14.9%)	93 (5.0%)	16 <sup>*</sup> (0.9%)	--
	Mahjong	1875	1010 (53.9%)	387 (20.6%)	365 (19.5%)	100 (5.3%)	13 (0.7%)
	Drag racing	1875	1595 (85.1%)	154 (8.2%)	95 (5.1%)	20 (1.1%)	11 (0.6%)
	Pornography	1875	1206 (64.3%)	258 (13.8%)	326 (17.4%)	65 (3.5%)	20 (1.1%)
	Betel nut	1875	1760 (93.9%)	57 (3.0%)	44 (2.3%)	14 <sup>*</sup> (0.7%)	--
	Illegal downloading	1875	796 (42.5%)	300 (16.0%)	489 (26.1%)	199 (10.6%)	91 (4.9%)

(continued on next page)

**Table A4 (cont.)** Fraction of respondents responding to each category of the Taiwan Youth Project survey delinquency items with categorical responses on a frequency scale.

Wave	Item	Number responding	% Not in last year	% Once or twice	% Sometimes	% Often	% All the Time
J3W1	Running away from home	2822	2677 (94.9%)	107 (3.8%)	30 (1.1%)	8* (0.3%)	--
	Skipping class	2825	2170 (76.8%)	481 (17.0%)	112 (4.0%)	54 (1.9%)	8 (0.3%)
	Vandalism	2824	2233 (79.1%)	470 (16.6%)	92 (3.3%)	18 (0.6%)	11 (0.4%)
	Theft	2824	2568 (90.9%)	213 (7.5%)	34 (1.2%)	9* (0.3%)	--
	Battery	2823	2448 (86.7%)	302 (10.7%)	56 (2.0%)	17 (0.6%)	--
	Smoking or Drinking	2822	2365 (83.8%)	275 (9.7%)	109 (3.9%)	49 (1.7%)	24 (0.9%)
	Betel nut	2820	2709 (96.1%)	77 (2.7%)	34 <sup>†</sup> (1.2%)	--	--
J3W2	Theft	2540	2485 (97.8%)	55 <sup>**</sup> (2.2%)	--	--	--
	Battery	2541	2376 (93.5%)	94 (3.7%)	71 <sup>***</sup> (2.8%)	--	--
J3W4	Running Away from Home	2039	1971 (96.7%)	68 <sup>†</sup> (3.3%)	--	--	--
	Theft	2039	1984 (97.3%)	55 <sup>†</sup> (2.7%)	--	--	--
	Fighting	2039	1956 (95.9%)	83 <sup>†</sup> (4.1%)	--	--	--
	Smoking or Drinking	2038	1709 (83.9%)	121 (5.9%)	112 (5.5%)	56 (2.7%)	40 (2.0%)
J3W6	Vandalism	1841	1756 (95.4%)	70 (3.8%)	15 <sup>†</sup> (0.8%)	--	--
	Fraud	1841	1575 (85.6%)	188 (10.2%)	67 (3.6%)	11* (0.6%)	--
	Mahjong	1841	1554 (84.4%)	163 (8.9%)	100 (5.4%)	24* (1.3%)	--
	Drag racing	1841	1671 (90.8%)	94 (5.1%)	51 (2.8%)	13 (0.7%)	12 (0.7%)
	Pornography	1841	1448 (78.7%)	208 (11.3%)	152 (8.3%)	25 (1.4%)	8 (0.4%)
	Drug use	1841	1779 (96.6%)	36 (2.0%)	26 <sup>†</sup> (1.4%)	--	--

(Continued on next page)

**Table A4 (cont.)** Fraction of respondents responding to each category of the Taiwan Youth Project survey delinquency items with categorical responses on a frequency scale.

Wave	Item	Number responding	% Not in last year	% Once or twice	% Sometimes	% Often	% All the Time
J3W7	Vandalism	1873	1716 (91.6%)	125 (6.7%)	32 <sup>†</sup> (1.7%)	--	--
	Theft	1873	1832 (97.8%)	41 <sup>‡</sup> (2.2%)	--	--	--
	Fraud	1873	1506 (80.4%)	275 (14.7%)	82 (4.4%)	10 <sup>*</sup> (0.5%)	--
	Mahjong	1873	1350 (72.1%)	275 (14.7%)	207 (11.1%)	41 <sup>*</sup> (2.2%)	--
	Drag racing	1872	1607 (85.8%)	141 (7.5%)	89 (4.8%)	28 (1.5%)	7 (0.4%)
	Pornography	1873	1220 (65.1%)	262 (14.0%)	323 (17.2%)	51 (2.7%)	17 (0.9%)
	Betel nut	1873	1762 (94.1%)	47 (2.5%)	46 (2.5%)	18 <sup>*</sup> (1.0%)	--

\* Due to too few respondents selecting some categories, the categories for “often” and “all the time” have been collapsed together.

\*\* Due to too few respondents selecting some categories, “all the time,” “sometimes,” and “once or twice” have been Collapsed together.

\*\*\* Due to too few respondents selecting some categories, “often” and “sometimes” have been collapsed together.

† Due to too few respondents selecting some categories, the categories for “sometimes,” “often,” and “all the time” have been collapsed together.

‡ Due to too few respondents selecting some categories, the categories for “once or twice,” “sometimes,” “often,” and “all the time” have been collapsed together.

**Table A5.** Fraction of respondents responding to each category of the Taiwan Youth Project survey delinquency items with dichotomous responses.

<b>Wave</b>	<b>Item</b>	<b>Number responding</b>	<b>Number responding no</b>	<b>Number responding yes</b>
J1W4	Smoking	2348	2224 (94.7%)	124 (5.3%)
	Drinking	2348	2198 (93.6%)	150 (6.4%)
	School Problem Behavior	2281	2175 (95.4%)	106 (4.6%)
	Skiping Class	2280	2088 (91.6%)	192 (8.4%)
	Fighting	2349	2222 (94.6%)	127 (5.4%)
J1W5	Smoking	2010	1841 (91.6%)	169 (8.4%)
	Drinking	2010	1806 (89.9%)	204 (10.1%)
	School Problem Behavior	1939	1819 (93.8%)	120 (6.2%)
	Fighting	2010	1902 (94.6%)	108 (5.4%)
J3W2	Smoking	2541	2369 (93.2%)	172 (6.8%)
	Drinking	2533	2432 (96.0%)	101 (4.0%)
	School Problem Behavior	2540	2470 (97.2%)	70 (2.8%)
J3W3	Smoking	2447	2200 (89.9%)	247 (10.1%)
	Drinking	2437	2217 (91.0%)	220 (9.0%)
	School problem behavior	2353	2242 (95.3%)	111 (4.7%)
	Running Away	2449	2389 (97.55%)	60 (2.45%)
	Skiping class	2356	1926 (81.7%)	430 (18.3%)
	Fighting	2446	2289 (93.6%)	157 (6.4%)

**Table A6.** Results of one-dimensional confirmatory factor analyses for each cohort-wave in the Taiwan Youth Survey, along with Cronbach's alpha reliabilities.

Wave	Variable	Standardized loading	Unstandardized loading	Error variance
J3W1 (alpha = .765)	Running away from home <sup>a</sup>	0.79	1.11 (0.04)	.37 (.05)
	Skipping class <sup>a</sup>	0.73	0.97 (0.04)	.47 (.04)
	Vandalism <sup>*</sup>	0.75	1.00	.44 (.03)
	Theft	0.77	1.03 (0.04)	.40 (.04)
	Battery	0.82	1.09 (0.04)	.33 (.04)
	Smoking or Drinking <sup>b</sup>	0.78	1.03 (0.04)	.40 (.04)
	Betel nut <sup>b</sup>	0.87	1.16 (0.05)	.24 (.05)
	J1W4 (alpha=.639)	Smoking	0.88	1.41 (0.13)
Drinking		0.79	1.25 (0.11)	0.38 (0.06)
School Problem Behavior <sup>*,a</sup>		0.63	1.00	0.61 (0.07)
Skipping Class		0.76	1.22 (0.12)	0.42 (0.06)
Fighting <sup>a</sup>		0.71	1.13 (0.09)	0.49 (0.07)
J1W5 (alpha = .579)	Smoking	0.88	1.07 (0.08)	0.22 (0.07)
	Drinking	0.73	0.89 (0.06)	0.46 (0.06)
	School Problem Behavior <sup>*</sup>	0.73	1.00	0.46 (0.07)
	Fighting	0.79	0.89 (0.08)	0.32 (0.07)
J1W6 (alpha = .566)	Smoking	0.87	1.11 (0.08)	0.24 (0.06)
	Drinking	0.74	0.74 (0.06)	0.45 (0.05)
	Running Away from Home	0.63	0.80 (0.08)	0.60 (0.09)
	Fighting <sup>*</sup>	0.79	1.00	0.38 (0.07)

(Continued on next page)

**Table A6 (cont.)** Results of one-dimensional confirmatory factor analyses for each cohort-wave in the Taiwan Youth Survey, along with Cronbach's alpha reliabilities.

Wave	Variable	Standardized loading	Unstandardized loading	Error variance
J1W8 (alpha = .478)	Fraud	0.49	0.73 (0.09)	0.76 (0.06)
	Mahjong	0.48	0.73 (0.08)	0.77 (0.05)
	Drag racing	0.82	1.23 (0.10)	0.33 (0.07)
	Pornography*	0.66	1.00	0.56 (0.06)
	Betel nut	0.71	1.06 (0.10)	0.50 (0.09)
J1W9 (alpha = .565)	Vandalism	0.62	1.29 (0.11)	0.61 (0.05)
	Theft	0.65	1.33 (0.16)	0.58 (0.09)
	Fraud	0.64	1.33 (0.11)	0.59 (0.05)
	Mahjong	0.45	0.93 (0.09)	0.80 (0.04)
	Drag racing <sup>a</sup>	0.67	1.39 (0.12)	0.55 (0.05)
	Pornography <sup>b</sup>	0.57	1.19 (0.08)	0.67 (0.04)
	Betel nut <sup>a</sup>	0.45	0.93 (0.13)	0.80 (0.06)
	Illegal downloading <sup>*,b</sup>	0.48	1.00	0.77 (0.04)
J3W2 (alpha = .372)	Smoking*	0.83	1.00	0.31 (0.08)
	Drinking	0.79	0.94 (0.10)	0.38 (0.09)
	School problem behavior	0.68	0.82 (0.09)	0.54 (0.08)
	Theft	0.26	0.32 (0.10)	0.93 (0.05)
	Battery	0.45	0.54 (0.07)	0.80 (0.05)

(Continued on next page)

**Table A6 (cont.)** Results of one-dimensional confirmatory factor analyses for each cohort-wave in the Taiwan Youth Survey, along with Cronbach's alpha reliabilities.

Wave	Variable	Standardized loading	Unstandardized loading	Error variance
J3W3 (alpha = .629)	Smoking	0.83	1.07 (0.06)	0.31 (0.05)
	Drinking	0.73	0.94 (0.06)	0.46 (0.05)
	School problem behavior	0.74	0.89 (0.07)	0.52 (0.06)
	Running Away	0.78	0.95 (0.08)	0.45 (0.09)
	Skipping class	0.69	0.91 (0.06)	0.50 (0.05)
	Fighting*	0.71	1.00	0.39 (0.06)
	J3W4 (alpha = .311)	Running Away from Home	0.81	1.25 (0.10)
Theft		0.75	1.17 (0.11)	0.44 (0.09)
Fighting		0.94	1.46 (0.12)	0.12 (0.07)
Smoking or Drinking*		0.64	1.00	0.59 (0.05)
J3W6 (alpha = .579)	Vandalism	0.57	1.07 (0.13)	0.68 (0.06)
	Fraud <sup>a</sup>	0.70	1.33 (0.12)	0.51 (0.05)
	Mahjong*	0.53	1.00	0.72 (0.05)
	Drag racing	0.69	1.32 (0.12)	0.52 (0.06)
	Pornography	0.77	1.47 (0.13)	0.41 (0.05)
	Drug use <sup>a</sup>	0.59	1.13 (0.14)	0.65 (0.07)

(Continued on next page)

**Table A6 (cont.)** Results of one-dimensional confirmatory factor analyses for each cohort-wave in the Taiwan Youth Survey, along with Cronbach's alpha reliabilities.

Wave	Variable	Standardized loading	Unstandardized loading	Error variance
J3W7 (alpha = .576)	Vandalism	0.70	1.51 (0.14)	0.51 (0.06)
	Theft	0.77	1.65 (0.17)	0.41 (0.09)
	Fraud <sup>a</sup>	0.57	1.24 (0.12)	0.67 (0.05)
	Mahjong <sup>*</sup>	0.47	1.00	0.78 (0.04)
	Drag racing <sup>b</sup>	0.65	1.40 (0.13)	0.58 (0.05)
	Pornography	0.70	1.52 (0.14)	0.51 (0.05)
	Betel nut <sup>a,b</sup>	0.55	1.18 (0.13)	0.70 (0.06)

\* Survey item used as a referent variable to define the scale for a particular wave.

<sup>a,b</sup> Denotes pairs of survey items with correlated errors in each wave.

**Table A7.** Discrimination and item difficulty parameter estimates, with 95% confidence intervals, for Taiwan Youth Project survey delinquency items with dichotomous responses.

<b>Wave</b>	<b>Item</b>	<b>Discrimination Parameter</b>	<b>Item difficulty parameter</b>
J1W4	Smoking	1.19 (1.02,1.38)	5.28 (4.74, 5.91)
	Drinking	0.97 (0.83,1.11)	4.40 (4.01, 4.81)
	School Problem Behavior	1.17 (1.00,1.37)	5.37 (4.82, 6.01)
	Skipping Class	1.03 (0.89,1.19)	4.14 (3.78, 4.52)
	Fighting	0.94 (0.80,1.09)	4.55 (4.16, 4.99)
J1W5	Smoking	1.69 (1.38, 1.98)	5.08 (4.39, 5.79)
	Drinking	1.16 (1.00, 1.34)	3.69 (3.34, 4.06)
	School Problem Behavior	1.06 (0.88, 1.26)	4.09 (3.71, 4.53)
	Fighting	1.33 (1.11, 1.58)	4.94 (4.44, 5.45)
J3W2	Smoking	1.20 (0.97, 1.46)	4.96 (4.33, 5.75)
	Drinking	1.11 (0.90, 1.34)	5.45 (4.79, 6.25)
	School Problem Behavior	0.86 (0.67, 1.07)	5.14 (4.60, 5.80)
J3W3	Smoking	1.00 (0.88, 1.13)	4.34 (3.98, 4.76)
	Drinking	0.80 (0.71, 0.90)	3.95 (3.65, 4.28)
	School problem behavior	0.72 (0.61, 0.84)	4.50 (4.13, 4.91)
	Running Away from Home	0.78 (0.64, 0.94)	5.52 (4.97, 6.14)
	Skipping class	0.80 (0.70, 0.90)	2.76 (2.56, 2.97)
	Fighting	0.86 (0.75, 0.97)	4.60 (4.21, 5.02)

**Table A8.** Discrimination and item difficulty parameter estimates, with 95% confidence intervals, for Taiwan Youth Project survey delinquency items with polytomous responses.

Wave	Item	Discrimination Parameter	$b_{\text{once or twice}}$	$b_{\text{sometimes}}$	$b_{\text{often}}$	$b_{\text{all the time}}$
J1W6	Smoking	0.38 (0.30, 0.47)	4.17 (3.86, 4.51)	1.44 (0.81, 2.14)	2.73 (1.98, 3.55)	3.14 (2.36, 4.03)
	Drinking	0.27 (0.23, 0.31)	1.97 (1.84, 2.11)	1.69 (1.42, 1.98)	4.17 (3.64, 4.77)	2.35 (1.63, 3.08)
	Running Away from Home	0.40 (0.31, 0.50)	5.31 <sup>‡</sup> (4.73, 5.94)	--	--	--
	Fighting	0.57 (0.46, 0.67)	5.80 <sup>‡</sup> (5.12, 6.55)	--	--	--
J1W8	Fraud	0.14 (0.11, 0.17)	3.01 (2.82, 3.22)	1.49 <sup>†</sup> (1.13, 1.87)	--	--
	Mahjong	0.10 (0.08, 0.12)	2.03 (1.89, 2.16)	0.71 (0.51, 0.91)	2.06 (1.71, 2.41)	2.19 (1.55, 2.85)
	Drag racing	0.26 (0.21, 0.32)	3.94 (3.63, 4.29)	2.43 (1.78, 3.17)	3.54 <sup>*</sup> (2.76, 4.36)	--
	Pornography	0.19 (0.15, 0.22)	2.88 (2.70, 3.06)	1.01 (0.68, 1.37)	3.36 <sup>*</sup> (2.84, 3.93)	--
	Betel nut	0.21 (0.16, 0.27)	5.08 (4.65, 5.57)	1.45 <sup>†</sup> (0.70, 2.27)	--	--
J1W9	Vandalism	0.10 (0.082, 0.12)	3.33 (3.09, 3.58)	2.91 <sup>†</sup> (2.38, 3.46)	--	--
	Theft	0.13 (0.091, 0.17)	4.97 <sup>‡</sup> (4.46, 5.54)	--	--	--
	Fraud	0.089 (0.075, 0.103)	2.21 (2.07, 2.36)	2.09 (1.78, 2.40)	3.14 <sup>*</sup> (2.61, 3.69)	--
	Mahjong	0.042 (0.034, 0.050)	1.16 (1.06, 1.26)	0.34 (0.21, 0.47)	1.67 (1.47, 1.87)	2.48 (2.01, 2.98)
	Drag racing	0.097 (0.081, 0.113)	2.95 (2.77, 3.14)	1.63 (1.25, 2.01)	3.17 (2.62, 3.74)	2.44 (1.75, 3.14)
	Pornography	0.084 (0.071, 0.096)	1.92 (1.79, 2.04)	0.43 (0.25, 0.62)	2.68 (2.37, 3.01)	2.56 (2.07, 3.05)
	Betel nut	0.064 (0.048, 0.079)	3.83 (3.59, 4.09)	0.97 (0.56, 1.40)	2.04 <sup>*</sup> (1.46, 2.65)	--
	Illegal downloading	0.065 (0.056, 0.075)	1.21 (1.10, 1.32)	-0.14 (-0.27, -0.01)	1.44 (1.27, 1.63)	1.58 (1.33, 1.86)

(Continued on next page)

**Table A8 (cont.)** Discrimination and item difficulty parameter estimates, with 95% confidence intervals, for Taiwan Youth Project survey delinquency items with polytomous responses.

Wave	Item	Discrimination Parameter	$b_{\text{once or twice}}$	$b_{\text{sometimes}}$	$b_{\text{often}}$	$b_{\text{all the time}}$
J3W1	Running away from home	.19 (.17, .22)	4.67 (4.32, 4.99)	4.27 (3.66, 4.97)	4.69* (3.97, 5.57)	--
	Skipping class	.15 (.13, .16)	2.21 (2.09, 2.32)	3.01 (2.72, 3.32)	3.09 (2.68, 3.51)	4.41 (3.82, 5.03)
	Vandalism	.15 (.13, .16)	2.31 (2.18, 2.44)	3.29 (2.98, 3.59)	4.03 (3.52, 4.56)	3.16 (2.57, 3.77)
	Theft	.16 (.14, .18)	3.54 (3.34, 3.72)	4.08 (3.55, 4.52)	4.02* (3.38, 4.67)	--
	Battery	.21 (.19, .23)	3.45 (3.24, 3.67)	4.77 (4.29, 5.29)	5.00* (4.43, 5.63)	--
	Smoking or Drinking	.16 (.14, .18)	2.99 (2.84, 3.14)	2.78 (2.38, 3.17)	3.46 (3.01, 3.96)	3.65 (3.07, 4.24)
	Betel nut	.26 (.23, .30)	5.99 (5.48, 6.54)	5.41† (4.56, 6.30)	--	--
	J3W2	Theft	0.32 (0.13, 0.52)	4.20** (3.85, 4.60)	--	--
Battery		0.33 (0.24, 0.42)	3.62 (3.40, 3.84)	.96*** (0.59, 1.36)	--	--
J3W4	Running Away from Home	.69 (.55, .84)	5.22‡ (4.64, 5.90)	--	--	--
	Theft	.56 (.43, .70)	4.96‡ (4.42, 5.58)	--	--	--
	Fighting	1.01 (.83, 1.20)	6.41‡ (5.58, 7.40)	--	--	--
	Smoking or Drinking	.30 (.23, .37)	3.05 (2.87, 3.24)	.82 (.50, 1.17)	1.85 (1.39, 2.37)	1.80 (1.26, 2.41)
J3W6	Vandalism	.19 (.15, .23)	4.10 (3.78, 4.45)	3.15† (2.46, 3.88)	--	--
	Fraud	.19 (.16, .22)	2.85 (2.67, 3.05)	2.50 (2.10, 2.94)	3.74* (3.09, 4.42)	--
	Mahjong	.12 (.10, .14)	2.69 (2.53, 2.85)	1.25 (0.98, 1.53)	2.50* (2.04, 2.96)	--
	Drag racing	.17 (.14, .20)	3.56 (3.32, 3.80)	1.94 (1.48, 2.42)	3.18 (2.52, 3.87)	2.12 (1.35, 2.88)
	Pornography	.22 (.18, .26)	2.65 (2.47, 2.84)	1.88 (1.48, 2.33)	4.16 (3.56, 4.80)	3.68 (2.91, 4.47)
	Drug use	.16 (.12, .21)	4.61 (4.26, 4.99)	1.69† (1.03, 2.40)	--	--

(Continued on next page)

**Table A8 (cont.)** Discrimination and item difficulty parameter estimates, with 95% confidence intervals, for Taiwan Youth Project survey delinquency items with polytomous responses.

Wave	Item	Discrimination Parameter	$b_{\text{once or twice}}$	$b_{\text{sometimes}}$	$b_{\text{often}}$	$b_{\text{all the time}}$
J3W7	Vandalism	.19 (.16, .23)	3.71 (3.42, 4.01)	3.54 (2.93, 4.20)	--	--
	Theft	.26 (.21, .33)	5.74 (5.08, 6.50)	--	--	--
	Fraud	.12 (.10, .14)	2.21 (2.07, 2.35)	2.14 (1.85, 2.44)	3.36 (2.75, 3.97)	--
	Mahjong	.082 (.069, .095)	1.89 (1.77, 2.02)	0.76 (0.58, 0.95)	2.29 (1.96, 2.62)	3.02 (2.84, 3.20)
	Drag racing	.14 (.12, .16)	3.02 (2.84, 3.20)	1.60 (1.25, 1.98)	2.82 (2.33, 3.34)	3.18 (2.45, 3.93)
	Pornography	.13 (.12, .15)	1.94 (1.82, 2.07)	0.55 (0.37, 0.75)	3.13 (2.78, 3.50)	2.76 (2.22, 3.32)
	Betel nut	.11 (.092, .14)	4.15 (3.88, 4.42)	1.03 (0.58, 1.49)	2.26 (1.67, 2.85)	--

\* Due to too few respondents selecting some categories, the categories for “often” and “all the time” have been collapsed together.

\*\* Due to too few respondents selecting some categories, “all the time,” “sometimes,” and “once or twice” have been collapsed together.

\*\*\* Due to too few respondents selecting some categories, “often” and “sometimes” have been collapsed together.

† Due to too few respondents selecting some categories, the categories for “sometimes,” “often,” and “all the time” have been collapsed together.

‡ Due to too few respondents selecting some categories, the categories for “once or twice,” “sometimes,” “often,” and “all the time” have been collapsed together.

**Table A9.** Number of independent draws from the posterior distribution of parameter estimates for each cohort-wave in the Taiwan Youth Survey.

Wave	Number of Independent Posterior Draws
J3W1	412 <sup>a</sup>
J3W2	3,174
J3W3	3,934
J3W4	3,075
J3W6	3,500
J3W7	2,998
J1W4	4,025
J1W5	2,972
J1W6	3,289
J1W8	3,000
J1W9	3,420

<sup>a</sup> The purpose of modeling deviance for the J3W1 cohort is to obtain a prior distribution in order to fit the other cohort-waves, so fewer posterior draws are needed for this wave.

## APPENDIX B

### Technical Details Concerning Implementation of the Bayesian Generalized Partial Credit Model for the TYP Data

#### Parameter Estimation Using Markov Chain Monte Carlo Methods

Following Patz and Junker (1999a:350), in order to make estimation algorithms easier to write and test, this research uses the following reparameterization of Muraki's Generalized Partial Credit Model:

$$P_{ijk}(\theta_i) = P(X_{ij} = k | \theta_i, \beta_{2j}, \beta_{\delta_{*j}}) = \frac{\exp \sum_{l=1}^k (\beta_{2j} \theta_i - \beta_{\delta_{lj}})}{\sum_{m=1}^k \exp \sum_{l=1}^m (\beta_{2j} \theta_i - \beta_{\delta_{lj}})} \quad (\text{Equation B1}),$$

where  $\beta_{2j}$  denotes the item discrimination parameter,  $\beta_{\delta_{lj}} / \beta_{2j}$  is the item step parameter  $\beta_{lj}$  in Equation 2.1,  $\theta_i$  refers to respondent  $i$ 's ability, and  $\beta_{\delta_{lj}}$  (that is, the first item step parameter for item  $j$ ) is set to zero for all  $j$  to make the model identifiable. Parameter estimation proceeds by sampling observations of  $\theta_i$ ,  $\beta_{\delta_{*j}}$ , and  $\beta_{2j}$  from a stationary Markov chain  $M_0, M_1, M_2, \dots$  such that  $M_k = \{\theta^k, \beta^k\}$ , and such that for the stationary distribution  $\pi(\theta, \beta)$  (Patz and Junker 1999b:151),

$$\int_{\theta, \beta} t[(\theta^0, \beta^0), (\theta^1, \beta^1)] \pi(\theta^0, \beta^0) d(\theta^0, \beta^0) = \pi(\theta^1, \beta^1), \quad (\text{Equation B2})$$

where  $t[(\theta^0, \beta^0), (\theta^1, \beta^1)] = P[M_{k+1} = (\theta^1, \beta^1) | M_k = (\theta^0, \beta^0)]$ , the probability of moving to a new state of the Markov chain given the value of the previous state, is the Markov chain's transition kernel.

To sample from the posterior distributions of the parameters, Patz and Junker cite findings that the parameters need not be drawn together at the same time from a joint posterior distribution: instead, they can be drawn in blocks in whatever order is convenient to the researcher (Chib and Greenberg 1995). Additionally, they make a few assumptions widely assumed in IRT research and survey research in general: first, individual abilities are conditionally independent given the item difficulties; and second, item difficulties are conditionally independent given individual abilities. They use the following general formula (Patz and Junker 1999a:344):

$$\begin{aligned}
p(X|\theta, \beta, \lambda, Y, Z) &= \int p(X|\theta, \beta, \lambda, Y, Z)p(\theta|\beta, \lambda, Y, Z)d\theta \\
&= \int p(X|\theta, \beta, Y, Z)p(\theta|\beta, Y, Z)d\theta \\
&= \int p(X|\theta, \beta, Z)p(\theta|\lambda, Y)d\theta \\
&= \prod_{i=1}^I \int \prod_{j=1}^J p(X_{ij}|\theta_i, \beta_j, Z_j)p(\theta_i|Y_i)d\theta_i, \quad (\text{Equation B3})
\end{aligned}$$

where in addition to the individual abilities  $\theta$  and item difficulties and item parameters  $\beta$  (which include both the discrimination parameters and item step parameters),  $\lambda$  refers to any hyperparameters associated with the distribution of individual abilities,  $Y$  denotes subject-specific covariates that may predict individual abilities,  $Z$  denotes item-specific covariates that may predict the difficulty of the test items, and  $X_{ij}$  refers to respondent  $i$ 's response to test item  $j$ . The present research does not make use of subject-specific covariates to predict abilities, saving them instead to explain variation in the latent variable for criminality during later chapters. Item-specific predictors  $Z$  are also not included, so Equation B3 reduces to

$$p(X|\theta, \beta, \lambda, Y, Z) = \prod_{i=1}^I \int \prod_{j=1}^J p(X_{ij}|\theta_i, \beta_j) p(\theta_i) d\theta_i. \quad (\text{Equation B4})$$

The joint posterior distribution  $p(\theta, \beta|X)$  can then be expressed as

$$\begin{aligned} p(\theta, \beta|X) &\propto p(X|\theta, \beta)p(\theta, \beta) \\ &= p(X|\theta, \beta)p(\theta)p(\beta), \end{aligned} \quad (\text{Equation B5})$$

where moving from the first line to the second line in Equation 2.6 involves assuming that the abilities and item difficulties are independent. Sampling from the joint posterior distribution is straightforward because  $\pi(\theta, \beta) = p(\theta, \beta|X)$  is the stationary distribution for the transition kernel  $t[(\theta^0, \beta^0), (\theta^1, \beta^1)] = p(\theta^1|\beta^0, X)p(\beta^1|\theta^1, X)$ , which corresponds to a Gibbs sampler (Patz and Junker 1999b:152). Additionally, some of the conditional independence assumptions used in this research facilitate a “block-at-a-time” implementation of the Metropolis-Hastings algorithm explained in Chib and Greenberg (1995). In brief, for two variables  $x_1$  and  $x_2$ , if two conditional transition kernels  $P_1(x_1, dy_1|x_2)$  and  $P_2(x_2, dy_2|x_1)$  can be specified, with invariant densities  $\pi_{1|2}^*(dy_1|x_2) = \int P_1(x_1, dy_1|x_2)\pi_{1|2}(x_1|x_2)dx_1$  and  $\pi_{2|1}^*(dy_2|x_1) = \int P_2(x_2, dy_2|x_1)\pi_{2|1}(x_2|x_1)dx_2$ , then the product of these two transition kernels has invariant density  $\pi(x_1, x_2)$ , and this property can be generalized to joint posterior densities of any number of variables. In other words, to get posterior draws of parameters of interest from their joint posterior density, a researcher can group the parameters into whatever blocks of conditional densities best facilitate convergence of the Markov chain, and then obtain draws from the conditional posterior densities. Drawing from the joint posterior distribution  $p(\theta, \beta)$  can thus be conducted by first drawing  $\theta^k \sim p(\theta|X, \beta^{k-1})$ , followed by  $\beta^k \sim p(\beta|X, \theta^k)$ . Equation B6 gives the probability distributions for  $p(\theta|X, \beta^{k-1})$  and  $p(\beta|X, \theta^k)$ :

$$p(\theta|X, \beta) = \frac{p(X|\theta, \beta)p(\theta, \beta)}{\int p(X|\theta, \beta)p(\theta, \beta)d\theta} \quad \text{and} \quad p(\beta|X, \theta) = \frac{p(X|\theta, \beta)p(\theta, \beta)}{\int p(X|\theta, \beta)p(\theta, \beta)d\theta}. \quad (\text{Equation B6})$$

In practice, the normalizing constants in the denominators in Equation B6 are intractable, especially since this research uses complex mixture models for informative prior distributions for  $\theta$ . A solution is to use a Metropolis-Hastings (M-H) algorithm to sample from  $p(\theta|X, \beta)$  and  $p(\beta|X, \theta)$ . The idea here is to choose a proposal distribution  $q$  with corresponding proposal transition kernel  $q[(\theta^0, \beta^0), (\theta^1, \beta^1)]$ , then select  $(\theta^k, \beta^k) = (\theta^*, \beta^*)$  with probability

$$\alpha[(\theta^0, \beta^0), (\theta^*, \beta^*)] = \min \left\{ \frac{\pi(\theta^*, \beta^*)q[(\theta^*, \beta^*), (\theta^0, \beta^0)]}{\pi(\theta^0, \beta^0)q[(\theta^0, \beta^0), (\theta^*, \beta^*)]}, 1 \right\}, \quad (\text{Equation B7})$$

and let  $(\theta^k, \beta^k) = (\theta^{k-1}, \beta^{k-1})$  otherwise. If  $q$  is a symmetric probability distribution, then  $q[(\theta^*, \beta^*), (\theta^0, \beta^0)] = q[(\theta^0, \beta^0), (\theta^*, \beta^*)]$  and the term drops out of Equation B7 (Chib and Greenberg 1995:329). Because  $\theta^*$  and  $\beta^*$  can be drawn during separate transition steps, a researcher can therefore choose different proposal distributions  $q$  for  $\theta$  and  $\beta$ , important because the choice of  $q$  affects the rate at which the Markov chain reaches a stationary state (Chib and Greenberg 1995; Patz and Junker 1999b). Also, given the assumption that individual abilities  $\theta_i$  are independent, as are the discrimination and item step parameters  $\beta_j$ , it becomes possible to sample parameters from the posterior distribution for one individual at a time and for one item at a time, a fact which also greatly speeds convergence of the Markov chain (Patz and Junker 1999b).

## Addressing Missing Values

In a follow-up to their first article outlining Bayesian Item Response Theory model estimation using MCMC, Patz and Junker (1999a) outline a technique for estimating ability scores for respondents who may not have responded to all the test items. The approach involves generating separating the set of responses to test items  $X$  into those responses that were observed,  $X_{obs}$ , and those that are missing,  $X_{miss}$ , letting  $M$  denote an  $I$  by  $J$  matrix of 0's and 1's, such that  $M_{ij} = 0$  if respondent  $i$  responded to item  $j$  and 1 if the response is missing. Incorporating  $X_{miss}$  into Equation B5 as a parameter to be estimated gives

$$p(\theta, \beta, X_{miss} | X_{obs}, M) \\ \propto p(M | \theta, \beta, X_{miss}, X_{obs}) p(X_{obs} | \theta, \beta, X_{miss}) p(X_{miss} | \theta, \beta) p(\theta, \beta). \quad (\text{Equation B8})$$

The first term drops out if a researcher makes the strong assumption that  $M$  is independent of  $\theta$ ,  $\beta$ ,  $X_{miss}$ , and  $X_{obs}$  (Patz and Junker 1999a:351), an assumption stronger than MCAR (missing completely at random) but which holds in the case of responses that are missing by design (Mislevy and Wu 1996) – such as, for Wave 4 of the J1 cohort study in Table 2.1, 137 out of 139 missing responses that in nearly all cases were associated with respondents who not in school who therefore did not respond to delinquency items concerning school offenses. Fortunately in this study, the Taiwan Youth Project team was extremely thorough in conducting interviews and few responses are missing except for those that are missing by design. In a few exceptional cases, however, a respondent demonstrated patterns of missingness (such as answering the first two questions about delinquency and skipping the rest) suggesting that his or her temperament may have led to the questions being missing. Lord (1983:477) points out that when they are due to temperament, such missingness patterns violate the IRT assumption that item difficulties for

separate questions are independent. Accordingly, in order for the model assumptions to hold I exclude from analysis the handful of respondents for which half or more of the responses are missing (no more than two or three per wave). The remaining missing cases always constitute fewer than half of one percent of respondents in each wave, so they are treated in the present research as a nuisance rather than explicitly attempting to model their missingness. When possible, missing values are first imputed using multiple imputation in PRELIS as an initial pre-processing step (Jöreskog and Sörbom 2002). For those few remaining cases (usually one or two test items for two or three respondents per wave), the present analysis assumes that  $M$  is independent of  $\theta$ ,  $\beta$ ,  $X_{miss}$ , and  $X_{obs}$  in Equation B8.

Patz and Junker (1999a) proceed by treating  $X_{miss}$  as a parameter and drawing it from the posterior distribution in Equation B8. Due to the assumption that  $X_{obs}$  and  $X_{miss}$  are independent, and also that  $X_{miss}$  and the ability and item parameters are independent, draws of  $X_{miss}$  can be treated as a separate step in a Gibbs sampler, involving a stationary distribution with transition kernel

$$\begin{aligned}
& t[(X_{miss}^0, \theta^0, \beta^0), (X_{miss}^1, \theta^1, \beta^1)] \\
&= p(X_{miss}^1 | \theta^0, \beta^0, X_{obs}, M) p(\theta^1 | X_{miss}^1, \beta^0, X_{obs}, M) p(\beta^1 | X_{miss}^1, \theta^1, X_{obs}, M) \\
&= p(X_{miss}^1 | \theta^0, \beta^0) p(\theta^1 | \beta^0, X_{miss}^1, X_{obs}) p(\beta^1 | \theta^1, X_{miss}^1, X_{obs}). \quad (\text{Equation B9})
\end{aligned}$$

Note that in Equation B9,  $M$  is assumed to be independent of  $\theta$ ,  $\beta$ ,  $X_{miss}$ , and  $X_{obs}$ ,

$$p(X_{miss}^1 | \theta^0, \beta^0, X_{obs}) = p(X_{miss}^1 | \theta^0, \beta^0) \quad (\text{Equation B10})$$

because  $X_{miss}$  and  $X_{obs}$  are assumed to be independent,

$$p(\theta^1 | \beta^0, X_{miss}^1, X_{obs}) \propto p(X_{obs} | \theta^1, \beta^0) p(X_{miss}^1 | \theta^1, \beta^0) p(\theta^1) \quad (\text{Equation B11})$$

because  $\theta$  and  $\beta$  are assumed to be independent, and

$$p(\beta^1 | \theta^1, X_{miss}^1, X_{obs}) \propto p(X_{obs} | \theta^1, \beta^1) p(X_{miss}^1 | \theta^1, \beta^1) p(\beta^1) \quad (\text{Equation B12})$$

for the same reason. Taking together Equations B9-B12, Patz and Junker (1999a:351) propose the following algorithm for obtaining posterior draws of  $\theta$  and  $\beta$  in the presence of missing data:

1. Draw  $\mathbf{X}_{miss}^m \sim p(\mathbf{X}_{miss} | \boldsymbol{\theta}^{m-1}, \boldsymbol{\beta}^{m-1})$ .
- 2.1. Draw  $\theta_1^m \sim p(\theta_1 | \boldsymbol{\beta}^{m-1}, \boldsymbol{\theta}_{>1}^{m-1}, \mathbf{X}_{miss}^m, \mathbf{X}_{obs})$ .
- 2.2. Draw  $\theta_2^m \sim p(\theta_2 | \boldsymbol{\beta}^{m-1}, \boldsymbol{\theta}_{<2}^m, \boldsymbol{\theta}_{>2}^{m-1}, \mathbf{X}_{miss}^m, \mathbf{X}_{obs})$ .
- ⋮
- 2.I. Draw  $\theta_I^m \sim p(\theta_I | \boldsymbol{\beta}^{m-1}, \boldsymbol{\theta}_{<I}^m, \mathbf{X}_{miss}^m, \mathbf{X}_{obs})$ .
- 3.1. Draw  $\boldsymbol{\beta}_I^m \sim p(\boldsymbol{\beta}_I | \mathbf{X}_{miss}^m, \mathbf{X}_{obs}, \boldsymbol{\theta}^m, \boldsymbol{\beta}_{>I}^{m-1})$ .
- 3.2. Draw  $\boldsymbol{\beta}_2^m \sim p(\boldsymbol{\beta}_2 | \mathbf{X}_{miss}^m, \mathbf{X}_{obs}, \boldsymbol{\theta}^m, \boldsymbol{\beta}_{<2}^m, \boldsymbol{\beta}_{>2}^{m-1})$ .
- ⋮
- 3.J. Draw  $\boldsymbol{\beta}_J^m \sim p(\boldsymbol{\beta}_J | \mathbf{X}_{miss}^m, \mathbf{X}_{obs}, \boldsymbol{\theta}^m, \boldsymbol{\beta}_{<J}^m)$ . (Equation B13)

In Equation B13,  $\boldsymbol{\beta}_j$  is a vector containing the item discrimination parameter and the item difficulty parameters ( $\beta_{2j}$  and  $\beta_{\delta_j}$  in Equation B1, respectively), for the  $j^{th}$  item. Drawing  $\mathbf{X}_{miss}^m$ ,  $\boldsymbol{\theta}^m$ , and  $\boldsymbol{\beta}^m$  in separate steps or “blocks” in Equation B13, constitutes a “block-at-a-time” implementation of the Metropolis-Hastings algorithm (Chib and Greenberg 1995).

In Equation B13,  $p(\mathbf{X}_{miss} | \boldsymbol{\theta}^{m-1}, \boldsymbol{\beta}^{m-1})$  in step 1 can be directly calculated using Muraki’s formula as parameterized in Equation B1, so values for  $\mathbf{X}_{miss}^m$  can easily be generated. In particular, for the  $i^{th}$  respondent answering the  $j^{th}$  question, Muraki’s formula gives us  $P(X_{ij} =$

k) for every item category  $k$  associated with question  $j$ . The researcher simply draws a pseudo-random number  $r$  between 0 and 1 and assign  $X_{ij}$  as follows:

$$X_{ij} = \left\{ \begin{array}{l} 1 \text{ if } r < p(X_{ij} = 1) \\ k \text{ if } \sum_{m=1}^{k-1} p(X_{ij} = m) < r < \sum_{m=1}^k p(X_{ij} = m) \end{array} \right\}. \quad (\text{Equation B14})$$

Determining values for the ability parameters  $\theta_i^m$ , item discrimination parameters  $\beta_{2j}$ , and item difficulty parameters  $\beta_{\delta,j}$  is more difficult, because the joint posterior distribution for these parameters is difficult or impossible to compute. Accordingly, following Patz and Junker (Patz and Junker 1999b:157-158) this research uses a Metropolis-Hastings step to draw each of these parameters. For person  $i$  in a sample with  $N$  respondents,  $\theta_i^m$  is computed using the following procedure:

1. Following Patz and Junker (1999b), draw  $\theta_i^*$  from a normal proposal density

$$q_{\theta}(\theta_i^{m-1}, \theta_i) = n(\theta_i | \mu, \sigma^2) = n(\theta_i | \theta_i^{m-1}, c_{\theta}^2), \quad (\text{Equation B15})$$

where  $c_{\theta}^2$  is a tuning parameter that affects the speed of convergence of the Markov chain.

2. Accept  $\theta_i^*$  with probability

$$\begin{aligned} \alpha(\theta_i^{m-1}, \theta_i^*) &= \min \left\{ \frac{p(X | \beta^{m-1}, \theta_{<i>_i^*}^k, \theta_{>i^*}^k) p(\theta_i^*) q_{\theta}(\theta^*, \theta^{m-1})}{p(X | \beta^{m-1}, \theta^{m-1}) p(\theta_i^{m-1}) q_{\theta}(\theta^{m-1}, \theta^*)}, 1 \right\} \\ &= \min \left\{ \frac{p(X | \beta^{m-1}, \theta_1^m, \dots, \theta_{i-1}^m, \theta_i^*, \theta_{i+1}^{m-1}, \dots, \theta_N^{m-1}) p(\theta_i^*)}{p(X | \beta^{m-1}, \theta_1^m, \dots, \theta_{i-1}^m, \theta_i^{m-1}, \theta_{i+1}^{m-1}, \dots, \theta_N^{m-1}) p(\theta_i^{m-1})}, 1 \right\} \\ &= \min \left\{ \frac{p(X_i | \beta^{m-1}, \theta_i^*) p(\theta_i^*)}{p(X_i | \beta^{m-1}, \theta_i^{m-1}) p(\theta_i^{m-1})}, 1 \right\}, \end{aligned} \quad (\text{Equation B16})$$

with the term  $q_{\theta}$  dropping out of Equation B16 because the proposal distribution is symmetric.

(The choice of a prior for  $\theta$  is covered later in this appendix.) The rest of the simplification is

due to the assumptions that the individual ability scores are *a priori* independent, as well as conditionally independent given the item discrimination and difficulty parameters. As the present research involves sampling one  $\theta_i$  parameter at a time, following Gelman et al. (2004:307), I select  $c_\theta^2$  such that  $\theta_i^*$  is accepted about 40% of the time.

In Step 3 in Equation B13, Patz and Junker (1999b) draw the discrimination and item difficulty parameters in separate steps for each item, and recommend drawing the  $k^{th}$  item difficulty parameter for item  $j$  from a normal proposal density. Accordingly, for the  $m^{th}$  step in the Markov chain, each item difficulty parameter is drawn from a proposal density

$$q_{\beta_{\delta_{kj}}}(\beta_{\delta_{kj}}^{m-1}, \beta_{\delta_{kj}}) = n(\beta_{\delta_{kj}}|\mu, \sigma^2) = n(\beta_{\delta_{kj}}|\beta_{\delta_{kj}}^{m-1}, c_{\delta_{kj}}^2), \quad (\text{Equation B17})$$

with  $c_{\delta_{kj}}^2$  allowed to be different for every item parameter as a tool for optimizing convergence of the chain.<sup>179</sup> Next, the item discrimination parameters are drawn from the proposal density

$$q_{\beta_{2j}}(\beta_{2j}^{m-1}, \beta_{2j}) = \text{lognormal}(\beta_{2j}|\mu, \sigma^2) = \text{lognormal}(\beta_{2j}|\beta_{2j}^{m-1}, c_{2j}^2), \quad (\text{Equation B18})$$

using a lognormal proposal as the discrimination parameters must always be positive.

Next, for each item, each iteration of the Markov chain involves jointly accepting or rejecting the discrimination parameter and item difficulty parameters as a group with probability

$$\alpha_j(\beta_j^{m-1}, \beta_j^*) = \min \left\{ \frac{p(X_j|\theta^m, \beta_j^*)p(\beta_j^*)\text{lognormal}(\beta_{2j}^*|\beta_{2j}^{m-1}, c_{2j}^2)}{p(X_j|\theta^m, \beta_j^{m-1})p(\beta_j^{m-1})\text{lognormal}(\beta_{2j}^{m-1}|\beta_{2j}^*, c_{2j}^2)}, 1 \right\}, \quad (\text{Equation B19})$$

where the additional lognormal densities in Equation B19 are due to the asymmetry of the proposal distribution  $q_{\beta_{2j}}$ . Gelman (2004) recommends acceptance rates of about 20% for

---

<sup>179</sup> In fact, Patz and Junker (1999b) use the same variance for every parameter, but I revised their code to use separate variances after finding that the Metropolis chain was not converging quickly enough for the TYP data.

parameters sampled jointly in order to speed convergence of the Markov chain, so in general I seek values of  $c_{2j}^2$  and  $c_{\delta_{kj}}^2$  that lead to a rejection rate of between 75% and 80%. The difficulty parameters for each item in the survey, however, were often correlated, and as the Metropolis jumps in this multidimensional parameter space were calculated independently for each parameter, the Markov chain was slow to converge. This slow convergence is due to the fact that when values of  $c_{\delta_{kj}}^2$  and  $c_{2j}^2$  were too high, the Metropolis-Hastings algorithm often explored portions of the parameter space associated with a low value for  $p(X_j|\theta^m, \beta_j^*)$ , the probability of that particular item actually fitting the observed data, leading to high rejection rates. Conversely, when  $c_{\delta_{kj}}^2$  and  $c_{2j}^2$  were too low, autocorrelation among consecutive draws from the posterior distribution became a serious problem, such that more than 1,000 draws from the posterior were often necessary just to obtain one pair of independent draws. Following recommendations by Peter Hoff (personal communication) and in Patz and Junker (1999b), instead of drawing each of the  $K-1$  parameters  $\beta_{\delta_{kj}}^*$  separately for a particular item  $j$ , where  $K$  is the number of categories in item  $j$ , I revised Patz's code to draw all of the parameters  $\beta_{\delta_{kj}}^*$  and the log of  $\beta_{2j}^*$  jointly, using multivariate normal Metropolis-Hastings jumps. Such jumps require defining a variance-covariance matrix  $\Sigma_j$  for all parameters to be drawn for item  $j$ . The approach denoted by Equations B17-B19 then generated initial draws for the item parameters, with their corresponding observed variances and covariances. To increase or decrease the "distance" of these Metropolis-Hastings jumps, I simply multiplied all elements in  $\Sigma_j$  by a scalar, or "tuning parameter." During the burn-in period, an adaptive algorithm increases or decreases the tuning parameter in order to arrive at a target rejection percentage, while decreasing the magnitude of the changes in the tuning parameter through an iterative process similar to but simpler than simulated annealing (Kirkpatrick, Gelatt, and Vecchi 1983).

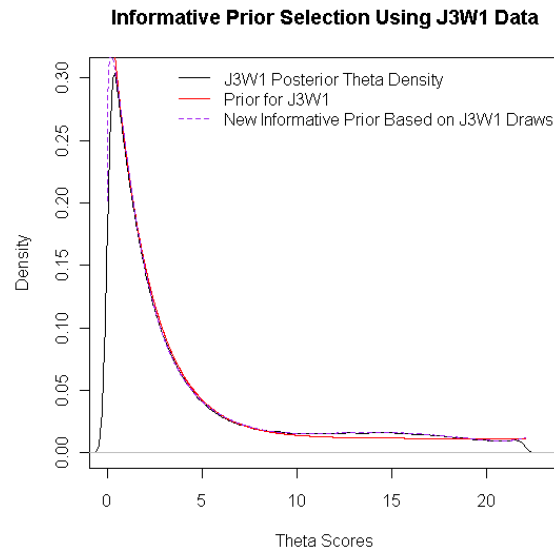
For the TYP data, even the above approach was slow to converge, so on top of the R function to identify a matrix  $\Sigma_j$  associated with a target rejection percentage, the current research implements a more complex adaptive algorithm that, for each item  $j$ , infers from the data a suitable target acceptance ratio  $\alpha_j$  that would lead to a speedy convergence of the Markov chain. In particular, beginning with a candidate acceptance ratio  $\alpha_j$ , the approach analyzes the autocorrelation among draws of each parameter from its posterior distribution, then determines the number of posterior draws necessary to obtain a pair of independent draws for each parameter and keeps track of the longest of these times, denoted here as  $t_{longest}$ . The procedure then identifies two target percentages,  $\alpha_j + \epsilon$  and  $\alpha_j - \epsilon$  (starting with  $\epsilon = .1$ ), runs the above algorithm to determine tuning parameters and variance-covariance matrices and  $\Sigma_j$  that yield those target percentages, and determines which of these three target percentages (including  $\alpha_j$  itself) is associated with the smallest value of  $t_{longest}$ . If that value of  $t_{longest}$  is associated with  $\alpha_j + \epsilon$  or  $\alpha_j - \epsilon$ , then the algorithm computes  $\alpha'_j = \alpha_j + \epsilon$  or  $\alpha'_j = \alpha_j - \epsilon$  as appropriate, then repeats the search procedure; if the smallest value of  $t_{longest}$  is instead associated with  $\alpha_j$ , then having found a local maximum, the algorithm decreases the value of  $\epsilon$  by 50% and repeats, stopping when the smallest value of  $t_{longest}$  is associated with  $\alpha_j$  and  $\epsilon = .025$  for every item  $j$ . The corresponding values of  $\alpha_j$  and  $\Sigma_j$  are then used to obtain draws of the item parameters from their posterior distribution after burn-in, and the complete procedure produces a fairly rapid convergence of the Markov chain (usually fewer than 500 posterior draws are needed to obtain pairs of independent draws).

### Determining Prior Distributions for Ability and Item Parameters

This section covers the details of obtaining an informative prior from Wave 1 of the J3 TYP cohort. Obtaining this distribution involves fitting the Bayesian GPCM to this wave, in turn specifying a prior distribution for individual abilities using findings from Western research into juvenile delinquency. Western research indicates that criminality roughly follows an overdispersed Poisson or negative binomial distribution, such that a few people commit a high number of crimes while most commit few or none. Accordingly, the prior for the J3 Wave 1 estimation is based on an exponential distribution to capture the fact that a high number of respondents appear to be answering that they commit zero crimes, and it is also mixed with a uniform distribution to capture a degree of uncertainty associated with the fact that the present sample is from a non-Western context. In particular, the mixture model incorporates an exponential density with rate parameter  $\lambda$  equal to 0.5 (implying an expected value for the exponential density corresponding to a raw score of 2), and a uniform prior. The range of the values for the prior helps to determine the scale for the ability scores, which in turn affects their interpretation; values for the uniform component of this prior range from 0 to the maximum observed raw score in a particular wave (22 in this case).<sup>180</sup> This mixture model – 75% exponential, 25% uniform – is shown by the solid red line in Figure B1.

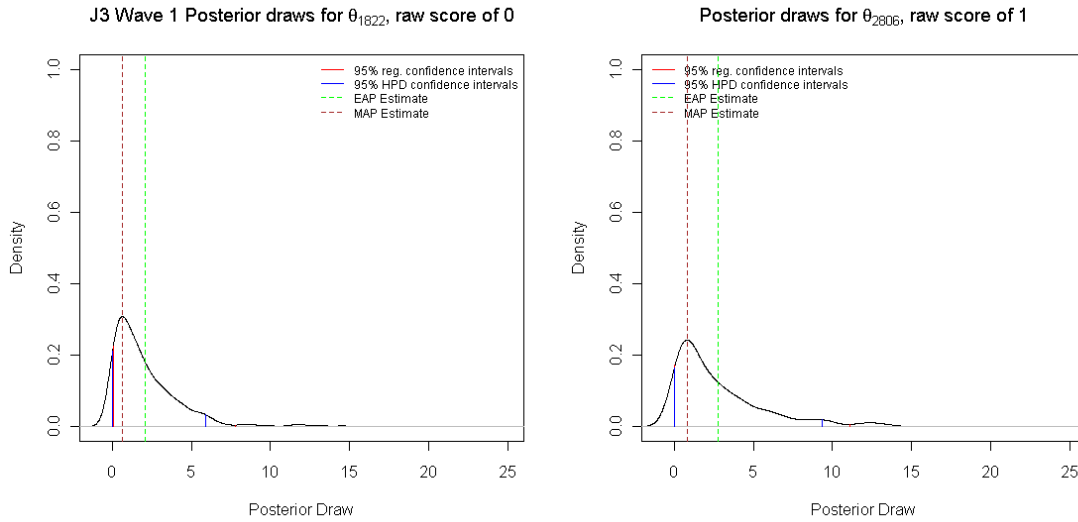
---

<sup>180</sup> Technically, an individual's estimated ability calculated with this prior could be greater than the maximum observed score, because for the exponential component of this density I used an exponential density rather than a truncated exponential distribution with a maximum value of 22. For all practical purposes, however, the maximum ability score obtainable using this prior is 22, because the probability of a random draw from an exponential density with rate parameter  $\frac{1}{2}$  being larger than 22 is only about 0.000017.

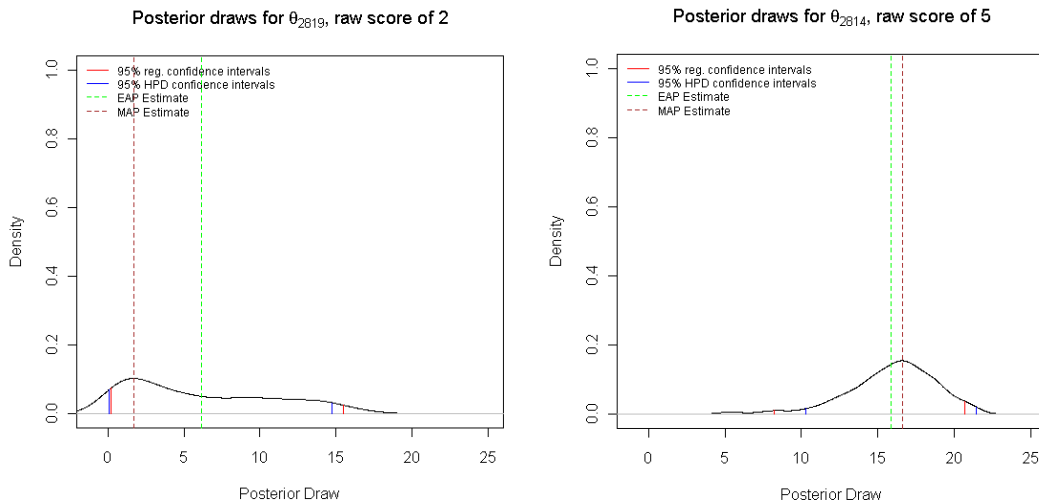


**Figure B1.** Derivation of informative priors to estimate latent constructs for criminality in the TYP data.

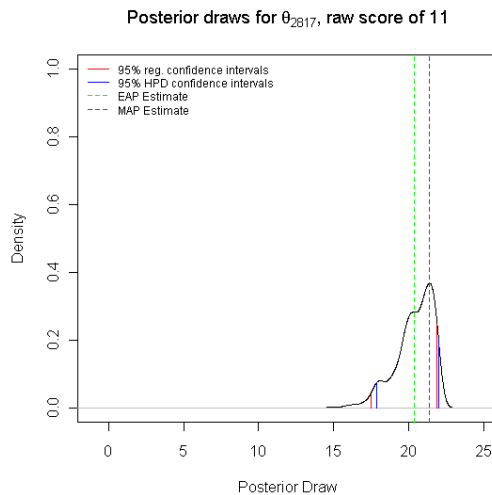
The black line in Figure B1 shows the distribution of Expected *A Posteriori* (EAP) estimates of individual abilities for the members of the J3 cohort responding to the Wave 1 survey. Its similarity to the prior distribution used to estimate ability scores belies the fact that the J3W1 responses – including zeroes – provide quite a bit of useful information to estimate ability scores. Figure B2 shows that knowing a respondent has a raw score of zero implies 95% confidence that his or her latent ability lies between 0 and about 6, compared to between 0 and roughly 19 for the prior; scores of 1 are associated with wider bands, between 0 and about 9. Looking to Figure 2.4, scores of 2 yield a posterior density similar to the flat prior, but scores of 5 and 11 provide enough information about deviance to rule out low ability scores associated with the exponential component of the prior distribution.



**Figure B2.** Distribution of posterior deviance estimates for respondent with raw scores of 0 and 1 in Wave 1 (J3 cohort). The “highest posterior density” (HPD) confidence intervals give the 95% confidence interval such that all observations within the interval are more likely to occur than those that are observed outside the interval; normally the blue lines would therefore be of equal height, but the left band is higher here due to the extreme skewness of the posterior densities.



**Figure B3.** Distribution of posterior deviance estimates for respondents with raw scores of 2 and 5 in Wave 1 (J3 cohort).



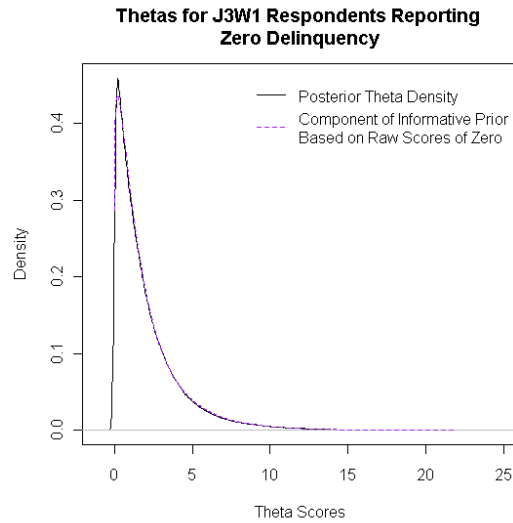
**Figure B4.** Distribution of posterior deviance estimates for a respondent with a raw score of 11 in Wave 1 (J3cohort).

Not every respondent reporting a raw score of 1 as in Figure B2, of course, will have an ability score 95% likely to be between 0 and 9; this interval may grow or shrink depending on the “difficulty” of the particular type of deviant action that the respondent reported doing. Generating hundreds of graphs such as those in Figures B2-B4 corresponding to all members of the J3W1 cohort, however, collectively demonstrated that individuals with the same raw score generally had posterior  $\theta$  densities with the same shape, and the densities of the expected a posteriori (EAP) values of all respondents with the same score therefore had a shape similar to the posterior density of theta values for each person in that group. To obtain a probability distribution that closely maps to the observed distribution of ability scores for the J3W1 survey, based on the shapes of the posterior densities the raw scores were assigned to five categories: 0, 1, 2-3, 4-6, and 7 and above, then the densities of observed EAP values for theta for each of those categories were plotted. Figures B5-B9 show these plots – each of which contains a probability density function in purple that is a mixture of two or more densities – along with a

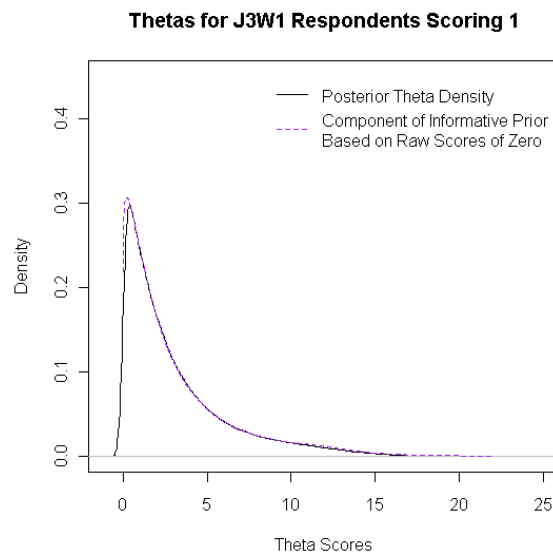
description of the components of these mixtures. The probability density function indicated by the dashed line in Figure B1 is itself a mixture of the densities in Figures B5-B9, where the weight associated with the contribution of each component to the total density is based on the proportion of each category of raw score observed in the sample (57.4%, 16.7%, 15.3%, 6.7%, and 3.7%, respectively). In obtaining these densities, only scores associated with respondents who did not leave any questionnaire items unanswered were used. Further, to restrict the ability scores to values between the minimum and maximum observed raw scores, these probability densities also use truncated normal, truncated exponential, and truncated lognormal distributions instead of their unbounded counterparts; the code for the truncated exponential and truncated lognormal distributions is derived from the truncated normal distribution in the `msm` package in R (Jackson 2010).<sup>181</sup> In practice in using probability densities this precise, the approach likely fitted a great deal of noise as well as signal to generate the composite informative prior, but the fitted noise is not particularly concerning as the goal is simply to maximize the information from the J3W1 survey used to generate the informative prior, regardless of how much noise is fit in the process.

---

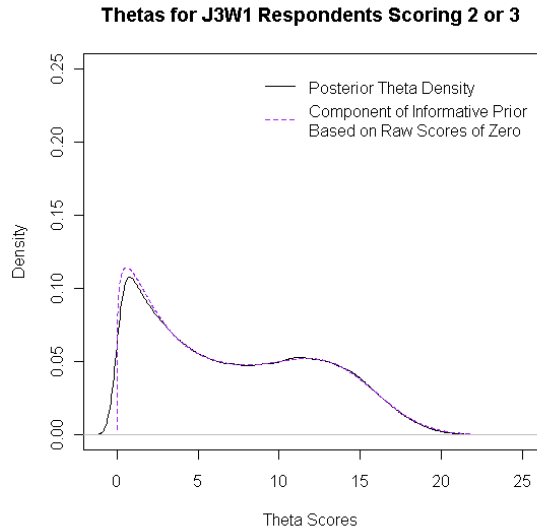
<sup>181</sup> A check on the coding confirmed that the area under the curve of each probability density function used here sums to exactly 1.



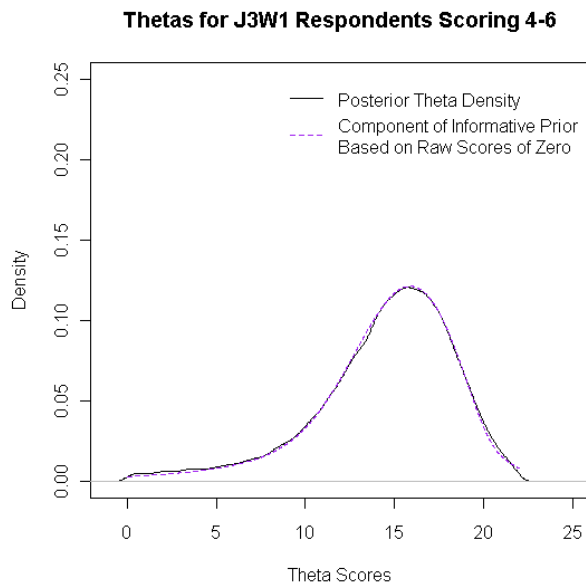
**Figure B5.** Distribution of EAP estimates for J3W1 survey respondents with raw scores of 0 (solid black line), along with a mixture of a truncated exponential distribution with a maximum value of 22 and rate parameter .41 (70% of the density), and a beta distribution with shape parameters 1.3 and 22 (30%).



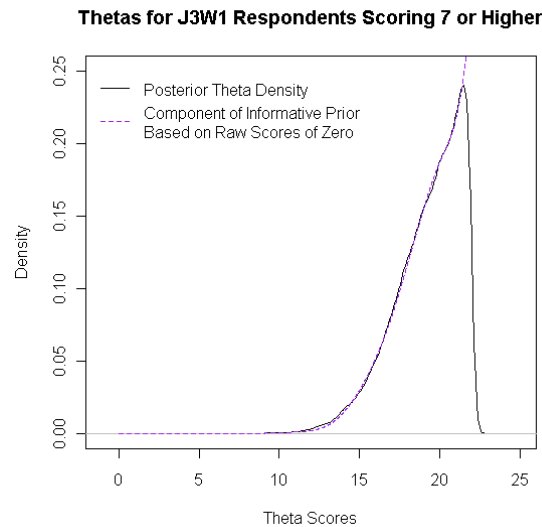
**Figure B6.** Distribution of EAP estimates for J3W1 survey respondents with raw scores of 1 (solid black line), along with a mixture of a truncated exponential distribution with a maximum value of 22 and rate parameter .28 (77.75% of the probability density), a beta distribution with shape parameters 1.3 and 19.5 (20%), and a truncated normal distribution with mean 11.3, standard deviation 2, and valid values between 0 and 22 (2.25%).



**Figure B7.** Distribution of EAP estimates for J3W1 survey respondents with raw scores of 2 and 3 (solid black line), along with a mixture of a beta distribution with shape parameters 1.2 and 9 (40% of the probability density), a truncated normal distribution with mean 12 and standard deviation 3.3 (38% of the density), a truncated normal distribution with mean 5.9 and standard deviation 2.9 (18% of the density), and a truncated normal distribution with mean 15 and standard deviation 2.2 (4% of the density). All truncated normal distributions are valid between 0 and 22.



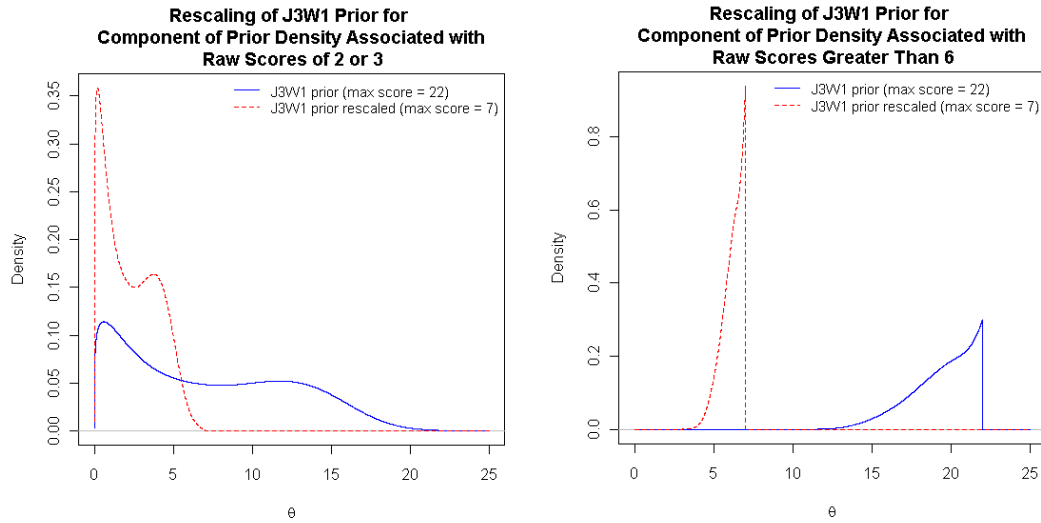
**Figure B8.** Distribution of EAP estimates for J3W1 survey respondents with raw scores between 4 and 6 (solid black line), along with a mixture of a truncated normal distribution with valid values between 0 and 22, a mean of 15.75, and standard deviation of 3 (54% of the probability density), and a “reflected” truncated lognormal distribution (46% of the density) equal to the maximum score (22) minus a truncated lognormal distribution with mean of  $\log(8)$  and standard deviation of  $\log(1.75)$  on a log scale, with a maximum value of 22. The resulting “reflected” lognormal density has an expected value of 14 and is left-censored at 0.



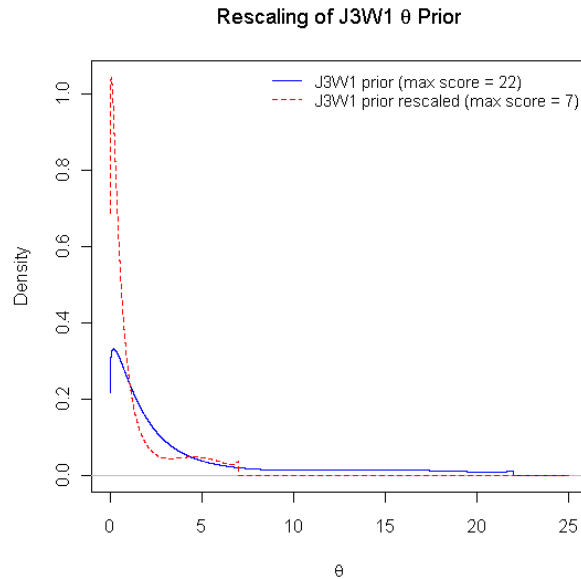
**Figure B9.** Distribution of EAP estimates for J3W1 survey respondents with raw scores of 7 or greater (solid black line). The component of the informative prior for this range of values, indicated by the purple dashed line, is a mixture of a truncated exponential distribution and four truncated normal distributions. The truncated exponential distribution has rate parameter 0.5 and valid values between 0 and 22, and has been “reflected” off the maximum score of 22 by subtracting its probability density function from that value (58% of the density); and the truncated normal distributions have valid values between 0 and 22 and have mean 17 and standard deviation 1.5 (16% of the density); mean 20 and standard deviation 1 (15% of the density); mean 18.5 and standard deviation 1 (10% of the density); and mean 14.5 and standard deviation 1 (1% of the density).

The scale for the ability scores is intended to facilitate their interpretation, with a one-unit increase in propensity for deviance roughly corresponding to an increase in one frequency category (e.g. from “once or twice” to “sometimes”) for a hypothetical “average” survey question and two categories differing in difficulty by an “average” amount. As the total observed raw score differs from one wave and cohort to the next, however, retaining this meaning for the ability scores requires the prior density to be rescaled for each wave in which it is used. Adjusting the informative prior distribution for use with each individual wave therefore involved rescaling the probability density function for each component of the mixture distributions by adjusting the bounds of the truncated distributions, multiplying the means and standard deviations of truncated normal distributions, multiplying the rate parameters of truncated

exponential distributions by a constant, and so on. Figure B10 gives two examples of such adjustments for a hypothetical wave with a maximum observed score of 7, and Figure B11 shows the end result.



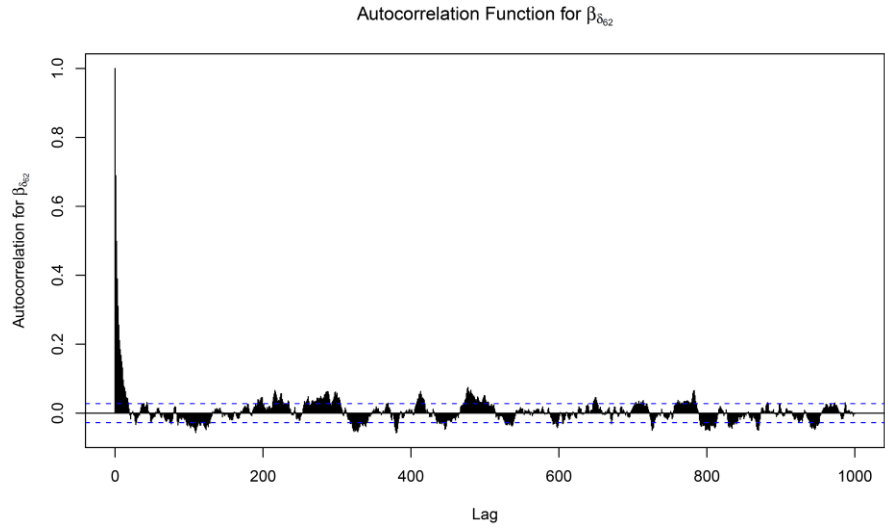
**Figure B10.** Example of rescaling the mixture densities in Figures B7 and B9 to suit a hypothetical wave with a maximum observed raw score of 7.



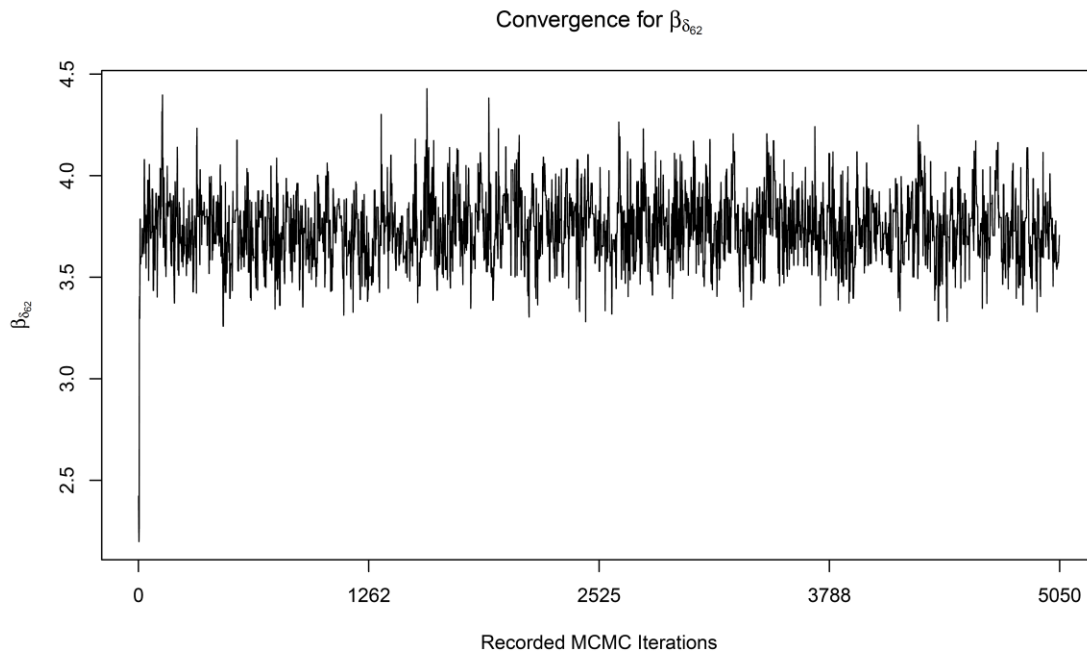
**Figure B11.** Example of rescaling J3W1 prior  $\theta$  density to suit a hypothetical wave with a maximum observed raw score of 7.

### Assessing Convergence of the Markov Chain

Plots of parameter values vs. the recorded number of MCMC chain iterations, along with plots of autocorrelation among the posterior draws, were used to determine convergence of the Markov chain. For Wave 7 of the J3 cohort, for example, after an initial run of the adaptive algorithm described previously to determine a variance-covariance matrix for multivariate normal Metropolis jumps for each set of item parameters, the longest time between independent draws from the posterior distribution for any parameter was 400 draws. The chain was then run again for 610,000 iterations, allowing 10,000 draws for the burn-in and saving every 100<sup>th</sup> draw. Plots similar to those in Figures B12 and B13, which show sample convergence and autocorrelation plots for an item difficulty parameter associated with Wave 2 of the J3 survey, confirmed that every fourth draw from the posterior distribution (after the initial burn-in) was independent. Accordingly, the chain generated 1,500 independent draws from the posterior distribution of the parameters. Repeating this procedure on a cluster of computers supported by the University of Washington's Center for Studies in Demography and Ecology (CSDE) generated a total of 4,500 draws from the posterior distribution. In this manner, with the aid of a cluster of eight computers and after more than 5,800 hours of computing time, I obtained at least 3,000 posterior draws for each parameter in all TYP survey waves, for both the J1 and J3 cohorts (see Table A9 in Appendix A).



**Figure B12.** Example diagnostic plot showing autocorrelation among posterior draws for an item difficulty parameter in Wave 2 of the J3 survey. Groups of draws that have peaks above or below the dotted lines represent draws that can be said to be correlated with 95% confidence; the fact that most draws fall within these two dotted lines indicates extremely low autocorrelation among the parameter draws.



**Figure B13.** Example diagnostic plot showing convergence for posterior draws for an item difficulty parameter in Wave 2 of the J3 survey. After an initial burn-in, subsequent draws of the parameter estimates both lie within a distinct interval and have high entropy, demonstrating low autocorrelation and thus convergence of the chain for that parameter.

## APPENDIX C

### Description of the Haebara and Stocking-Lord Methods for Vertical Scaling

This appendix presents details of the Haebara and Stocking-Lord procedures for vertical scaling. As described in Chapter 2, during vertical scaling a researcher must decide whether to compute the error in transforming an ability score from an “old” scale to the “new” scale, from the “new” scale to the “old” scale, or a “symmetric” implementation that does both. This dissertation uses a symmetric implementation as the present research may require rescaling an ability score in either direction. For each item, the Haebara method calculates the difference in the expected response to the item on the old scale and on the new scale, squares these differences, then sums them over all items (Equation C1). Following Kim and Kolen (2005),

$$Q = Q_1 + Q_2, \quad (\text{Equation C1a})$$

where

$$Q_1 = \frac{1}{S_1} \sum_{i=1}^{G_O} \left\{ \sum_{j=1}^J \sum_{k=1}^{K_j} [\hat{P}_{jkO}(\theta_{iO}) - \hat{P}_{jkN}^*(\theta_{iO})]^2 \right\} W_1(\theta_{iO}), \quad (\text{Equation C1b})$$

$$Q_2 = \frac{1}{S_2} \sum_{i=1}^{G_N} \left\{ \sum_{j=1}^J \sum_{k=1}^{K_j} [\hat{P}_{jkN}(\theta_{iN}) - \hat{P}_{jkO}^\#(\theta_{iN})]^2 \right\} W_2(\theta_{iN}), \quad (\text{Equation C1c})$$

and

$$S_1 = \sum_{j=1}^J K_j \left[ \sum_{i=1}^{G_O} W_1(\theta_{iO}) \right], \quad S_2 = \sum_{j=1}^J K_j \left[ \sum_{i=1}^{G_N} W_2(\theta_{iN}) \right]. \quad (\text{Equation C1d})$$

Equation C1 indicates a symmetric implementation of the Haebara method. In Equation C1,  $J$  is the number of test items used for the vertical scaling, each of which contains  $K_j$  categories, and

$\hat{P}_{jk}(\theta_i)$  denotes the probability that a hypothetical respondent with ability score  $\theta_i$  responds affirmatively to category  $k$  of item  $j$ , plugging into Muraki's formula for the Generalized Partial Credit Model (Equation 2.2) the estimated values for the item discrimination and difficulty parameters held invariant for the vertical scaling.  $\hat{P}_{jkO}(\theta_{iO})$  therefore refers to the estimated probability that a respondent with ability  $\theta_{iO}$  (as measured on the "old" scale) responds to category  $k$  of item  $j$ , plugging  $a_{jO}$  and  $\{\beta_{2jO}, \beta_{3jO}, \dots, \beta_{kjO}\}$  into Muraki's formula.  $\hat{P}_{jkN}^*(\theta_{iO})$ , in contrast, also refers to the probability that a respondent with ability  $\theta_{iO}$  (as measured on the "old" scale) responds to category  $k$  of item  $j$ , but transforms the item parameters on the new scale to the old scale in order to do so, using the equations

$$a_{jO} = \frac{a_{jN}}{A} \quad \text{(Equation C2)}$$

and

$$\beta_{vjO} = A\beta_{vjN} + B \quad \text{(Equation C3)}$$

in the case of the GPCM. Similarly, computation of  $\hat{P}_{jkO}^\#(\theta_{iN})$  involves converting the item parameters on the old scale to the new scale, using  $a_{jN} = Aa_{jO}$  and  $\beta_{jvN} = \frac{\beta_{jvO} - B}{A}$ .  $S_1$  and  $S_2$  in Equation C1 are standardization factors that in theory can be ignored when estimating the scaling parameters  $\hat{A}$  and  $\hat{B}$ , although the software package for vertical scaling used for the present research includes the factors as "the minimization algorithm used for nonlinear problems is affected by the magnitude of a criterion function due to its stopping rules" (Kim and Kolen 2003:14-15). Finally,  $\theta_{iO}$  and  $\theta_{iN}$  in Equations C1(a-c) denote a set of  $G_O$  and  $G_N$  "proficiency points" chosen at various locations along the ICC curve, each of which is assigned a weight

$W_1(\theta_{iO})$  for values of  $\theta_{iO}$ , or  $W_2(\theta_{iN})$  for  $\theta_{iN}$  (Kim and Kolen 2005). A distribution of proficiency points and their weights can be selected in any number of ways; one approach is to select a one that corresponds to the distribution of ability scores in a particular wave, but this research uses a set of 100 points that are equally weighted and spaced at equal intervals along the ICC curve between the minimum and maximum computed ability score, in order to emphasize a rescaling of ability scores for respondents of all abilities – including those with a relatively high propensity to offend – and without attempting to fit unobserved regions of the ICC curve.

The Stocking-Lord method is quite similar to the Haebara method, but it first sums the differences in expected responses over all items, then squares the sum:

$$F = F_1 + F_2, \quad (\text{Equation C4a})$$

where

$$F_1 = \frac{1}{S'_1} \sum_{i=1}^{G_O} \left[ \sum_{j=1}^J \sum_{k=1}^{K_j} U_{jk} \hat{P}_{jkO}(\theta_{iO}) - \sum_{j=1}^J \sum_{k=1}^{K_j} U_{jk} \hat{P}_{jkN}^*(\theta_{iO}) \right]^2 W_1(\theta_{iO}), \quad (\text{Equation C4b})$$

$$F_2 = \frac{1}{S'_2} \sum_{i=1}^{G_N} \left[ \sum_{j=1}^J \sum_{k=1}^{K_j} U_{jk} \hat{P}_{jkN}(\theta_{iN}) - \sum_{j=1}^J \sum_{k=1}^{K_j} U_{jk} \hat{P}_{jkO}^\#(\theta_{iN}) \right]^2 W_2(\theta_{iN}), \quad (\text{Equation C4c})$$

and

$$S'_1 = \sum_{i=1}^{G_O} W_1(\theta_{iO}), S'_2 = \sum_{i=1}^{G_N} W_2(\theta_{iN}). \quad (\text{Equation C4d})$$

The terms in Equation C4 are similar to that of the Haebara formula, except for the inclusion of  $U_{jk}$ , a scoring weight for category  $k$  of item  $j$ . This weight is necessary in order for the test characteristic curve to be defined (Kim and Kolen 2003), and is usually set to  $k$  or  $k-1$ ; the current work uses  $k-1$ , because it is the default setting in the POLYST software package.

## APPENDIX D

### **Design of a Filter to Create Data for Work Intensity, Work Extensity, and Whether Respondents Worked During the Previous Year, Academic Year, and Holiday Breaks**

This section describes the data filter used in this dissertation to determine whether respondents to Waves 8 and 9 worked during the year prior to the Wave 8 survey and during the year prior to the Wave 9 survey, as well as whether they worked during the academic year and summer. The filter also determines how intensively and for how many weeks they worked (see the right three columns of Table 4.2). The filter first merges responses to Wave 8 and Wave 9 into an array of up to five unique jobs that each respondent had held.<sup>182</sup> Next, as the timing of summer vacation varies for students in high school and college (generally it runs from July 1<sup>st</sup> through August 31<sup>st</sup> for high school students and from the second or third week in June through the second week in September for college students), a pre-processing step first determines whether respondents were in high school, college, or neither during Waves 8 and 9. In many cases, resolving discrepancies among the grades in which respondents stated they were enrolled in different years required consulting the name of the school. The TYP project team matched

---

<sup>182</sup> To identify duplicate jobs in Waves 8 and 9 and not count them twice when computing time worked during the year before the Wave 8 survey, I wrote some code to match jobs based on the name of the job, start dates, and end dates (if applicable). Respondents stated their job title for each job they listed in each wave, and the Taiwan Youth Project survey team then used a job classification schema developed by Taiwan's Department of Labor to assign a four-digit code to each job. The four-digit codes start with extremely general categories, and progressively assign more specific categories within each larger category, until the final digit is job-specific. To continue using an earlier example, a job such as selling snacks at a crowded night market during the Lunar New Year holiday while keeping passers-by from stealing the food might have been treated as "cashier" one year and "sales" the next; sometimes the same job might even be referred to as "security services – other." Matching jobs based on their belonging to the same general category of work made more sense than using the entire four-digit code. To determine exactly how many digits to use for the matching, while writing the data filter I selected a set of respondents' answers to use as test cases, and then examined the effect of using different matching criteria to minimize both true negatives and false positives. In practice, matching jobs based on their sharing just the first digit in these four-digit codes increased the number of matches that I found for the test cases without introducing false positives, so I used this somewhat broad criterion for matching. Allowing for a small degree of recall measurement error as well, I required the beginning dates for the two jobs being matched to be separated by at most two months, and required the ending dates to also fall within two months of each other (if the respondent had already left the job in Wave 8).

numeric codes from official government registries to school names, and those lists were used here.<sup>183</sup> A few coding assumptions were necessary concerning jobs that started or ended when the academic year was starting: first, if a job ended in September and a respondent was in college, the job is treated as not having overlapped the academic year. Similarly, if the job began in June and a respondent had been in college that year, the job is treated as not having overlapped the academic year that ended in June.

Assumptions concerning winter break were not as straightforward, as many jobs that a respondent may have held for several months would normally take off during the break (which runs for about three weeks between mid-January and mid-February), while other jobs taken up by youth are often temporary, seasonal jobs designed around the holiday schedule. Accordingly, the filter identifies respondents who held a job during the January or February corresponding to the break and who started the job prior to January or ended it after February, then codes the amount of time these respondents spent working during summer and winter breaks in two ways. The first approach treats them as not having worked through the break, such that the resulting measure of having worked during the summer or winter corresponded to anyone who either worked during the summer months or who appeared to intentionally take a temporary job during that January or February. The second approach codes respondents as having worked through the break, to create a less conservative estimate of the number of respondents who worked during the summer or winter.<sup>184</sup>

---

<sup>183</sup> Reproducing research methodologies at this level of detail is what makes replication difficult, and indeed nearly impossible. I therefore plan to make all my source code available to the TYP office in the Institute of Sociology at Academia Sinica in order to aid future efforts to replicate these results in Taiwan, Mainland China, or elsewhere. Researchers may have specific questions about how I handled various types of missingness patterns to infer a respondent's current grade in school, type of school (general track or vocational), and so on. I encourage researchers seeking to reproduce this methodology to refer to my comments in the SPSS syntax files, which are more detailed than this filter description.

<sup>184</sup> To reduce the amount of time spent writing the data filter, I used the first, more conservative measure of work during the summer or winter in fixed effects models examining the effect of summer work on deviance that covered

Generally, when determining how many weeks a respondent worked, the filter uses an assumption that he or she started the job halfway through the month and ended it halfway through the month. For respondents who started and ended a job during the same month, the filter counts the job as lasting for two weeks, and if they started and ended a job in either June or September (and their summer break extended into that month), the job is coded as having lasted for two weeks and having taken place during the summer break. Similarly, jobs that started and ended in January are coded as having taken place during the last two weeks in January, and jobs that started and ended in February are coded as having taken place during the first two weeks in February.

The filter was designed to reduce the number of respondents to the Wave 8 survey who stated that they had held two jobs, one of which began after September 2006 and one of which ended prior to July 2006 (and hence for whom whether they had worked at least during the previous summer, and possibly also during the academic year, was unclear). It fulfilled its purpose, reducing this number from 498 to just nine respondents for whom missing values prevented a determination of exactly when they had worked, and a few dozen additional respondents missing from the Wave 9 survey and for whom answering just two questions about their work history in Wave 8 may have provided incomplete information for the 12-month period between June 2006 and June 2007.<sup>185</sup> This approach may, however, overestimate the number of

---

both high school and college years. A few respondents to the Wave 8 and Wave 9 surveys who only worked during the winter and not the summer may therefore have been introduced, but this number was no more than 42 respondents out of 648 coded as working during the summer before Wave 8 using the more conservative measure, and that is assuming that none of these 42 respondents also worked during the summer. Accordingly, especially as my focus was on the academic year, I did not take extra time to produce a new variable for Waves 8 and 9 unique to the summer for this single model.

<sup>185</sup> Readers might find a little further detail useful in assessing exactly how effective the filter was. In writing the filter, I first flagged 498 respondents to the Wave 8 survey who listed their primary job as having started after the beginning of their academic year in 2006, and thus who may also have held a job during the previous summer (but the data had not yet resolved that question). After I included the second job detailed in the Wave 8 survey – for which the question asked respondents to state the first job they ever had held – I unflagged all of the 498

hours that respondents worked, or the number of weeks that they worked per year, if the algorithm used to identify duplicate jobs referenced in both waves 8 and 9 had a high true negative rate.

Once the filter produces an array containing the starting and ending times for each unique job, it computes the total number of weeks worked during the academic year, year, and the set of weeks comprised of the summer and winter breaks joined together. For time periods during which a respondent may have held more than one job, the filter counts the weeks only once but sums the number of hours worked per week. This approach also serves as a rough way to identify respondents who had really referred to the same job in Waves 8 and 9 but who the filter had not flagged as having described the same job twice, as the filter could potentially produce excessive work intensity values (e.g. 120 hours per week). Of course, some large estimates of work intensity may simply be due to respondents who exaggerated their working hours, or who perhaps viewed being on call as the same as working. After computing work intensity values for each relevant period, the filter truncates work intensity at a maximum value of 100 hours per week; this truncation only affected about half a dozen cases, providing some evidence that the true negative rate probably was not too large.

---

respondents who either did not list a second job (and hence presumably had never had another job), or who listed a job that started prior to September 2006. Only 17 remaining respondents had started both jobs after the beginning of the previous academic year, and I coded them as not having worked during the previous summer. While I did not specifically flag the number of respondents who answered questions about the second job and ended prior to the beginning of their summer break (as that would have required extra coding based on the type of school respondents were attending), I know that out of 615 respondents who answered questions about the second job and were not flagged for manual coding due to missingness in the responses, 276 ended the job prior to the start of 2006. A number of these respondents would have been missing from Wave 9, and as 11.96% of respondents to the J1 Wave 8 survey were missing from the J1 Wave 9 survey, my best guess is that for a few more than  $(498 * 276/615 * .1196) = 26.72$  respondents – so about 30 to 35 people – I have to assume that they did not hold a third job that took place during Summer 2006. For the remaining respondents, I can use their responses to Wave 9 to gain a clearer picture of whether they did additional work during Summer 2006 and the 2006-2007 academic year and winter break that was not captured in the Wave 8 survey. In other words, the filter helped to retrieve additional information about roughly 450 people, and reduced the level of uncertainty about when and how long respondents actually worked to a few dozen instead of several hundred cases.

One last detail is worth mention: the variables covering whether respondents worked during the previous year covers July 2006-June 2007 (for Wave 8) and July 2008-June 2009 (for Wave 9). The portion of the variable covering whether respondents worked during the summer and winter breaks corresponding to the summer, however, occasionally covers mid-June. As high schools do not start their summer breaks until about July 1<sup>st</sup>, the filter codes students who were in high school or freshmen in college in waves 8 and 9 as not starting their breaks until July 1<sup>st</sup>, and one associated assumption is that all students who were freshmen in college were in high school immediately beforehand. In any case, it is not obvious that anyone in the labor force or in military service during the year prior to their freshman year in college had enjoyed a traditional summer break, yet the questions about the timing of work make that assumption; as measurement error is associated with these respondents anyway, this particular coding strategy probably does not change the situation very much. For students who were sophomores in college or later, the filter codes their summer breaks as starting in mid-June – meaning that for these students, the time frames corresponding to working during the previous 12 months and the sum of the academic year and summer and winter breaks are not equivalent; the first corresponds to July-June, and the second to mid-June through mid-June, or July through mid-June. The overall goal for this coding was to allow the summer and winter break questions to be comparable to those asked of respondents to the survey in waves 4 and 5.

## APPENDIX E

### **Time-Varying Covariates Used to Model Selection into Work<sup>186</sup>**

Tables 3.1 and 4.1 list the variables in the TYP youth survey for modeling selection into work and deviance in Chapter 4. I include these covariates in propensity score models for selection into work for a number of reasons. Household income is included as it no doubt affects selection into part-time work and into cram school; “Adverse SES” captures whether respondents are satisfied with their family’s financial situation (Waves 3 and 6) and whether they worry about their family’s economic situation (Wave 5), and is a measure of strain. The question is not asked of respondents in college, and as the value of this variable may well be affected by whether students have been working part-time, I do not lag the Wave 6 variable in models for selection into work in Waves 8 and 9. Strain at school is measured in terms of overall concern about grades and concern about teachers. Yi, Wu, Chang and Chang (2009) examined the effect of perceived unfair treatment by teachers, but the particular questions asked in the J1W4 survey are not robust enough to create a strong variable for analysis. School commitment is captured by time spent on homework in middle school and college, and by class rank in middle and high school; lagged commitment is unavailable for Wave 5 because students in Wave 4 were not asked about homework and were interviewed at the beginning of the academic year, when class rankings were unavailable. Given the sizeable transition between middle and high school, interpolating time spent on homework or class rankings using data from middle school would

---

<sup>186</sup> The writing in this appendix is particularly rough, as it essentially consists of the author’s research notes while filtering the TYP data, making extensive use of the first person as with a diary. The level of detail in this appendix is not necessary for most readings of the present dissertation, and is meant to assist researchers who may seek to replicate these findings in the future. As the number of readers who therefore would give this appendix a close read can probably be counted on one hand, the author has not given the text much editing and appreciates the reader’s understanding.

also be inappropriate, so school commitment is not used for modeling selection into work in Wave 5. While family cohesion is not a strong predictor of delinquency net of family intactness to my knowledge, it has been linked in Taiwan to moral beliefs (Wu and Jou 2009), which is not included in the Taiwan Youth Survey often enough to be a time-varying covariate; as moral beliefs themselves are a consequence of deviance (Matsueda 1989), it would be inappropriate for me to treat them as time-invariant, and thus my next best strategy is to include family cohesion in my models. Class cohesion is measured differently in J1W3 vs. in college; the college variable is a measure of attachment to one's classmates ("Are you satisfied with your relationship with your classmates?") while the J1W3 variable more closely captures cooperation within the classroom instead of competition.

The "strain from family" indicator has to do with questions such as "Do you feel like you'll never be able to live up to your parents' expectations about your grades?" on the J1W3 survey; questions of this nature are only asked intermittently, and I decide not to directly include them as strain due to school and parents' expectations for their child's educational attainment also capture this concept. Additionally, because attachment to other family members such as grandparents and siblings is not measured as regularly as attachment to parents, and because attachment to parents has a more direct effect on crime, I leave out attachment to grandparents (although it could be used as an additional covariate in the models for college students, as a sensitivity check). Strain due to school (a perceived inability to keep up with schoolwork) is only available for students' middle and high school years; in college, only grades and time spent doing homework are available, so I use those covariates only (both measures of school commitment). Following Sampson and Laub's use of the Glueck data (1993), I combine harsh discipline and erratic discipline into one scale capturing harsh and erratic discipline; discipline

questions are not asked in Wave 8, so for that wave I lag the Wave 6 response. Whether respondents have an allowance was not included in any studies, but presumably it could affect selection into work during high school, as most Taiwanese youth seek employment in order to buy things; while the question is not asked in college, presumably for most families students in college stop receiving an “allowance” in the traditional meaning of the word. In college, however, having student loans could push students into working, as much as having an allowance allows students in high school to not have to work. Concerning self-control, the survey only contains one or two questions in J1 Wave 3 that come close to measuring self-control: “Since starting junior high school, have the following happened? 1) Unable to concentrate or focus my attention for a long period of time; 2) Fell asleep in class.” The second condition could be due to working long hours, and so really only the first question has to do with self-control, and I feel squeamish about treating a variable measured by a single item as a time-invariant covariate predicting deviance. Instead, Gottfredson and Hirschi (1990) theorize that minor deviance is a great way to measure self-control, so I simply include lagged deviance in the selection models as a control. Depression is included because it is correlated with deviance, and because it could also lead students to give up school and work instead (although no literature in Taiwan supports or refutes this argument). Experiencing strain due to working part-time is sadly only observed during one wave, but is highly likely to affect part-time work during a subsequent wave, and could affect delinquency as well; unfortunately it is only asked of Wave 5 respondents working full-time (only about 6.6% of the sample), so I omit the variable from the selection model. I do not include beliefs at all, as they would be lagged by two years in the single wave in which I could use them, and beliefs are prone to change based on deviance. Social disorganization, observed only twice (in Waves 3 and 6), should ideally be interpolated to obtain

values for Waves 4 and 5, but as some respondents doubtless move between Waves 3 and 6, interpolation is risky, and instead I use lagged social disorganization, and assume that disorganization is somewhat stable over time. Finally, materialism could potentially be treated as time-invariant, but as it is not observed until Wave 8, I choose not to include it in models of selection into work, lest an interest in keeping up with the latest fashion trends also be a consequence of working for money to buy things. I do keep materialism in the model for selection into Wave 9 work.

#### *Additional Details of the TYP Covariates Relevant to Data Cleaning*

Mother's education and father's education is a time-invariant variable based on Wave 1 (if available) and on later waves (if not). The category "elementary school" in Waves 1 and 2 is coded as "elementary school or less" in Waves 3 and 6, so in a few cases (10 cases for mothers and 15 for fathers), by treating the Wave 3 or Wave 6 version of the variable as the same as the version used in Waves 1 and 2, I may be overestimating a parent's educational level. The variable in Wave 9 contained different categories for "elementary school" and "less than elementary school," but since only about half of respondents to the Wave 1 survey also answered the Wave 9 survey, I just use the Wave 1 survey for all respondents. The Wave 9 survey does suggest that about 2/3 of mothers who had an elementary school education or less in fact completed elementary school, so this inconsistency between the category wording in Waves 1 and 2 compared to Waves 3 and 6 should really only affect around three cases for mothers and five cases for fathers. I report this information here to aid anyone seeking to replicate the work.

Household income is unavailable in Waves 4, 5, and 7, so I obtain estimates for income during these years by conducting a linear interpolation of parental income in adjacent waves. In

doing so, I place priority on using responses from the parent surveys, then place priority on using waves proximate in time to the wave for which I am interpolating values. Household income is reported by both the parent and child in Waves 3 and 8, so for these waves I use the child's response if the corresponding response to the parent survey is missing. In carrying out the linear interpolation, I also account for the fact that the survey waves are spaced apart by intervals of unequal length. Finally, I only interpolate values for Waves 4, 5, and 7; if household income is missing for Waves 3, 6, or 8, I impute the missing values instead of interpolating them using neighboring waves.

School attachment is calculated as follows: for Wave 3, it is a scale created from the following questions, after each is standardized:

1. "Do you like your school?"
2. "Do you feel a part of your school?"
3. "Does your school bring you honor?"
4. "How many of your teachers know you?"
5. "How many of your teachers give you a good impression?"
6. "How many of the teachers at your school are worthy of respect?"
7. "Do you think your head teacher cares about your class?"
8. "Do you like your teachers?"
9. "Are you satisfied with your relationship with your teachers?"

The scale has a Cronbach's alpha reliability of .757. As most of the respondents to Wave 4 completed a shorter telephone survey, the only real question about school attachment asks if respondents are satisfied with their current school. Unfortunately the question is asked immediately after respondents state their scores on the high school entrance examination and are asked if their parents and siblings like the school they are in, so youth respondents are likely primed to assess their educational attainment, rather than their degree of emotional attachment to the school. As a further indication of how problematic this question is with respect to measuring

school attachment, Wave 5 respondents are also asked the question, but only if they have delayed their education and are entering 10<sup>th</sup> grade at the time of the survey. Accordingly, for Wave 5 I use “How satisfied are you with your head teacher?” as a measure for school attachment. These two questions do not necessarily measure the same thing, as indicated by the fact that when in Wave 7 respondents are asked both how satisfied they are with their college and how satisfied they are with their teachers, a scale comprised of both questions has an alpha reliability of just 0.464. Due to this low reliability and the fact that satisfaction with the school is again asked in association with examination scores, I use attachment to teachers as my measure for school attachment in Wave 7. In Wave 8, school attachment is a scale, with an alpha reliability of .762, constructed from “How satisfied are you with the content of the courses you have taken?” and “How satisfied are you with your teachers in college?”

One disadvantage of the TYP survey design is that respondents in Waves 7 and 8 who were either still in high school or in a five-year *wuzhuan* (junior college) did not answer the survey questions related to school attachment and commitment. Accordingly, I carry out two analyses for students of college age: one that investigates the effect of work on delinquency for youth in college, and one that investigates the effect of work on delinquency for youth of college age, but which does not control for school attachment and commitment for youth in school (including youth in college).

For school commitment in Wave 3, I use four questions:

1. School grades as represented by ranking in one’s junior high school classroom
2. “Since starting ninth grade, do you think have you studied hard?” (This is a yes/no question.)
3. “Since starting ninth grade, aside from attending classes, how many hours per day on average do you spend studying, attending cram school, and doing homework?”
4. “Since starting ninth grade, on the weekends and on holidays, how many hours per day do you spend on average studying and doing homework?”

Following Xie and Zhou (2013), the present research assigns responses to the first question to four levels: “superior” (grades within the top five students in the respondent’s classroom), “above average” (the top six to 10 students), “average” (the top 11-20 students), and “below average” (ranked 21<sup>st</sup> or lower). (A question on the Wave 4 survey indicates that the average 10<sup>th</sup> grade class size was 41.74 and only 1.6% of classes had fewer than 21 students, so this coding strategy is robust to the fact that class sizes differ.) A principal components analysis indicates that all four of these questions load onto a single latent factor for school commitment, with the fourth question loading the highest. I therefore construct a single scale with these four items (after standardizing them), and the scale has a Cronbach’s alpha reliability of .669. (The reliability would be higher were I to only use homework hours, but I leave the other questions in as they are the only available ones in Waves 4 and 5.) Wave 4 respondents are only asked if they studied hard, so that becomes my measure of school commitment. Wave 5 respondents are again asked to state if they studied hard, and give their class rank as well as how many students are in their class. Using class size to transform class rank into a percentile increases the correlation of class rank with studying hard from .245 to .259 while only losing four cases to missingness, so I perform the transformation. Including class rank as a percentile in a scale with the dichotomous question about whether students studied hard produces an alpha reliability of only .41. The question about homework is a less precise but more accurate measure of school commitment than class ranking: it is less precise because it is a dichotomous variable to which about 2/3 of respondents answer “yes,” but it is more theoretically valid because it demonstrates a strong commitment to work hard to get ahead, while some students with a high class ranking could simply be gifted and not work hard for the high grades. As one measure is continuous and the

other discrete, constructing a single scale from the two could be problematic, so I opt to include both measures of school commitment in the propensity score model separately as I am more interested in controlling for school commitment than in interpreting the effect on propensity to work of school commitment as a single theoretical construct in the selection model. For Wave 7, school commitment is the number of hours per week respondents spend preparing for class, and the Wave 8 survey contains the same question as well as an additional question about how many hours per week respondents spend studying during examination periods. To reduce measurement error, I construct a measure for Wave 8 school commitment from the average of the standardized values for both questions, and the resulting scale has an alpha reliability of .76.

Parental attachment is rather straightforward. In Waves 3, 4, 7, and 8, respondents are asked if they are satisfied with their relationship with their mother and if they are satisfied with their relationship with their father, giving their responses on a Likert scale. In Wave 5, they are asked the following questions, also measured using Likert scales:

1. "Does your mother care about your feelings or problems?"
2. "Does your father care about your feelings or problems?"
3. "Does your mother help you when you run into trouble in your daily life?"
4. "Does your father help you when you run into trouble in your daily life?"
5. "Does your mother understand you?"
6. "Does your father understand you?"

I create unstandardized scales from the questions for each wave. All have high reliability (greater than 0.7), and the scale for Wave 5 is particularly high with an alpha reliability of 0.83.

Concerning parental monitoring, the youth surveys for Waves 3, 4, 6, and 8 each contain two questions – one for the mother and one for the father – asking if the parent knows all the places where the respondent goes each day. Scales constructed from these variables all have alpha reliability scores greater than 0.7. Wave 5 contains the following six questions:

1. "Does your mother know where you go every day?"
2. "Does your father know where you go every day?"
3. "Does your mother know who you are with when you are away from home?"
4. "Does your father know who you are with when you are away from home?"
5. "Does your mother keep after you all day, asking where you are and what you are doing?"
6. "Does your father keep after you all day, asking where you are and what you are doing?"

The last two questions are highly correlated with each other but not to the rest of the questions; a scale constructed with all six questions has an alpha reliability of .705, which rises to .795 with these two questions removed. Hirschi (1969:88) theorizes that the importance of parental monitoring lies in the parent being psychologically present at the moment that a child is presented with the opportunity to commit a crime. The last two questions have much more to do with whether the parent physically communicates with the child than whether he or she is psychologically present, so I only use the first four to construct a scale for Wave 5. Finally, parental monitoring is not asked in Wave 7, so I have the option of lagging parental monitoring for an additional year, or using monitoring in Waves 6 and 8 to interpolate parental monitoring in Wave 7. I choose to lag Wave 6 monitoring, as Wave 8 parental monitoring may be affected by work during Wave 7. This approach may slightly overestimate the effect of parental monitoring in college, but at least it won't be polluted by college work.

Family intactness, in the Western literature, is negatively associated with deviance and is usually coded as the youth respondent living with both biological parents. In the case of the Taiwan Youth Project survey, only respondents to the Wave 3 survey are asked if they are living with both biological parents, while Waves 4, 5, and 6 simply ask if the respondent is living with both parents, Wave 7 asks if the respondent is living at home, and Wave 8 asks if the respondent is living at home and if he or she is living with each parent in three separate questions. I code "living with parents" as living with both biological parents for Wave 3, living with both parents

for Waves 4 and 5. For Wave 8, I derive living at home with both biological parents from the three separate questions. Finally, for Wave 7, if the respondent is living at home that year, then if the respondent was living with both parents in Wave 6 (or in Wave 5 if they did not answer the Wave 6 question), I assume they were still living at home with both parents the following year. For 42 respondents who were living at home in Wave 7 but did not respond to either the Wave 5 or Wave 6 survey questions (2.4% of the Wave 7 sample), I simply assumed they were living at home with both parents in Wave 7 (as 72% of the respondents to the Wave 6 survey were living with both parents). I coded the remaining Wave 7 respondents who were not living at home as perforce not living at home with both biological parents that year.

Peer deviance is measured by two questions in Wave 3: “Since starting ninth grade, how often has this person who you describe as your best friend among your classmates skipped class?” and “Since starting ninth grade, has this person...caused problems in class (such as arguing, fighting, or otherwise causing trouble)?” Treating two respondents with no friends as the same as respondents who have non-delinquent friends<sup>187</sup>, I create a scale from the two questions. In Wave s 4 and 5, only one variable for peer deviance is available: a question concerning whether any of respondents’ friends or classmates have used drugs. Only about 12% in Wave 4 and 15% in Wave 5 answer yes, with about 300 Wave 4 and 150 Wave 5 respondents stating that they do not know. Those responses I code as missing, then impute. As peer deviance is not observed in Wave 7, I use lagged Wave 6 responses. Wave 6 respondents are asked to describe how often their three best friends smoke, drink, run away from home, steal, fight, and use drugs, on frequency scales ranging from “Never” (1) to “All the time” (5). I standardize each measure, then create a scale from these standardized variables with an alpha reliability of .728.

---

<sup>187</sup> In general treating zeroes with different meanings as the same can be problematic, but in the present case, two cases would not bias the results very much at all.

In Wave 8, peer deviance is a composite measure of 10 questions, asking respondents how often in the past year respondents' good friends were involved in any of the following types of deviance: vandalism, theft, cheating or fraud, gambling (playing Majiang), drag racing, viewing pornographic CDs or magazines, illegal drug use, chewing betelnut, smoking, and drinking. The alpha reliability for this measure is .616, but increases to .668 if I standardize each question before creating a scale, so I do the latter. Doing so reduces the interpretability of the peer deviance measure, but the corresponding reduction in measurement error may make a difference.

For prior deviance I use the EAP estimates for deviance derived in Chapter 2, except for deviance in J1 Wave 3 (which I did not include in the Chapter 2 analysis). For that wave, I use a scale based on the following questions:

“Thinking about ninth grade, how often have you done the following?”

1. “...Not followed school rules”
2. “...Skipped class”
3. “...Smoked, drank, or used drugs”
4. “...Cheated on a test”
5. “...Caused trouble at school (such as causing a disturbance in class, arguing, or fighting)”

Possible responses are “never” (0), “occasionally” (1), “sometimes” (2), or “often” (3). I use a scale based on the sum of the unstandardized questions with a Cronbach's alpha reliability of .772. An increase of 1 on this scale therefore corresponds to moving up one category for one of these questions. The variables are exceptionally clean, as only 10 respondents were missing on any of the items, and nine of them were only missing on one item. For these nine respondents I weighted the sum for the remaining four items by 5/4 to allow comparability, and left the scale value missing for the final respondent.

Peer attachment questions are only included in Wave 3 and 6. I could interpolate to derive values for Waves 4 and 5, but for the fact that work during Waves 4 and 5 could have an effect

on peer attachment in Wave 6. Accordingly, I use peer attachment to model selection into Wave 4 work, and peer attachment lagged by two years to model selection into Wave 5 work. Waves 3 and 6, respondents are asked how much the following statements describe their friends:

- “1. They care about me.”
- “2. They often help me with things.”
- “3. They comfort me when I run into difficulties.”

The Wave 3 and Wave 6 scales have alpha reliability scores of 0.88 and 0.92, respectively.

Each wave, respondents are asked if they have experienced a number of stressful life events of differing severity, and are asked different questions each year. Determining weights for these events would constitute a research project in its own right, and following other research with these data (e.g. Lin and Li 2013) I simply convert the responses to binary variables and sum them to create scales for each survey wave. The questions are as follows:

Wave 3:

- “Thinking about ninth grade, have the following happened?”
- “1. I had a falling out with a good friend.”
- “2. My grades and class ranking significantly dropped.”
- “3. I started wearing glasses or braces.”
- “4. My family started to have less money.”
- “5. My parents argued more and more.”
- “6. My father or mother spent more and more time out of the house.”
- “7. I had an ongoing conflict with one of my classmates.”
- “8. I developed a serious illness or sustained a serious injury.”
- “9. I changed schools.”
- “10. I was not allowed to participate in an important school activity (sports, music, club, etc.)”
- “11. I was intentionally bullied<sup>188</sup> or hurt by a classmate.”
- “12. My father or mother became unemployed.”
- “13. I fought with my parents more and more.”
- “14. A close relative passed away.”

---

<sup>188</sup> The Chinese expression for “bullying” (勒索) has a slightly different meaning from the English and literally translates to “extortion.” The idea is that rather than simply being put down, the recipient of the bullying is forced to perform some favor for or give some tangible property to the person doing the bullying.

Wave 4:

“In the last six months (since March or April), have the following happened?”

“1. Did you develop a serious illness or sustain a serious injury?”

“2. Were you intentionally bullied or hurt by a classmate?”

Wave 5:

“In the last year (since last November), have the following happened?”

“1. Did your family’s economic status change?”<sup>189</sup>

“2. Did you develop a serious illness or sustain a serious injury?”

“3. Were you intentionally bullied or hurt by a classmate?”

“4. Did any of your friends or classmates pass away?”

“5. Did you move?”

Wave 6:

“In the last year, have the following happened?”

“1. I had a falling out with a good friend.”

“2. My parents argued more and more.”

“3. My father or mother spent more and more time out of the house.”

“4. My mother or father had a violent interaction with another family member.”

“5. I had an ongoing conflict with one of my classmates / coworkers.”

“6. I developed a serious illness or sustained a serious injury.”

“7. I changed schools / changed jobs.”

“8. I lost a job.”

“9. I stayed away from home.”

“10. My mother or father became unemployed.”

“11. My mother or father passed away.”

“12. A close relative passed away.”

“13. My pet died.”<sup>190</sup>

Wave 8:

“In the past year, have the following happened?”

“1. I had a broken heart.”

“2. I / my girlfriend accidentally got pregnant.”

“3. I lost my job or quit school.”

“4. I had a falling out with a good friend.”

“5. My parents argued with each other more and more.”

“6. My father or mother spent more and more time out of the house.”

“7. My mother or father became unemployed.”

“8. I did not get along well with a coworker / classmate.”

---

<sup>189</sup> To keep the structure of the question the same as that of the other stressful life events questions, I coded “substantially improved,” “improved a little,” and “stayed about the same” to 0, and “worsened a little” and “substantially worsened” to 1.

<sup>190</sup> The fact that I assign the same weights to the deaths of both a respondent’s parent and pet is comical to be sure, but I do so to avoid arbitrarily picking and choosing weights, and future research could certainly improve upon this methodology.

- “9. I developed a serious illness or sustained a serious injury.”
- “10. I took a leave of absence from school.”
- “11. I argued more and more with my parents.”
- “12. My mother or father passed away.”
- “13. A close relative passed away.”
- “14. My pet died.”

Concerning self-esteem, the Wave 3, 6, and 8 questions are as follows:

- “1. I am unable to solve some of my problems.” (Reverse-coded)
- “2. I am unable to control some things that happen to me.” (Reverse-coded)
- “3. I view myself with a positive attitude.”
- “4. I am quite satisfied with myself.”
- “5. Sometimes I think I’m useless.” (Reverse-coded)
- “6. Sometimes I think I have no redeeming qualities.” (Reverse-coded)

Scales created from these questions for Waves 3, 6, and 8 all have reliabilities above .7. Wave 4 includes the same questions, but the third and sixth are only included in the mail survey (administered to about ¼ of the survey respondents). An additional two questions are included in the phone survey (given to the other ¾ of respondents):

- “7. Are you a person of value (useful person)?”
- “8. Do you not have anything to be proud of?” (Reverse-coded)

Scales created from the mail and phone survey responses for Wave 4 have alpha reliabilities of .711 and .630, respectively, and are in comparable units.

Educational aspirations in Wave 3 are a combination of four questions:

1. “How far do you think you can go in school?” (Options are middle school, high school, junior college, college or technical institute, master’s degree, and doctoral degree.)
2. “Do you currently plan to go to college?” (Options are “no,” “it does not matter to me if I go to college or not,” “I have not decided,” and “yes.”)
3. “How important is getting a college education to you?”
4. “Generally speaking, how important is it to work hard for good grades?”

I standardize these four questions and create a scale from them that has an alpha reliability of .742. Respondents to Waves 6 and 8 are asked how far they believe they can go in school, as well as how far they would aspire to go if not restricted in any way. In Wave 6, questions about educational aspirations are unfortunately only asked of the 94.6% of respondents who are in school, and imputing the rest would be inappropriate, given that youth who are not in school no doubt have lower educational aspirations than those who are. Indeed, 78% of the respondents not in school in Wave 6 who answered the Wave 3 questions about educational aspirations had aspirations below the mean at that time. For the 99 respondents not in school in Wave 6, instead of lagging the Wave 3 values I therefore impute their Wave 6 educational aspirations using the Wave 3 cases, in order for all resulting Wave 6 values to be on the same scale. The Wave 6 scale (before imputation) and the Wave 8 scale have alpha reliabilities of .719 and .820, respectively.

Autonomy is measured in Waves 3, 7, and 9, and new scales to capture individuating and relating autonomy (Yeh, Liu, Huang, and Yang 2007; Yeh and Yang 2006) are not available until Wave 9. The Wave 3 autonomy questions ask respondents to indicate, on a four-point Likert scale, the extent to which they agree with the following:

- “1. I like to share some things that happen to me with my friends.”
- “2. I like to be able to go wherever I want to go.”
- “3. I like to become the center of attention in a group.”
- “4. I like to be free, doing whatever I like.”
- “5. I would rather do things together with my friends, and don’t like doing things by myself.”
- “6. I like participating in a group that has warmth and camaraderie.”
- “7. When I make plans, I hope to obtain the opinions of people whom I respect.”
- “8. When out in public, I like people to pay attention to my appearance.”
- “9. I like to say things that people think are smart and witty.”
- “10. I like to do things my own way, and don’t care what other people think.”

An additional question in Wave 3 asks respondents to state if in the future they would prefer to be someone who respects elders, someone good at making friends, or an independent person. The trouble with this question and Items 5 and 10 above is that they force respondents to choose between individuating and relating autonomy, when in fact these types of autonomy are in fact highly correlated and both benefit personal development in different ways (Yeh, Liu, Huang, and Yang 2007). A principal components analysis of the 10 Wave 3 autonomy questions in the above list bears this out by identifying three factors with eigenvalues greater than 1: 32.6% of the variance is explained by a factor positively associated with all questions, while two additional factors that appear to map to individuating and relating autonomy explain another 14.9% and 12.9% of the variance, respectively. For the present research I therefore autonomy as unidimensional, use the eight survey items that do not force respondents to choose between individuating and relating autonomy, and create a single scale with an alpha reliability of .751.

As indicated in the literature review, work can increase individuating autonomy, which in turn can lead to crime. Relating autonomy, in contrast, is associated with lower deviance, but is unaffected by work (at least among students of college age). Given the absence of yearly information about autonomy in the present study – much less individuating and relating autonomy – it is nonetheless useful to control for baseline individuating autonomy in Wave 3, and include this lagged autonomy in each year in high school. In this manner, the only way that autonomy could increase deviance is via the unobserved change in autonomy due to work, a change that is likely due to changes in individuating autonomy (though the present research could not prove such a mechanism). I would be comfortable in treating this unobserved mechanism as part of the causal effect of work for the present study, to be examined further in later research.

In Wave 7, respondents are asked, “Compared to the past few years, do you think that these days you are more able to do what you would like to?” Responses are on a four-point Likert scale. The question is problematic in that respondents with high autonomy who also had high autonomy in the past may answer “never” or “occasionally” as opposed to “sometimes” or “often.” Respondents are also asked if they still rely on their families, and if so if it is for financial assistance, emotional support, help with daily living, or discussing opinions. These questions are basically uncorrelated with the first question, and indeed involve a hidden assumption that individuating autonomy and relating autonomy are separate concepts; indeed, a sense of greater independence is weakly *positively* correlated with discussing ideas with family, supporting findings that individuating and relating autonomy are positively related (Yeh, Liu, Huang, and Yang 2007). Even though the question “are you more able to do what you would like to?” is problematic in its wording, it still has face validity and is positively correlated with the more reliable Wave 3 autonomy (Pearson’s  $R = .114$ ), so I use this measure for autonomy in Wave 7. Finally, Wave 8 respondents are asked, “Do you feel like you are an ‘adult’ now?” As up to 1/3 of respondents answer “I don’t know,” I do not use this question as a measure of autonomy in Wave 8, and instead lag Wave 7 autonomy in predicting selection into Wave 9 work.

Harsh and erratic discipline, following Glueck and Glueck (1950), for Waves 3, 4, and 6 is a sum of two measures: “For the same misdeed, does your [mother / father] sometimes punish you and sometimes not?” along with “In being disciplined by my [mother / father], sometimes I will get hit.” Glueck and Glueck theorize that discipline can be criminogenic when it is both harsh and erratic, so I sum these two questions to create a scale for each parent, then take the mean of the two scales to create a measure for parental harsh and erratic discipline. (When one parent is

deceased or otherwise not present in the child's life, I use the scale for the present parent.) In Wave 5, the only question on the survey involving discipline is "If you do something wrong, does your [mother / father] physically hit you, send you out of the house, or not allow you to come home?" As the answer is dichotomous, compared to the Likert scale questions in the other three waves, and as a question about erratic discipline is missing, I opt to use lagged Wave 4 harsh and erratic discipline as the Wave 5 measure. Discipline questions are missing in Waves 7 and 8, so to predict work in Waves 8 and 9 I use the lagged Wave 6 measures.

Monthly allowance is given in Waves 3, 4, and 6; I use these waves to interpolate allowance in Wave 5, then later impute missing values for approximately 200 respondents not observed in Wave 6. Additionally, as a few respondents in Wave 4 give numbers for their daily allowance that translate to unrealistically high monthly numbers, I right-truncate monthly allowances at NT\$15,000.

In Waves 7 and 8, respondents are asked if they have taken out student loans. Unfortunately, only respondents in four-year general-track colleges and science and technology institutes are asked if they have loans – a fact that can be problematic because families of students attending high schools and junior colleges in Taiwan must also pay tuition, and because students in four-year colleges may also leave school to work prior to graduation precisely because the loan is too much of a financial burden and they must pay it off.

Cram school is a straightforward variable to code: for Waves 3, 4, and 5, it is an indicator variable set to 1 if a respondent attended a cram school program separate from his or her school. For Wave 7, cram schooling also includes continuing education courses. The question was not asked in Wave 8.

Concerning social disorganization, Wave 3 and 6 respondents are asked to rate the extent to which the neighborhood in which they live is safe, clean, and suitable for raising children. Reverse-coded, these responses represent a community's level of social disorder (Sampson and Raudenbush 1999), with alpha reliabilities of .819 and .870, for Waves 3 and 6. Because I believe that the respondent's working or deviance would have minimal impact on overall disorder in the community, and because respondents can move between Waves 3 and 6, in modeling selection into work in Wave 6, I use contemporaneous social disorder rather than the Wave 3 measure.

Neighborhood attachment, as asked in Wave 3, is only two questions: "In general, do you feel an intimate connection to your neighborhood?" and "Overall, are you satisfied with your neighborhood?" Respondents answer on a four-point Likert scale. These two questions have a correlation of .368, not enough to create a scale (alpha reliability of only .538). Accordingly, I use the more widely-used measure, "Overall, are you satisfied with your neighborhood?" In Wave 6, an additional question, "In general, do you like your neighborhood?" is included. "Do you feel an intimate connection to your neighborhood?" is less correlated with the other two questions, so I create a scale out of standardized responses to the questions about liking and being satisfied with one's neighborhood, with an alpha reliability of .813. No questions are asked about neighborhood attachment in Waves 4 and 5. Because respondents can move, instead of using Waves 3 and 6 to interpolate responses that may have no meaning for these two waves, I lag the Wave 3 responses to model selection into Wave 5 work, and I use contemporaneous Wave 6 responses to model selection into Wave 6 work. (An assumption that I make in the process is that neither work nor deviance in Wave 6 affects the respondent's overall satisfaction with their neighborhood.) The effect of neighborhood attachment on deviance or selection into

work should quickly attenuate for respondents who move away from their old neighborhoods, but the construct would still have meaning. Wave 8 respondents are also asked about neighborhood attachment: if they have moved away from home (in order to attend college, for example), they are asked how satisfied they are with their present neighborhood. All respondents are also asked how satisfied they are with their original neighborhood where they family lives. Accordingly, I use both questions to create a single variable for Wave 8 neighborhood attachment, which references attachment to the neighborhood in which they currently live; as both questions use the same four-point Likert scale, all responses for this new variable are on the same metric. Because neighborhood attachment is not available in Wave 7 and because many respondents have likely moved between Wave 6 (high school) and Wave 8 (college), I use contemporaneous Wave 8 neighborhood attachment to model selection into Wave 8 work, and lagged Wave 8 neighborhood attachment to model selection into Wave 9 work. 44 Wave 8 respondents state that they don't know how satisfied they are with their present neighborhood, so I mark them as missing values and impute neighborhood attachment for these respondents.

School involvement is measured in Waves 3, 6, 7, and 8. In Wave 3, respondents are asked if they served as a class cadre, club leader, or assistant to the teacher, and if they represented their school in a competition in 9<sup>th</sup> grade. I convert these questions to indicators and sum them to a single scale with a maximum score of 4. In Wave 4, respondents are not asked about their involvement in school except on the written survey, only administered to about 500 students, so I use lagged Wave 3 school involvement to approximate school involvement in those two waves. In Wave 6 I could use contemporaneous school involvement to predict selection into work, but consider that problematic because contemporaneous part-time work is probably even more likely to predict selection into leadership positions in after-school activities. Accordingly, given no

better available measure, I still use lagged Wave 3 school involvement to predict Wave 6 work. School involvement in Waves 7 and 8 consists of the number of hours per week that respondents spent participating in club activities during the previous year.

Questions about neighborhood involvement are only asked in Wave 3 of respondents who have recently moved, so I use the Wave 2 questions, which ask how often respondents spend time with other children in the neighborhood, and how often their families spend time with other neighborhood families, to create a scale with an alpha reliability of .773. Using these values, I then impute neighborhood involvement for Wave 3 respondents not observed in Wave 2. Neighborhood involvement is also observed in Waves 6 and 8, but I do not include it as a contemporaneous predictor of work because the decision to work likely reduces time available for involvement in the community. Similarly, as many respondents move out of their neighborhoods to attend college, I do not use lagged Wave 6 involvement to predict work in Wave 8, so work during Wave 8 has no neighborhood involvement predictor. The Wave 8 neighborhood involvement measure is the frequency with which respondents participate in activities in the neighborhood in which they are currently living, on a four-point Likert scale.

For family involvement, Waves 3 and 4 ask how many days respondents have dinner with their families during an average week. The Wave 4 survey also contains the question, “Do you often spend time with your family on weekends? (Time with any of your family members counts, not just the whole family.)” The two variables are moderately correlated (Pearson’s  $R = .34$ ), but not quite enough for a convincing scale (Cronbach’s  $\alpha = .53$ ). Perhaps participation in cram school may limit many students’ ability to have dinner with their families, while they make up the time on weekends. The Wave 5 survey replaces the question about dinner with family members with one asking how frequently respondents have breakfast with their families, but the

correlation with time spent with family on weekends is even lower. For Wave 5 I stick with time spent with family on weekends, as Hirschi's control theory (1969:21-23) postulates that involvement in conventional activities reduces an individual's opportunity to commit crime, and crimes are more likely to take place on weekends than during breakfast hours. Family involvement in college is harder to measure, as only respondents living away from home are asked how often they contact their family, and no respondents are asked if they regularly interact with family members on weekends. Waves 7 and 8 do not include questions about family involvement, aside from questions about how often students not living at home contact family members and whether they have met with them for meals at any time during the semester. As the former question only encompasses respondents not living at home, and as the latter question does not ask about frequency of interaction, they are not suitable as measures for involvement.

Classroom cohesion in Wave 3 consists of the following questions:

“Thinking about the following descriptions of classrooms, do you think they apply to the present situation in your own classroom?”

“1. The students in my class always help each other.”

“2. I do not like talking with my classmates very much.” (Reverse coded)

“3. My classmates love each other devotedly, like a family.”

“4. Sometimes my classmates do not cooperate with each other.” (Reverse coded)

Together, these questions form a scale with an alpha reliability of 0.70. In Waves 7 and 8, respondents in college simply state, on a five-point Likert scale, how satisfied they are with their relationship with their classmates.

Four questions measure family cohesion in Wave 3:

- “Thinking about the following descriptions of family life, do you think they apply to yours?”
- “1. Family members will talk among each other when making decisions.”
  - “2. Family members like to spend their free time together.”
  - “3. When there are family events, everyone in my family participates.”
  - “4. Family members get to know each other’s friends.”

These questions comprise a scale with a reliability of .760. Wave 6 only contains the first and third questions, and the scale retains a reliability of .735.

Parents’ expectations for how far their children will go in school may also affect children’s decisions to work while in school. The Wave 6 and 8 parent surveys contain questions explicitly asking parents both to state the highest degree they wish their child could attain given no limitations on their advancement, and the highest degree they expect their child to realistically attain. These two variables are more highly correlated in Wave 8 than in Wave 6 (with a Pearson product moment correlation of .653 vs. .519), suggesting that parental expectations tend to become more realistic as their children get older. About 60 more parents in each wave answer the question about realistic expectations about educational attainment than the more aspirational wording, so I use the question that has fewer missing values in order to impute the rest with greater accuracy. Parents are not asked this question on the Wave 3 parent survey, so I use the equivalent question on the child survey that asks children how far in school they think their parents expect them to advance. About 590 (20.7%) of respondents state that they do not know, however, and for these cases I attempt to impute missing values using the rest of the sample. Finally, to measure parental educational expectations in Waves 4 and 5, I use a question in the student surveys asking respondents what approximate class ranking their parents expect them to attain, with responses that include (“It’s fine as long as I keep up,” “Average,” “Above average,” “Excellent,” and “Among the top three students in class.”

Strain associated with a child's dissatisfaction with his or her family's economic situation, over and above the family's actual household income, may spur the child to either find a part-time job or to commit crime. The Wave 3 youth survey contains the following three questions, which I standardize and use to construct a scale with alpha reliability .744:

- "1. How does your family's current economic situation compare to most families?"
- "2. What influence do your family's economic conditions have on your development?"  
(Responses range from "very restrictive" to "very helpful.")
- "3. Since starting ninth grade, how satisfied have you been with your family's economic situation?"

Wave 4 contains no questions about strain due to low income, while Wave 5 contains a single question asking youth respondents if family financial problems cause them worry, and Wave 6 contains one question asking respondents if they are satisfied with their family's financial situation. As respondents in college are not asked about strain due to low income, I lag the Wave 6 variable for selection models for these waves.

Strain due to school in Taiwan is difficult to measure because while poor academic performance might be considered a source of strain, in Taiwan the better students tend to internalize a greater degree of pressure. In particular, in Wave 3, a dichotomous question, "Do you feel pressure concerning the problem of getting into higher education?" is negatively correlated with responses to "Since starting ninth grade, have you been able to keep up with the pace of your classes?" A third question, "Do you feel that you never will be able to live up to your parents' expectations about your grades?" is positively correlated with both. Empirical tests of General Strain Theory have shown that strains including "negative secondary school experience such as low grades and negative relations with teachers" lead to crime, even while "the inability to achieve educational and occupational success goals appear to be unrelated to

crime” (Agnew 2006a). Experiential pressure due to the college examinations, according to Agnew, would likely not lead to crime because the strain is not seen as unjust (Agnew 2006a:102), whereas the negative emotions associated with an inability to keep up with coursework and a sense of not being able to live up to parents’ unrealistic expectations could indeed lead to crime. Accordingly, for Wave 3 I create a scale from these last two questions, and I hedge my bets in capturing negative emotions in the model for selection into work and deviance by also controlling for depression. In Wave 4, for school strain I use the only available related question, “Have you been unable to grasp your course material?” Responses are worded as yes/no. In Wave 5, three questions are available asking respondents if they have been unable to grasp the course material, if they feel that studying is taxing for them, and if they are worried about their grades. All three questions are positively correlated but together create a scale with a reliability of only 0.517, so I use the first question for consistency and because it is the most likely of the three questions to be linked to feelings of injustice. As the Wave 6 questions related to school strain are all linked to the pressure of entering college, I do not lag them to predict work and delinquency for respondents who have already entered college, and Waves 7 and 8 contain no questions about strain due to school.

Depression is measured using a shortened version of the SCLR-90 scale. The following questions, measured on a five-point Likert scale, are asked in Wave 3:

“In the last week, how much have you experienced the following uncomfortable feelings?”

“1. Headaches”

“2. Dizziness”

“3. Loneliness”

“4. Depression”

“5. Overly worried”

“6. Muscle aches”

“7. Insomnia, not able to get to sleep”

“8. Parts of my body feel numb or as if pierced by needles”

- “9. It feels like something is caught in my throat”
- “10. Parts of my body feel weak”
- “11. Wanting to violently beat or hurt someone else”
- “12. Unstable sleep or constantly waking up”
- “13. Often arguing with others”
- “14. Loudly shouting or breaking things”
- “15. Not wanting to live”

Waves 4, 5, 7, and 8 ask subsets of these questions, and the alpha reliabilities for the depression measure are .88, .74, .71, .79, and .90 for Waves 3, 4, 5, 7, and 8, respectively.

I use a single variable for online activity for Waves 3, 4, 5, 7, and 8. In Waves 3, 4, 5, and 8, the measure is a Likert scale-type indicator for how many times per week respondents go online, ranging from “never” to daily. For Wave 7, the measure is the number of hours per week that respondents go online. For Waves 4 and 5 I also have a measure for how often respondents frequent Internet cafés, on a five-point and six-point frequency scale.

Materialism, measured in Wave 8, is a scale derived from six questions, measured on four-point Likert scales:

- “To what extent do the following statements fit with your way of thinking?”
- “1. As soon as new products come out, I want to quickly go and buy them.”
- “2. I often wish I can keep up with fashion trends.”
- “3. It is important for clothing and other products to all be famous brands.”
- “4. To use fake knock-offs is to not have face.”
- “5. It’s worth spending time and money to keep up with fashion trends.”
- “6. Wearing famous brands of clothing and shoes makes people dazzled.”

The scale has a reliability of .87.

## APPENDIX F

### Missing Value Imputation Procedure for Chapter 4

For each wave used in Chapter 4, the present analysis imputes missing values using the `ice` multiple imputation extension in the `Stata` software package, written by Dr. Patrick Royston. `ice` works off the principle of imputation using chained equations, also known as fully conditional specification (FCS). Algorithms implementing FCS specify a conditional probability density for each covariate to be imputed, and iterate over these probability densities (van Buuren and Groothuis-Oudshoorn 2011). The approach then generates several new imputed data sets, and a researcher then conducts the original analysis on each data set separately and combines the estimates, such that each point estimate in the combined analysis is the mean of the corresponding individual point estimates, and the variance of each point estimate is a weighted sum of the within-imputation variance and between-imputation variance for the corresponding point estimates for each data set (White, Royston, and Wood 2010). `Stata`'s `mim` function implements the combined estimation of multiple imputed data sets.

The present analysis presents a few challenges often encountered during multiple imputation: most impose an assumption that these densities are multivariate normal distributions, but some of the variables in the present study (such as deviance) are heavily skewed, even after transforming the data. `Stata`'s `ice` algorithm relaxes this assumption with an approximate Bayesian bootstrap as an option to users. The present analysis also uses predictive mean matching (PMM) for continuous variables, an ad hoc procedure through which the only imputed missing values allowed are those that are observed in the original data. This procedure helps to preserve departures from normality observed in the original data (White, Royston, and Wood

2010:383) while also preventing impossible results such as negative income. By default `ice` imputes missing values for each covariate using every other covariate, but it also allows the user to directly specify equations used for regressions, and to impute values only for cases conditional on other variables which may or may not be themselves imputed. Using this implementation of conditional imputation, the present analysis performs an imputation on the full TYP data set, while only imputing values for each variable conditional on a respondent's having responded to the survey wave. Additionally, for questions only asked of respondents who are in school, in college in particular, have parents, or who have at least one friend, I only impute missing values if the respondents have the appropriate status. In each case, in the regressions predicting missing values for each covariate, I include as predictors all available covariates in the same wave such that every respondent who qualified to answer the question(s) related to the dependent variable in the regression also qualified to answer the question(s) associated with the predictive covariate (e.g. school attachment would be predicted by neighborhood attachment but not parental attachment, as some respondents in school did not have parents).<sup>191</sup> (Also, in a coding error I omitted family cohesion in Wave 3 – a variable containing no item non-response – from the regressions predicting missing values in other Wave 3 covariates to be imputed, and decided not to correct this error as it likely affected the results in only an extremely minor manner given the number of other predictors in the model.) `ice` allows users to specify the functional form of the regression for imputation of each covariate, and so I use logistic regression for binary variables,

---

<sup>191</sup> While extremely useful, `ice` is not perfect: a minor bug in its implementation entails that if a variable (such as hours worked per week during the academic year) is imputed conditionally based on a second variable (such as whether respondents worked during the academic year), and this second variable is itself conditionally imputed based on a third variable (such as whether respondents attended school), then only in the case of large syntax files or complex data sets, `ice` will incorrectly impute missing values of the first variable (hours worked per week) for every case in the data set. For Waves 8 and 9, I therefore run `ice` in two separate files: one that imputes a few additional cases for whether respondents worked, and one that does not impute additional cases for whether respondents worked but does impute work intensity and extensity for those who worked. I then use these files for separate regressions.

OLS regression with predictive mean matching for continuous variables, ordered probit regression for ordinal variables measured on scales similar to Likert scales, and multinomial logistic regression for the educational tracking variable. A useful additional feature, passive imputation, allows me to propagate imputed values for the tracking variable into dummy variables for the type of tracking without separately imputing missing variables for the dummy variables, and then I use the dummy variables to impute missing values for some variables (such as school commitment) for which some types of tracking status (such as military service or participating in the labor force but not attending school) are nonsensical.

One additional feature of `ice` is important to note: occasionally, in imputing a categorical variable, levels of a discrete variable will perfectly predict levels of a second discrete variable. The gender and educational tracking variables in the present analysis present one example: female respondents never enter military service, but knowing a respondent's gender could potentially predict selection into academic tracking. Perfect prediction can create enormous point estimates and standard errors of the coefficients for the variables perfectly predicting the outcome variable during one of the "chained equations" in the multiple imputation process (White, Daniel, and Royston 2010). `ice` addresses perfect prediction by automatically checking for such cases, alerting the user with a message when they are encountered, then conducting an augmented regression in which a small number of observations are included in the unobserved category (women in mandatory military service in the present case), and assigned low rates so as create only a negligible impact on the actual imputed values. In the present case, perfect prediction occurs when predicting Wave 8 and Wave 9 educational tracking with gender. In what appears to be a bug in `ice`, I found that in the context of my dataset containing about 200 variables and a single call to the program several hundred lines in length, the package only alerts

the user to the presence of one of the two perfect predictions, implying that it may not properly perform the augmented regression in the second case. As a workaround, I omit gender from the regression predicting educational tracking in Wave 8, as more time-varying covariates are available for that regression than for Wave 9, and some of those additional covariates in Wave 8 likely pick up the effect of gender in predicting tracking.<sup>192</sup>

White, Royston and Wood (2010) recommend a rule of thumb that the number of data sets users generate should be equal to the percentage of missing values for the covariate with the largest missing value percentage. In the present case, Wave 3 parental aspirations for how far their children will go in school are missing for 28.6% of cases, as many children simply do not know what their parents' aspirations are. The single-most missing item is allowance in Wave 5, which is 32.1% missing (although as mentioned previously, although I impute values for allowance for all respondents, I only use this variable for respondents in school as the Wave 4 question applies only to respondents in school). Accordingly, I generate 33 data sets.

---

<sup>192</sup> I have reported this bug and the second involving nested conditional imputations to the author of `ice`, Dr. Patrick Royston. Most but not all of the features of `ice` have been incorporated into the `-mi impute chained-` command in `Stata 13`, so users may wish to consider the features available in the `mi` package before proceeding to use `ice`.

## APPENDIX G

### Covariates Used to Model Survey Attrition

In selecting a set of Wave 1 covariates to predict survey attrition, I use the same variables that Wu, Chiang, and Hwang (2013) found predictive of survey attrition in the J1 cohort for the TYP survey. What follows is a description of each variable used.

Family cohesion and attachment is a scale comprised of nine questions, measured with a four-point Likert scale:

“To what extent do the following descriptions of family life apply to your family?”

“1. Relationships among family members are closer than their relationships with people outside the family.”

“2. Family members feel close to each other.”

“3. When making decisions, family members discuss them with each other.”

“4. My family members like to spend their free time together.”

“5. When there are family activities, everyone in my family participates.”

“6. We like to do some things together as an entire family.”

“7. People in my family make friends with each other’s friends.”

“8. When I feel frustrated, I can always be comforted by my family members.”

“9. When I need help or advice, I can count on my family.”

The scale has a Cronbach’s alpha reliability of .867.

School involvement is measured through leadership, a sum of three indicator variables concerning whether the respondent was a leader in their classroom, club leader, or assistant to the teacher in the last year. Additionally school commitment is measured by the student’s class rank (among the top five students, 6<sup>th</sup>-10<sup>th</sup>, 11<sup>th</sup> – 20<sup>th</sup>, 21<sup>st</sup> -30<sup>th</sup>, or below the top 30). A higher score denotes a higher rank (i.e. better grades). Community involvement is a scale comprised of three indicator variables for participating in community events, participating in community religious activities, and using community infrastructure (such as parks or athletic equipment).

Educational aspirations denotes the respondent's highest level of education that he or she would like to obtain in the absence of practical constraints such as economic limitations, converted into years. Finally, deviance for the J1 cohort in Wave 1 is a scale derived from the same set of nine survey questions asked of the J3 cohort in Wave 1 and listed in Table 2.1. I take the mean of all responses to deviance questions that the respondent answered, leaving the scale missing if the respondent did not answer any of the deviance questions.

## APPENDIX H

### Principal Components Analysis of Criteria for Marriage

Respondents to Wave 6 of the Taiwan Youth Project survey were asked to state which of 17 possible criteria they considered important when thinking about whom to marry. In comparing the criteria listed in Table H1 were compared to the six love types referred to in Cho and Cross (1995), it is evident that none are associated with mania, so a five-factor model (instead of one with six factors) is used here. Table H1 shows the rotated results of a principal component analysis in SPSS using varimax rotation with Kaiser normalization. The first factor, which might be referred to as “obligatory” and explains 17.2% of the variance in the model, seems to refer to individuals who marry based on common religion, politics, or because the families are well-matched or the matching is otherwise auspicious. The second factor, labeled “pragma (education)” in Chapter 5, captures 10.5% of the variance in the model and includes individuals who marry based on the respondent’s appearance and education. The third factor, “pragma (resources),” captures 8.0% of the variance and has a correlation of .55 with gender (with female coded as 1). The factor is accordingly generally associated with either female respondents who prefer someone with poor appearance and a good job and solid economic foundation, or else male respondents who prefer someone with a good appearance and weak economic foundation. The last two categories explain 7.2% of the variance and 5.8% of the variance, respectively, and both capture a mixture of eros (love based on attraction) and storge (love based on friendship). Of course, reifying these factors is generally a bad idea, as the loadings for these factors vary by individual, and so the labels are intended only to help to relate these dimensions to the theorized love attitudes.

**Table H1.** Marital Qualification Factor Analysis Component Score Coefficient Matrix.

Which of the following criteria are important to you when thinking about who to marry?	Component				
	1	2	3	4	5
(01) Good appearance	-.072	.460	-.355	-.030	.185
(02) Healthy	.000	.007	.080	-.126	.406
(03) Good moral character	.004	-.135	.048	-.108	.495
(04) Well-educated	-.077	.463	.053	-.018	-.058
(05) Good family background	-.013	.427	.094	-.002	-.178
(06) We're attracted to each other	-.046	-.010	-.145	.350	.112
(07) Good personality	.009	-.058	-.067	-.027	.415
(08) Economic foundation	-.039	-.038	.518	-.006	-.020
(09) Has a good job	-.043	.023	.450	-.101	.120
(10) Our families have similar socioeconomic statuses	.176	.161	.019	.030	-.062
(11) Suitable age	.033	.084	-.123	.195	.154
(12) Auspicious (horoscope)	.279	.039	-.056	-.018	-.002
(13) Good world-view	-.001	-.074	.111	.526	-.190
(14) From same province	.399	-.087	-.041	-.058	.016
(15) Similar religious views	.356	-.124	-.073	-.013	.091
(16) Compatible interests	-.028	-.006	-.029	.529	-.125
(17) Similar political views	.375	-.090	.034	-.008	-.047

## APPENDIX I

### Nonresponse among the J3 Variables Predicting Selection into Marriage

This section covers an analysis of item non-response among the covariates predicting selection into marriage. Table I1 presents the percentage of respondents who did *not* respond to each survey item in each wave associated with the corresponding row and column names. The denominator associated with each percentage, also given in the row labeled “N” at the bottom of the table, is the total number of respondents who both completed the retrospective questions about smoking and drinking in the 2011 survey and the survey associated with the corresponding column in Table I1; this is the total set of respondents whose survey respondents could theoretically be used for the causal models in this chapter. Notice that this number decreases with increasing survey wave, a fact that may be counterintuitive given the nature of survey attrition and which is caused by the retrospective nature of the survey questions used to measure smoking. The closer to 2011 that a survey was administered, the more likely that a respondent was observed in both survey years and the larger the sample size at the bottom of Table I1.

**Table 11.** Item nonresponse patterns for variables affecting selection into marriage for the TYP survey.

<b>Time-Varying Covariates</b>	<b>J3W6</b>	<b>J3W7</b>	<b>J3W8</b>
Household income	1.87	0.70	
Tracking	0.81	0.16	0
School / teacher attachment	0.24	0.23	
School commitment: grades	0.41	1.09	
School commitment: time spent on homework	0.41		
Parental attachment	0.16	0.16	0.31
Parental monitoring	0.24		
<i>Guan</i>	0	0.08	
Live with parents <sup>a</sup>	0.16	0	0
Peer deviance	1.06		
Deviance (IRT measure)	0.24	0	
Stressful life events	0.16	0.08	
Self-esteem		0	
Autonomy	0.16	0	0
If have student loans	0.49	0	
Cram school		0.08	0
Parental educational aspirations	17.3*	46.4*	
Depression	0.16	0	0
How often online	0.81		
Materialism	0.16		
Peer sexual behavior	37.9*		
“Dating and marriage are the same”	0.33		
“If I love someone, then I will love them my whole life”	0.41		
Cohabitation	0.41	0.31	
Maximum number of concurrent sexual partners respondent has ever had	1.14	0.39	
Currently in a relationship	0.24	0.47	0
Attachment to current partner	0.41	0.47	0
Length of current relationship	0.57	0.39	0
Ever had sex	0.24	0	
At what age do you want to get married?	0.16	0	2.15

(Continued on next page)

**Table I1 (cont.)** Item nonresponse patterns for variables affecting selection into marriage for the TYP survey.

<b>Time-Varying Covariates</b>	<b>J3W6</b>	<b>J3W7</b>	<b>J3W8</b>
Intention to marry current partner	0.24	0.55	0
Currently working	0.16	0	0
Marriage	0	0.08	0
N	1228	1283	1305
Percentage answering all questions	93.81 <sup>b</sup>	96.96	96.24

<sup>a</sup> “Live at home” for J3W8

<sup>b</sup> Leaving out parental aspirations for the child’s educational attainment and peer sex.

Only a few survey items in Table I1 exhibit high item nonresponse rates. As was the case for the J1 cohort (see Table 4.6), a large proportion of responses concerning parental expectations for the child’s educational attainment are missing, as the youth provided this information and many responded that they did not know what their parents’ expectations were. Similarly, a great deal of information about peer sexual behavior is missing, as a large proportion of respondents also stated that they did not know if their friends had had sex. Parental aspirations for their child’s educational attainment do not significantly predict marriage for any survey year, so the regressions predicting marriage do not include this variable. Peer sexual behavior turns out to matter, however, so this variable is simply coded to an arbitrary number for respondents who did not know about their friends’ sexual behavior, then interacted with an indicator variable for whether respondents stated whether they knew about their friends’ sexual behavior. This dummy variable is then included as a main effect in the logistic regressions predicting marriage, and the interaction effect for peer sexual behavior and knowledge of peer sexual behavior is also included in the model without a main effect for peer sexual behavior. Without loss of sample size due to nonresponse, then, the interaction effect represents peer sexual behavior for respondents who know about their peers’ sexual activity.

## VITA

Richard Callahan, a native of Maryland, is a quantitative sociologist with a specialization in criminology, Bayesian methods, and Chinese language. He holds a bachelor's degree in Mathematics & Social Science from Dartmouth College, where he studied social network analysis under Dr. Joel Levine. After graduating he worked as a police consultant in Los Angeles for Linder & Associates, where he contributed to a Plan of Action for the Los Angeles Police Department. While a student of Dr. Ross Matsueda at the University of Washington he conducted research for a year at Academia Sinica in Taipei as a U.S. Fulbright Fellow, and also studied Japanese in Yokohama and Mandarin Chinese in Beijing and Taipei. A recipient of the Western Society for Criminology's Miki Vohryzek-Bolden award for best graduate student paper, he plans to conduct policy-relevant research into social problems until there aren't any left.