Examining Gender Differences in Writing Skill with Latent Factor Modeling

Laurel Woods

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Education

University of Washington

2016

Committee

Deborah McCutchen

Robert Abbott

Program Authorized to Offer Degree:

Education

University of Washington

Abstract

Examining Gender Differences in Writing Skill with Latent Factor Modeling

Laurel Woods

Chair of the Supervisory Committee:

Dr. Deborah McCutchen

College of Education

Male students have historically under-performed in writing tasks compared to their female counterparts. This study examined writing skill differences between the male and female subjects of a sample of 4th and 5th grade students in the Pacific Northwest (n = 309) using an analysis of variance (ANOVA) to test for a possible gender effect on writing skill. A multiple-group confirmatory analysis (CFA) was used to examine possible structural differences between writing factors across gender groups, and a structural equation model (SEM) was used to examine writing factors predicting an essay task outcome for both gender groups. Despite evidence of a significant main effect of gender on writing within the sample, the multiple-group CFA indicated no significant structural differences between writing factors across the gender groups. The SEM results were not reported for this study because of the extremely high covariance between the writing factors in the model. The results of both the ANOVA and CFA are similar to other studies in the literature, and support a gender similarities hypothesis.

An examination of the Common Core standards will reveal that acquisition of the skills that define the college- or career-ready student necessitate an early and vigorous start in writing instruction. A portrait of the students who meet the Common Core standards is described in the handbook, *Common Core State Standards for English Language Arts & Literacy* (2010) to be students who "comprehend as well as critique" and "respond to the varying demands of audience, task, purpose and discipline" (p. 7). Looking past the portrait of the successful student to the year-to-year itemized standards, the path to comprehension, critical reading and meeting audience demands begins in kindergarten. The Common Core writing standards waste no time. From kindergarten onwards, they require exposure to and writing in varied genres, each accompanied by different sets of literary tools and skills. We can assume that the new standards require a constant upward climb towards complex and competent authorship. In meeting these standards, the teachers', administrators' and districts' concern becomes modifying instruction to meet the needs of the students to whom these standards are applied.

The achievement of intermediate elementary school boys (4^{th}-5^{th} grade) in comparison to their female counterparts is the concern of this study. Historically, boys in all age groups have under-performed girls of the same age on standardized writing measures (Bourke & Adams, 2011; Fearrington et al., 2014; Hyde & Linn, 1988; Lee & Otaiba, 2015; Maynard, 2002; Scheiber et al., 2015). More specifically, many boys struggle with compositional fluency, an essential skill for the many timed writing tasks (Berninger et al., 1996; Berninger & Fuller, 1992) that are common in today's standardized testing batteries.

The current emphasis of high writing standards in public schools as described in the Common Core standards (2010) warrants an examination of the issue in order to identify the contributing factors to boys' relatively lower performance on writing assessments than their

female counterparts. Since there is a lack of evidence that males and females differ at any age in general intellectual ability (GIA), it is unlikely to be a cognitive deficit (Camarata & Woodcock, 2006). Therefore, the present study asks (a) if there is a significant difference between the boys' and girls' overall writing skill in the sample of 4th and 5th grade students and (b) if there is a difference between the underlying relationships among the factors that contribute to the writing skills of the two gender groups. The following section will provide information on the present study's conceptualization of the writing process for developing writers, and define the reading and writing skills before examining their relationships across gender groups.

Cognitive Factors in Writing Production

The theoretical models of the writing process, as described by the work of Hayes and Flower (1980) and the modified model adapted for developing writers (Berninger & Swanson, 1994) have provided myself, and most of the literature cited here, a framework on which to hang our proverbial analytical hats. The following section will provide information about the elemental factors of the writing process, and how the process differs for developing writers, the present study's group of interest.

Working Memory

Working memory, as described by Lee Swanson (2011) involves the "preservation of information while simultaneously processing the same or other information" (p. 481). The completion of any function requires a distribution of limited resources to the concurrent tasks embedded within each function, and in the context of the present study, working memory is the limited resource of concern.

In the conceptual model of "The Simple View of Writing" (Berninger et al., 2002) the resource of working memory is required for (a) text generation, which results in the language

that is written, (b) transcription, which is the encoding process of the language produced during text generation, and (c) any executive processes such as revision, planning or self-regulation that are necessary for the completion of a writing task. These three processes are in constant competition for the resource of working memory, and as with any situation of limited resources and high demands, "something's got to give". The Simple View model illustrates the tenuous balance of working memory that beginning writers, or writers with disabilities (Graham,1990) must maintain to continue the act of writing. Generally, this balance ceases to impede the overall writing process once transcription is automatized (McCutchen, 1996). However, until the writer has automated the transcription process, the function of executive processes is compromised in favor of directing working memory to transcription, which results in words on the page. In extreme cases, re-direction of the crucial resource away from both executive processes and text-generation to the process of transcription leaves the writer in an even tighter bind, as the work of putting words to the page impedes the work of assembling the words to transcribe.

Olive and Kellogg (2002) examined these trade-offs and the conceptual distribution of working memory with both adult and third grade writers in France. By comparing the two groups on three different writing tasks, they supported their hypothesis that the main struggle in young writers' composition was the constant balance of resources, including working memory, between the higher-level or executive processes and the low level process of motor transcription. This was in contrast with the adult subjects in their sample, who, with automated transcription skills, seemed to devote most of their cognitive resources to the higher-level processes. The ability to direct more resources to executive processes (revision, planning, self-regulation, goal-setting) would explain a hypothetically mature writer's thematically, grammatically and expressively superior writing output, in contrast to a developing writer's struggle to strike a balance in the

distribution of cognitive resources, the same struggle writers with learning disabilities face (Graham, 1990). Olive and Kellogg's (2002) findings fit with previous comparisons of mature and developing writers, which have established that while younger writers harness resources to put words on the page, maturing writers who have mastered the transcription process gradually transition to attending to their writing goals, including the organization of the text they produce and the needs of the reader or audience (Bereiter & Scardamalia, 1987).

While there are established differences in how beginning and mature writers distribute working memory to cognitive processes, the difference between male and female beginning writers is less clear. This gender difference becomes less clear when working memory is considered as a collection of memory skills that can contribute the performance of writing or reading related tasks. For instance, Adams, Simmons and Willis (2014) found no significant difference in general working memory across gender, but did find that there was a significant correlation between writing fluency and verbal working memory for boys, and a significant correlation between writing fluency and visuo-spatial working memory for the girls in their sample. Their finding regarding the correlation between visuo-spatial memory and writing fluency for the female subjects in their sample is especially interesting, as a 2010 study found visuo-spatial memory skills of the same age group (4-5 year olds) to be crucial in predicting whether the subjects would meet U.K. writing standards (Bourke & Adams, 2010). Unlike the Adams, Simmons and Willis (2014) study, however, Bourke and Adams (2010) did not examine gender differences.

Although the present study did not include a measure for working memory, it is the central resource in the Simple View of Writing as well as the Hayes and Flower (1980) model described below, which is the foundation for the present study's analysis design.

The Hayes and Flower Model of Writing

Planning and Revision

In the model specified by Hayes and Flower (1980) through the examination of protocol analyses, writers' task environment and long-term memory provide context for their writing process by which the writer might set goals for the writing task. During the writing task, writers recursively generate content by retrieving information from their context, which is then organized according to the goals framed within the context. This process within the model is defined as "planning".

The content generated and organized within planning is read and edited by detecting weaknesses and evaluating the content and structure of the text already written in light of the goals set within the planning process, thereby reaffirming or modifying those goals within the process called "revision". All of the writing processes in this model happen recursively, but together, planning and revision conceptually represent a pair of bookends to the process of "translating" by providing a context for the words written.

However, planning and revision were not included in the model of writing skill for the present study because of the 10-minute time constrain on the key writing task (the WIAT-III writing subtest, Pearson, 2010). The time constraint on this task might suggest that there would be little time and limited incentive for the subjects to either plan or revise, even if they were inclined to do so.

Translation

The third component of the Hayes and Flower (1980) model of the writing process is translation, defined by the transformation of predicated material from (working) memory into acceptable sentences under the guidance of the writing plan as defined in the planning process. It

also involves the act of putting words to page, whether in handwriting, keyboarding, or other form of transcription. However, this aspect of translation was not emphasized in the Hayes and Flower (1980) model, since the act of transforming the fully formed English sentences into their orthographic forms on the page is not as challenging to mature writers as the processes of planning and revising.

Hayes and Flowers' (1980) global, undifferentiated description of translating ideas into text was not appropriate, however, for the developing writers whom Berninger, Whitaker and Swanson studied (1992). Berninger and colleagues (1992; Berninger & Swanson, 1994) instead proposed a modification to the Hayes and Flower (1980) model for developing writers wherein the translation process functions as two process for developing writers; text generation and transcription. This split accounts for the distribution of working memory resources to the process of transcription, which is separate from producing content until the writer's transcription skills are fully automatized (Berninger et al., 1992). In their study of lower-level developmental skills in beginning writing, Berninger et al. (1992) proposed a developmental progression of lower-level skills that complements the Hayes and Flower (1980) cognitive model of the writing process and introduced the idea that lower-level skills might constrain the transcription process of developing writers. In their discussion of a split in the translation factor, Berninger and Swanson (1994) defined "text generation" as the act of transforming ideas into language representations in working memory, and "transcription" as the act of translating the language representations into written symbols. For the age group in the current study, 4th and 5th graders, Berninger and colleagues found that lower-level linguistic skills for words, sentences and texts contributed to the translation process, and that orthographic-motor integration was no longer contributing to the writers' ability to generate text, meaning that translation was no longer

functioning as two separate factors within the 4th and 5th grade students. This might mean that for the present study, a model of the factors contributing to the 4th and 5th grade subjects' performance on the writing measure of interest might not include distinct text-generation and transcription factors.

However, studies that have examined gender differences in writing suggest that boys' performance on writing measures is due to lower compositional fluency or automatic letter-writing skills rather than writing quality after controlling for compositional quality (Berninger & Fuller, 1992; Berninger, Whitaker & Feng, 1996). So while studies that have examined writing skills at the 4th and 5th grade level suggest that the translation process works similarly to a mature writer's because the transcription skill has become automatized, studies that have looked at gender differences suggest that females' transcription skills might become automated sooner than their male counterparts' (Adams, Simmons & Willis, 2014). Because of this possible gender difference in the relationship between transcription and writing skill, the present study examines a factor model that included transcription and text generation, hypothesizing that there may be a significant difference across genders in the correlation between transcription and the writing measure due to boys' lower performance in compositional fluency within the literature.

The following section will discuss the literature on the intersection of gender and writing that guided the analysis of the present study, specifically literature that investigates the existence of the gender gap in writing.

Relevant Research in Gender and Writing

While the present study will investigate a possible mean difference between the male and female subjects' performance on an extended writing measure, findings from studies looking for a gender gap in writing achievement have been mixed (Lee & Otaiba, 2015, p. 41).

The gender similarities hypothesis (Hyde, 2005) suggests that males and females are more alike than they are different. Put into terms of effect sizes, the gender similarities hypothesis posits that the gender effect on most psychological measures would be close to zero, with increasingly rare occurrences of gender effect sizes in the small, moderate and large ranges. A meta-analysis of 46 studies (Hyde, 2005) that examined many psychological variables (aggression, leadership, reasoning, spatial reasoning, spelling, e.g.), indicated that most (78%) of the psychological variables showed gender differences in the close-to-zero or small range. This meta-analysis did not include a writing measure, but it did include studies that compared gender groups from "all ages" (Hyde, 2005, p. 583) on vocabulary skills, reading comprehension (close-to-zero effect) and speech production (small effect). The meta-analysis also included one study that compared spelling skill of adolescents, which indicated a moderate effect of gender (d = .45). A decade following the meta-analysis (Hyde, 2005), a different study (Reynolds et al., 2015) examined gender similarities across spelling and written expression using a nationally stratified sample (N= 2,027). This study found moderate gender effects for written expression (KTEA-II Comprehensive Form) and spelling across age groups, indicating a consistent gender gap in written expression from ages 7-19 in their sample. The consistency of the female advantage in written expression from the Reynolds study (2015) is in contrast to findings from Scheiber and colleagues' (2015) study. With their (Scheiber et al., 2015) sample, the female advantage on the KTEA-II Brief Form writing measure increased with age (6-21 years). In conclusion, while there is documentation of a female advantage in written expression, the developmental progression of the advantage is unclear. The potential existence of a persistent and growing gap (Scheiber et al., 2015) lends the examination of gender differences an air of

immediacy, but even the temporary existence of a gap during writing development is worth examining for the purposes of understanding developing cognition.

Studies on the female advantage in writing achievement also include qualitative studies that examine gender differences in genre-specific preference or performance. For instance, both Maynard (2002) and Fletcher (2006) comment on the particular difficulties that boys have with writing personal narrative texts. Unlike the present study, this literature is frequently framed within a motivation and achievement context, suggesting that the gender gap is connected to attitudinal and motivational differences rather than cognitive differences between groups within these studies.

Possible cognitive origins of the gender gap in writing achievement

If there is a gender effect on writing skill in the sample of the present study, what is at the root of the writing achievement difference across genders? The literature on cognitive writing processes provides the present study with two possible explanations for intermediate-elementary aged male subjects' comparatively lower writing skill in other studies: (a) the relationship between word-level language skills and transcription and (b) the relationship between transcription and orthographic fluency. Both of these reasons are inter-related, as they both concern a difference in the underlying relationship of a linguistic skill and at least one sub-component of translation process (text-generation and transcription) from the Berninger and Swanson (1994) adaptation of the Hayes and Flower (1980) writing process model. However, the present study will address these two options separately for the sake of clarity.

The relationship between translation and language skills

Berninger and Swanson's (1994) analysis of a multivariate battery administered to a sample of intermediate grade subjects (n = 300) led them to theorize that for intermediate grades,

the language skills for words or for smaller units of language constrained transcription, while the language skills for words, sentences and texts constrained text generation at this stage of development. For the present study, this might indicate that a gender difference in the relationships between either word-skills or sentence or text-level skills could contribute to a mean gender difference in written composition.

The relationship between transcription and orthographic fluency

The work of Berninger and Fuller (1992) suggested that gender differences in writing achievement are related to automatic letter writing skills, and that automatic letter writing skills are in turn closely related to orthographic skills, which are a skill set that underlies student performance on handwriting, composing and spelling, all areas where men and boys have shown poorer performance in comparison to their female counterparts (Berninger et al, 2008). In other words, orthographic skills can be conceptualized as those skills that contribute to the translation factor of the Hayes and Flower model (1980).

In their study with first through third grade subjects from the same general geographic area as the present study, Berninger and Fuller (1992) found an "asymmetry in gender differences" (p. 375) wherein boys were at an advantage in verbal fluency while girls excelled at orthographic fluency. The authors concluded that the male disadvantage for orthographic fluency in their sample was associated with their relative disadvantage in composition. Since the authors (Berninger & Fuller, 1992) did not attempt to make any developmental claims about orthographic skill, one may be cautious in generalizing the relationship of orthographic skills with transcription and its subsequent constraint on composition to the present study's older sample of fourth and fifth grade subjects.

The Present Study

The present study addresses three questions. The first question asks if there is a gender effect on the WIAT essay content scores of the sample. In other words, is there a significant difference between the male and female subjects' content scores that can be attributed to gender? The second question of the present study asks if there is difference in the underlying relationships among the writing factors across the gender groups. This question is intended to explore the possible differences in the relationships between the factors in a structural model across the gender groups. The third question asks if the two factors of writing processes (Text Generation and Translation) uniquely predict the Essay factor within either gender group. This question, like the second, concerns the relationships between factors, but it is going beyond the correlations between factors by asking if the writing process factors directly contribute to the essay outcomes of the gender group.

In summary, the present study will first address the question of the existence of a gender difference on the WIAT Essay (Pearson, 2010) task for this sample, and will then proceed to examine the relationships between the factors for significant structural differences across gender groups. Finally, the present study will test the hypothesis that either of the writing factors uniquely predicts the essay outcome for either gender group.

Method

Participants

The participants for this study come from a larger sample of participants in the Academic Language Project. Subjects from the larger sample with complete data for the measures used in the analyses were selected for this study. The subjects (n = 309) were fourth and fifth grade students (4^{th} grade, n = 113, 5^{th} grade, n = 196) from 6 public schools and 1 parochial school in

the Pacific Northwest. Data was gathered from subjects in 14 different classrooms: 5 fourth grade classrooms, 2 mixed fourth and fifth grade classrooms, and 7 fifth grade classrooms. There are 147 boys and 162 girls in the study sample. The mean age of the fourth grade participants was approximately 9 years and 8 months while the mean age of the fifth grade participants was approximately 10 years and 9 months. Although no data was collected on SES, our subject pool was 75.7% White/European American, 6.8% Asian, 7.40% more than one race, 4.2% Black/African American, 1.3% Native American/Alaska Native, 0.6% Pacific Islanders, and 0.3% other. Most students (90.3%) speak English only, and the remainder spoke one or more languages other than English.

Measures

Assessments were administered by trained research assistants during the beginning of the academic year. The students completed the WJ-III Subtests individually with a research assistant, and completed the rest of the assessments in a group (class) setting at their schools. The analyses conducted included the raw scores from each of the assessments described below.

Essay writing quality. The subjects completed the standardized essay subtest from the *Wechsler Individual Achievement Test, Third Edition* (*WIAT-III*, Pearson, 2010) as a measure of extended writing skill. This measure was the outcome variable for the present study. For this task, the subjects were given 10 minutes to write an essay by hand in response to the instructions: "Write about your favorite game and include at least three reasons why you like it." At the conclusion of a 10-minute session, the students were instructed to stop writing. Although the scoring of the essays produces three scores (content and organization, word count, and mechanics), the content and organization score was used as the outcome variable for this study because it was considered the best indicator of essay quality. The content and organization score

ranges from 0 to 20 points, reflecting five criteria: introduction/thesis statement, conclusion, paragraphs, transitions and reasons/elaborations. To score well on the essay's content score, students are required to introduce the essay with a clear thesis and provide clear, identifiable reasons and elaborations that support the thesis. In addition to these requirements, an essay including transition words and a clear conclusion to the essay would also earn a higher score. The test manual for the *WIAT-III* (Breaux & Frey, 2010) reports a reliability of .82 for fourth and fifth graders. The number of words in each essay was also assessed (see Table 1).

Vocabulary. As part of group testing, the subjects completed the nationally-normed vocabulary task from *Gates-MacGinitie Reading Tests* (MacGinitie et al., 2007) as a measure of reading vocabulary knowledge. The subjects were asked to respond to 45 multiple choice questions in the allotted 20 minutes by selecting a word from five options that best matches the meaning of an underlined word in a clause. The publisher reports that the task shows adequate reliability (Kuder–Richardson Formula 20 reliability = .86; alternate forms r = .77) (Lesaux et al, 2014). For the present study, this measure was included in the Text Generation factor because it reflects the transformation of ideas into linguistic representations in working memory by asking the subject to connect two different linguistic representations of the same concept. As a multiple choice task, this task requires minimal transcription, since it doesn't require the subject to translate the language into written symbols.

Writing fluency. This measure from the *Woodcock-Johnson III Tests of Achievement* (*WJ-III,* Woodcock et al., 2001) was included in the analysis to assess writing fluency, which includes writing speed, legibility, and syntactic accuracy. The subjects were instructed to write one sentence for each item by using the short list of stimulus words. The subjects were instructed that the goal of the task was to complete as many items as possible during the 7-minute timed

task. At the end of the allotted time, the subjects were instructed to stop writing. The test manual (McGrew et al., 2007) for the *Woodcock-Johnson III Tests of Achievement* (*WJ-III,* Woodcock et al., 2001) reports a reliability of .79 for children ages 7-11. This measure was included in the model of the present study as a variable contributing to the Transcription factor, since the main focus of the task is quickly transcribing the language (the stimuli words) into written symbols (the subject's sentence). Since this task requires the subject to use the stimuli words exactly as they are presented, the process of text generation is limited.

Morphological skill (sentence combination) The subjects completed this researcher-designed sentence combination task during group testing as a measure of syntactic flexibility and morphological awareness. Each of the eight items includes three sentences that the students are instructed to combine into one sentence without using the word "and" by changing the form of any of the words necessary. This last instruction is in contrast to the writing fluency task described above, where the subjects are instructed not to change the form of the words in any way. After 16 minutes, the subjects were instructed to stop writing. Inter-rater reliability between two scorers for scores used in the present study as well as a subset of scores from a second data collection with a similar sample was r=.994 (Pearson's *r*). Twelve percent of the items were independently scored by both scorers to ensure inter-rater reliability. This task was included in the Text Generation factor because it reflects the subject's ability to transform the three sentences within each item into a different syntactic structure than they were presented. This task differs from the writing fluency task, included in the Transcription factor, because it asks the subject to transform the linguistic representations by manipulating the clausal and morphological structures within the sentence.

Morphological skill (morphological production) The subjects also completed a researcher-designed morphological production task. Students were presented with a stem word followed by an incomplete sentence. Students were given examples and were instructed to complete the sentence by changing the underlined (prompt) word (e.g. *memory*: The party was _____.). Participants received one point for every word they correctly changed and phonological misspellings were allowed (e.g. *memorible* instead of *memorable* and *explotion* instead of *explosion*). This task was modeled after other sentence-completion tasks used to assess a construct that is commonly referred to as morphological awareness in the literature (e.g. Carlisle, 2000; McCutchen et al., 2009). For the purposes of this study it was included in the Text Generation factor because it involves the transformation of ideas into linguistic representations. Unlike the other variable within the Text Generation factor, this task requires the subject to write their response, meaning that it requires minimal transcription of the language within working memory into written symbols. Still, the present study includes this measure in the Text Generation factor because the scoring guidelines for this untimed measure prioritize the linguistic transformation of the items (e.g., *memory* to *memorable* or even *memorible*) over writing speed or spelling accuracy. Cronbach's alpha to test internal reliability for the morphological production task was .92, indicating very high internal reliability.

Spelling. The subjects completed a group-administered adaptation of the WIAT III spelling subtest as a measure of spelling skill. For this task, the subjects were instructed to listen to the research assistant read the items from the protocol: the word, the word in a sentence, and the word once again. After all the 38 items had been read, the subjects were allowed to ask the research assistant to repeat items. The WIAT III reports reliability between .87 and .96. This

measure was included in the Transcription factor because it requires the subject to translate the spelling items into written symbols without any kind of transformation or manipulation.

Analysis

In addition to an analysis of variance (ANOVA) to confirm the presence of a mean score difference across genders, the present study employs a test of invariance with a multi-group confirmatory factor analysis (CFA), as well as a structural equation model (SEM) drawn from Berninger and Swanson's adaptation of the "Translation" factor (1994), which includes a Transcription and Text Generation factor as well as an Essay factor predicted by a single variable, the WIAT Essay content score. The correlations among the six measures for girls (n = 162) and boys (n = 147), means and standard deviations can be found in Table 1.

Table 1
Correlations, Means and Standard Deviations for Measures Across Male and Female Subjects

																				Females	
Measure		1.		2.		3.		4.		5.		6.		7.		8.		9.		10.		11.		*M*	*(SD)*
1. Essay Content Score		--		.588	**	.433	**	.324	**	.366	**	.338	**	.421	**	.249	*	.341	**	.241	*	.405	**	8.44	(3.68)
2. Essay Word Count		.637	**	--		.404	**	.375	**	.301	**	.327	**	.438	**	.259	**	.347	**	.293	**	.328	**	109.38	(34.84)
3. Writing Fluency		.352	**	.483	**	--		.503	**	.371	**	.401	**	.454	**	.405	**	.375	**	.293	**	.328	**	19.23	5.5.5
4. Spelling		.190	*	.349	**	.418	**	--		.550	**	.625	**	.732	**	.610	**	.746	**	.631	**	.565	**	30.19	(7.14)
5. Sentence Combining		.177	*	.344	**	.459	**	.597	**	--		.638	**	.671	**	.417	**	.512	**	.413	**	.566	**	31.83	(18.59)
6. Vocabulary		.157		.190	*	.403	**	.527	**	.576	**	--		.771	**	.339	**	.671	**	.558	**	.657	**	30.28	(8.23)
7. Morph. Production		.294	**	.402	**	.559	**	.671	**	.709	**	.721	**	--		.453	**	.761	**	.603	**	.720	**	23.23	(8.59)
8. Orthographic Choice		.197	*	.352	**	.458	**	.603	**	.495	**	.377	**	.538	**	--		.399	**	.247	**	.372	**	27.27	(7.22)
9. Word ID		.102		.207	*	.382	**	.698	**	.555	**	.593	**	.637	**	.495	**	--		.761	**	.630	**	57.81	(5.61)
10. Word Attack		.120		.196	*	.306	**	.661	**	.452	**	.390	**	.594	**	.498	**	.686	**	--		.583	**	24.47	(4.79)
11. Comprehension		.183	*	.299	**	.421	**	.557	**	.576	**	.671	**	.647	**	.453	**	.593	**	.508	**	--		31.62	(3.88)
Males	*M*	7.01		88.37		17.80		29.52		28.19		30.03		22.86		25.20		57.72		24.77		31.65			
	(SD)	(3.70)		(34.91)		(5.17)		(7.38)		(17.49)		(8.61)		(8.98)		(6.59)		(6.38)		(4.68)		(3.54)			

Note. N=309. Morph. Production = Morphological Production Task
* $p < .05$, ** $p < .01$, *** $p < .001$.

Analysis of Variance

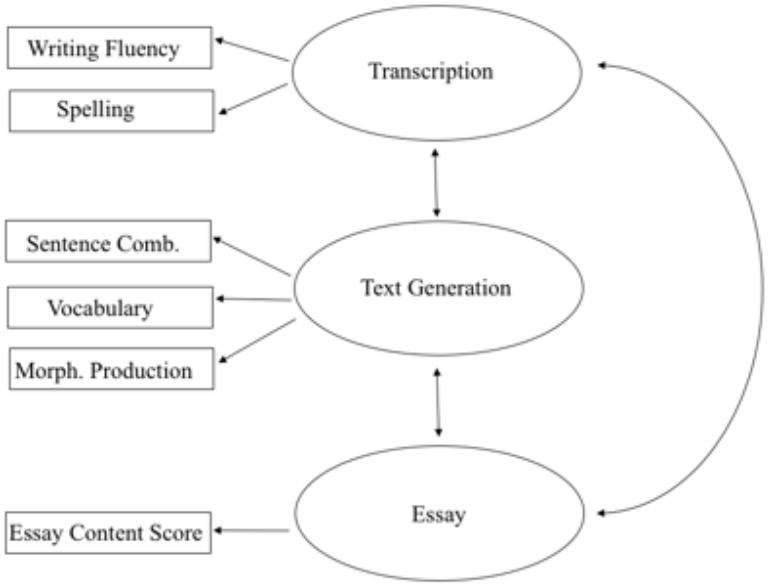
A one-way, between-subjects, fixed-factor ANOVA showed that there was a significant effect of gender on the subjects' WIAT Essay Theme/Organization sub-score, $F(1, 304) = 12.94$, $p < .001$, $\omega^2 = .04$. This effect size is small according to Cohen's (1988) standards, indicating that 4% of the variation in the essay sub-score was accounted for by gender.

After confirming the gender effect in the essay quality score, the present study proceeded to test for invariance of the underlying relationships between writing factors across both gender groups using three multi-group confirmatory factor models.

Tests of Invariance

The present study asks if there is a difference between the underlying relationships

Figure 1. Three-Factor Confirmatory Factor Analysis across Male and Female Subjects



among writing factors. This study examined the relationships among literacy measures within writing factors for both two gender groups by examining the fit of a factor structure using a multiple-group confirmatory analysis of the covariance matrices for boys and girls (Figure 1). The correlations among the six measures for girls (n = 162) and boys (n = 147) and standard deviations can be found in Table 1.

The results from the multiple-group confirmatory factor analysis and information about the factors that are modeled are shown in Table 2. Indicators of the Transcription factor were Writing Fluency and Spelling task. The indicators for the Text Generation factor were Sentence Combining, Gates-MacGinitie Vocabulary task, and the Morphological Production task. The WIAT Essay content score was the sole indicator for the Essay task factor.

Table 2.
Standardized and Unstandardized Path Coefficents with Test Statistics (Z) and Squared Correlations (R^2) for Confirmatory Factor Analysis of Male and Female Subjects

	Males				Females			
Relationship	Std.	R^2	Unsd.	*Z*	Std.	R^2	Unsd.	*Z*
Factor → measure								
Transcription → Writing Fluency	0.60	0.37	3.12	7.31	0.58	0.33	3.10	7.44
Transcription → Spelling	0.71	0.50	5.21	8.56	0.87	0.77	6.25	11.58
Text Generation → Sentence Combining	0.76	0.58	13.32	10.53	0.73	0.54	13.60	10.48
Text Generation → Vocabulary	0.76	0.57	6.51	10.43	0.83	0.69	6.81	12.52
Text Generation → Morph. Production	0.94	0.89	8.46	14.48	0.93	0.87	8.01	15.06
Essay Task → Essay Score	1.00	--	1.00	--	1.00	--	1.00	--
Transcription ↔ Text Generation	1.00	--	1.00	16.63	0.88	--	0.88	17.85
Transcription ↔ Essay Task	0.39	--	1.43	3.76	0.42	--	1.55	4.91
Text Generation ↔ Essay Task	0.29	--	1.07	3.41	0.45	--	1.65	5.73

In the unconstrained multiple-group model of Translation, Text Generation and Essay task, the variances of the Translation and Text Generation were fixed to 1.0 for identification, and correlations among the factors were freely estimated for both gender groups. χ^2 (14, N = 309) = 30.35 and the comparative fit index was .98, indicating an excellent fit for the data.

To examine whether factor structures for the two gender groups differed, a second multiple-group factor analysis was conducted with the factor loadings of each indicator and the covariance among the three factors constrained equal for girl and boys. For this constrained model, χ^2 (22, N = 309) = 37.816 and the comparative fit index was .981, indicating an excellent fit for the data. The difference in chi-square between the constrained and unconstrained was 7.466 (df = 8, N = 309, p < 0.487). A follow-up Lagrangian multiplier test indicated that none of the constraints added accounted for the change in fit.

Structural Model

A structural model (Figure 2 & 3) was tested to answer the third research question of the present study, do writing factors uniquely contribute to an Essay content score across gender groups? However, the model with these two predictors will not be reported, because of the

extremely high correlation between the Translation and Text Generation factors (Table 3)

Table 3.
Standardized and Unstandardized Path Coefficents with Test Statistics (Z) and Squared Correlations (R^2) for a Structural Equation Model of Male and Female Subjects

	Males				Females			
Relationship	Std.	R^2	Unsd.	Z	Std.	R^2	Unsd.	Z
Factor → measure								
Transcription → Writing Fluency	0.61	0.37	3.13	10.59	0.58	0.34	3.13	10.59
Transcription → Spelling	0.75	0.56	5.77	14.32	0.83	0.69	5.77	14.32
Text Generation → Sentence Combining	0.77	0.59	13.53	14.96	0.73	0.53	13.53	14.96
Text Generation → Vocabulary	0.77	0.59	6.71	16.37	0.82	0.68	6.71	16.37
Text Generation → Morph. Production	0.93	0.86	8.21	20.83	0.94	0.88	8.21	20.83
Essay Task → Essay Score	1.00	--	1.00	--	1.00	--	1.00	--
Transcription ↔ Text Generation	0.99	--	0.99	19.55	0.90	--	0.90	18.70
Transcription ↔ Essay Task	3.38	--	12.15	47.36	0.24	--	0.89	0.76
Text Generation ↔ Essay Task	2.94	--	-10.91	47.35	0.23	--	0.84	0.74

Figure 2. Standardized Path Coefficients for Structural Equation Model for Male Subjects

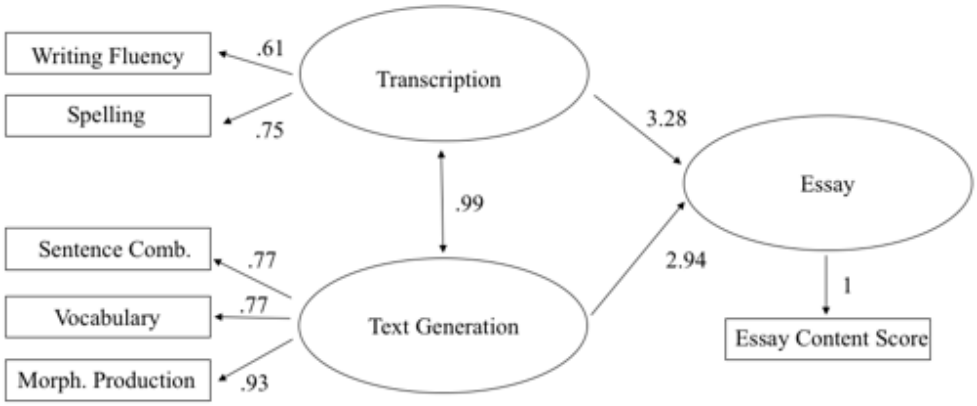
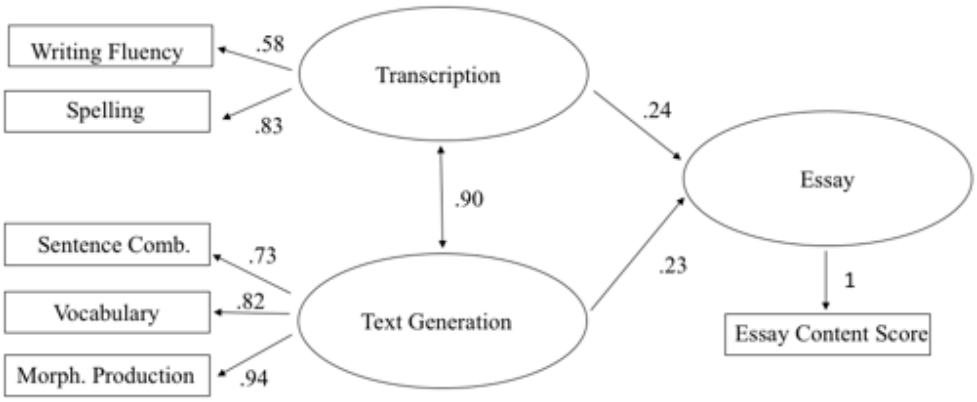


Figure 3. Standardized Path Coefficients for Structural Equation Model for Female Subjects



Discussion

Although an analysis of variance confirmed that there was a gender effect on the mean WIAT Essay content score, the underlying relationships among the writing factors were similar

in girls and boys, indicating that there was is no significant difference in factor relationships between the two groups. As a result, this study is unable to identify structural differences between the relationships among the writing factors across the gender groups.

A possible reason the analyses did not reveal any structural differences in the model could be the small effect size of gender on the WIAT Essay content score as reported for the analysis of variance. The analysis also shows that a possible reason for the lack of unique contributions from the writing model factors (Text Generation and Transcription) to the Essay factor is that the two factors are not distinct factors (Table 3). In the next section, I will draw from the literature on gender differences in writing to discuss the highly correlated factors and the structural invariance of the writing model that these analyses indicate.

Discussion of the sample

Based on the findings from the present study and the work of others, the following section will discuss some assumptions that can be made about the sample. Given that the self- and teacher-reported demographic information leads us to assume that this sample is largely typically developing (see Table 1 for means and standard deviations of measures in the testing battery), the findings from this study are only generalizable to students like them: typically developing fourth and fifth graders in public and parochial Seattle area schools. It is important to keep in mind that findings regarding the existence, onset, and severity of the gender gap in writing achievement might vary depending on the subject population. Lee and Otaiba's (2015) study using latent factor modeling to examine the contributions of both age and SES status to writing skill supported Entwisle and colleagues (2007), who found that there is an interaction between poverty and gender for writing performance. Specifically, Lee and Otaiba (2015) looked at a sample that was bifurcated into high- and low-poverty groups and found that the gender

effect was larger for boys who experienced high levels of poverty in contrast to both girls who experienced high levels of poverty and subjects of both genders from the low-poverty group. The present study found a significant, if small gender effect on writing performance, but it did not have access to any socio-economic status (SES) data on the sample, although the demographic is largely thought to be of middle or upper-middle SES. As a result, the present study cannot comment on an interaction between gender and SES for the sample; however, examining a gender and SES interaction across different levels of SES might be a possible avenue for future research.

Discussion of measures

In examining the tools this study used to measure the students' performance on writing, particularly the essay task (Pearson, 2010), we might assume that the students have full content knowledge of their essay topic, since it is an essay on their favorite game, or at least a game that they like for which they can think of three supporting reasons for their affinity. The opinion essay as a writing task reduces the concern that there might be a genre and gender interaction as qualitatively observed by Maynard (2002) and Fletcher (2006), who both comment on the particular difficulties that boys have with writing personal narrative texts.

However, this does not mean that the writing measure for the present study was the universal ideal of writing measurement. To echo the concerns voiced by Berninger and Fuller at the beginning of their 1992 study, many writing measures either emphasize product over the process of generating, or the process of sentence composition over discourse composition. By saying this, the authors are pointing out the imperfections in writing measurement as a whole, but not with the intention of turning the reader away from trusting writing measures. In pointing out the differing priorities of writing measures used in the field, Berninger and Fuller are

drawing attention to the compromises that exist within measures, compromises that are then reflected on to the findings of a study that employs them. In other words, the scoring rubrics used for many writing measures either emphasize the cohesion of the whole writing product or prioritize the sentence-level structural integrity within the essay over its thematic cohesion. In reflection, the variable of interest, the WIAT-III (Pearson, 2010) essay content score heavily emphasizes the five-paragraph structure of argument, which might suggest that it emphasizes the product over process. Additionally, the gender difference on mean word count on the essay task as well as the high correlation between the WIAT content score and essay word count (Table 1) suggest that the content score is tied to the length of the essay the subject writes. The mean difference between female word count (109.38) and male word count (88.37) might account for some of the gender difference in the WIAT essay content scores. The present study must take into consideration that with this particular writing measure, the content score reflects a mean difference between genders on a proxy for the five-paragraph structure, an important convention of academic writing. The measure may not reflect a mean difference between genders in holistic writing aptitude, creativity or linguistic skill.

Discussion of the analyses

Kline (2011, p. 94) stated that the ultimate goal of structural equation modeling was statistical beauty, defined in part by a theoretical rationale which differentiates between what is known and what is unknown, and sets conditions for posing new questions. While it may have been possible to create models different from the one examined here that showed significant structural differences between the two gender groups, the model examine here was theoretically justifiable.

The structural equation model, which drew upon Berninger and Swanson's (1994) model of writing processes, featured a Transcription and a Text Generation factor. These two factors were included in the model following the research that suggested that boys and girls' transcription skills are not automated at the same rate (Adams, Simmons & Willis, 2014), a developmental difference which might be reflected in the relationships of the measures that correlate to the transcription factor for each gender group. However, this was not the case, as the multi-group confirmatory factor analysis found no structural difference between the present study's male and female groups for the Translation and Text-Generation factor model. Looking back at the model of the writing process that Berninger and Swanson (1994) provide for this age group in particular, we might infer from the closely bound Transcription and Text Generation factors (see covariance in Tables 2 and 3) that the transcription process was not distinct from text generation process at this stage of our sample's writing development, meaning the act of assembling the language to write is not significantly differentiated from the act of putting the words on paper. This puts the present study's sample developmentally closer to the writers described in the Hayes and Flower (1980) model of writing, where transcription and text generation are merged into a process of translation.

The following section will explore a possible reason for the lack of evidence supporting the existence of structural differences between the gender groups' writing processes, namely, the possible structural similarity of boys and girls' writing processes. As discussed at the beginning of the discussion section, it is possible that the small overall gender effect on writing skill for this sample meant that the analysis was unable to capture any significant structural differences between the two groups. However, it is also possible that the mean gender differences observed by both this study and others are not related to structural differences. That is, girls and boys

employ similar processes, with girls simply employing them more effectively. The factor analysis approach of Lee and Otaiba's study (2015) with kindergarten students was unable to find any structural difference across genders and high- and low-poverty SES groups, with differences emerging only in a gender and SES interaction on writing measures. Their conclusion discussed the possibility that the difference was due to within-group differences, as the constructs were similarly correlated to each other across the groups overall, which implies that there were no measured structural differences in the subjects' writing achievement. While the sample of the Lee and Otaiba (2015) study is considerably different from the present study's, the structural similarity across genders may be a common finding for the two studies. Additionally, the common finding of a structural similarity between genders for writing skill links both the present study and Lee and Otaiba (2015) to the gender similarities hypothesis (Hyde, 2005; Hyde et al., 1988).

Discussion of future research

In the light of the structural similarity of the writing process across genders, one is left to consider other possible explanations for the writing achievement gap. While the present study did not include any qualitative observations about the sample, collecting data beyond quantitative data on writing performance in future research might reveal possible attitudinal, motivational or socio-cultural patterns for each gender group and contribute to the existing qualitative and mixed-methods research on gender differences in writing.

In his book, Fletcher (2006) describes the struggle that his male subjects face in writing. An overall pattern in their difficulty was feeling misunderstood as writers and being limited by classroom policies. Fletcher proposes that teachers allow those male students who feel discouraged to have more freedom to use slap-stick humor, violence, or fantasy in their writing.

While it is disconcerting that students feel discouraged from writing, Fletcher's suggestions for adapting instruction echoes an argument against the "feminization" of education (Jones & Myhill, 2007), an argument that reaches back to before women were admitted into the realm of formalized education (Maynard, 2002) but misses the larger point, that engaging in a discourse community such as writing is probably as individualized as it is gendered, and that gender identity is varied, meaning that ways of engaging in writing are different across individuals, not simply across two homogenous gender groups.

In addition to considering gender identity as a spectrum, not a dichotomous grouping, we must consider that each classroom has its own community of practice (Lave & Wenger, 1991) that determines which writing practices are valued. While one of the goals of formal writing instruction is to induct the students into a discourse of academic writing, the ways that the instruction is enacted and the writing that is rewarded vary across classrooms. In short, the issue of writing achievement is not a simple conflict between a feminized academic system and masculine boys who are unable to adapt, but rather an interaction among a multitude of varied gender identities interacting within a complex social practice. Transitioning to thinking about writing achievement in the context of this interaction led Jones and Myhill (2007) to suggest that perhaps this complex interaction contributed to the small and inconsistent effect sizes of gender on writing performance within the literature. In other words, the inconsistency of gender gap presented by the literature may be an artifact of the variation within the population of both gender identity and classroom communities of practice, as well as their interaction, upon writing practices. In turn, these writing practices are measured by writing measures that prioritize different aspects of the writing product or process, meaning that the final score an individual receives reflects not their writing skill, but a complex inter-weaving of many factors which

includes, but is not limited to, his or her writing skill as measured by a specific writing measure within the context of a given classroom community.

In summary, the present study suggests two possible directions for future research: examining the interaction between gender and SES with a stratified SES sample to expand upon Lee and Otaiba's work (2015), or examining the interaction between gender identity, classroom communities of practice and writing achievement with the use of ethnographic or mixed-methods methodologies. Either of these future directions would provide insight that could be harnessed for educational equity. Either of these future directions would provide insight into how the interaction of environmental and social constructs, acting within and upon the students, affects their achievement on the increasingly academically and vocationally important writing skills outlined by the Common Core Standards (2010).

References

Adams, A. M., Simmons, F., & Willis, C. (2015). Exploring relationships between working memory and writing: Individual differences associated with gender. *Learning and Individual Differences, 40,* 101-107.

Bereiter, C., & Scardamalia, M. (1987). The psychology of written composition. *The psychology of education and instruction series.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Berninger, V., & Fuller, F. (1992). Gender differences in orthographic, verbal, and compositional fluency: Implications for assessing writing disabilities in primary grade children. *Journal of School Psychology, 30*(4), 363-382.

Berninger, V. W., and Swanson, H. L. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. In Carlson, J. S. (Series Ed.) and Butterfield, E. C., (Vol. Ed.), *Advances in cognition and educational practice, vol. 2: Children's writing: Toward a process theory of the development of skilled writing* (57-91). Greenwich, CT: JAI Press.

Berninger, V., Vaughan, K., Abbott, R., Begay, K., Coleman, K., Curtin, G., Hawkins, J. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology, 94*(2), 291-304.

Berninger, V., Whitaker, D., Feng, Y., Swanson, H., & Abbott, R. (1996). Assessment of planning, translating, and revising in junior high writers. *Journal of School Psychology, 34*(1), 23-52.

Bourke L., & Adams, A. (2010). Cognitive constraints and the early learning goals in writing. *Journal of Research in Reading, 33*(1), 94-110.

Bourke L., & Adams, A. (2011). Is it differences in language skills and working memory that account for girls being better at writing than boys? *Journal of Writing Research, 3*(3), 249-277.

Breaux, K. C. & Frey, F. E. (2010). Reliability coefficients and standard errors of measurement for subtest component scores. Retrieved May 20, 2016, from http://images.pearsonclinical.com/images/assets/WIAT-III/WIAT3_TechReport2_Fnl.pdf.

Camarata, S., & Woodcock, R. (2006). Sex differences in processing speed: Developmental effects in males and females. *Intelligence*, *34*, 231–252.

Carlisle, J. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing: An Interdisciplinary Journal*, *12*, 169-190.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2007). Early schooling: The handicap of being poor and male. *Sociology of Education*, *80*, 114–138.

Fearrington, J., Parker, P., Kidder-Ashley, P., Gagnon, S., McCane-Bowling, S., & Sorrell, C. (2014). Gender differences in written expression: Curriculum-based measurement in third- through eighth-grade students. *Psychology in the Schools, 51*(1), 85-96.

Fletcher, R. (2006). *Boy writers: Reclaiming their voices*. Portland, Me.: Stenhouse Publishers.

Graham, S. (1990). The role of production factors in learning disabled students' compositions. *Journal of Educational Psychology, 82*(4), 781-91.

Hayes, J. & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.) *Cognitive processes in writing* (3-30). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hyde, J. (2005). The gender similarities hypothesis. *American Psychologist. 60*(6), 581-592.

Hyde, J., Linn, M., & Masters, John C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1), 53-69.

Kaufman, A. S., & Kaufman, N. L. (2005). *Kaufman test of educational achievement (KTEA-II): Brief form* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kline, R. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Lee, J.A.C., & Al Otaiba, S. (2015). Socioeconomic and gender group differences in early literacy skills: A multiple-group confirmatory factor analysis approach. *Educational Research and Evaluation, 21*(1), 40-59.

Lesaux, N., Kieffer, M., Kelley, J., & Harris, J. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents. *American Educational Research Journal, 51*(6), 1159-1194.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2007). *Gates-MacGinitie reading tests* (4th ed.). Rolling Meadows, IL: The Riverside Publishing Company.

Maynard, T. (2002). *Boys and literacy: Exploring the issues* (Language and literacy in action series). London; New York: Routledge/Falmer.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*(3), 299-325.

McCutchen, D., Logan, B. & Biangardi-Orpe, U. (2009). Making meaning: Children's sensitivity to morphological information during word reading. *Reading Research Quarterly, 44*(4), 360-376.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). Technical manual. *Woodcock-Johnson III normative update*. Rolling Meadows, IL: Riverside.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects. Washington D.C.: National Governors Association Center for Best Practices.

Olive, Thierry, & Kellogg, Ronald T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory & Cognition, 30*(4), 594-600.

Pearson (2010). *Wechsler individual achievement test* (3rd ed.). Bloomington, MN: PsychCorp.

Reynolds, M., Scheiber, C., Hajovsky, D., Schwartz, B., & Kaufman, A. (2015). Gender differences in academic achievement: Is writing an exception to the gender similarities hypothesis? *The Journal of Genetic Psychology, 176*(4), 211-234.

Scheiber, C., Reynolds, Matthew R., Hajovsky, D.B., Kaufman, A.S. (2015) Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychology in Schools, 52*(4), 335-348.

Swanson, H.L. (2011). Intellectual growth in children as a function of domain specific and domain general working memory subgroups. *Intelligence, 39*(6), 481-492.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside.