

© Copyright 2015

Hiroki Oura

Scaffolding Student Learning in the Discipline-Specific Knowledge through Contemporary Science Practices: Developing High-School Students' Epidemiologic Reasoning through Data Analysis

Hiroki Oura

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Philip Bell, Co-Chair

Learning Sciences & Human Development

Stephen T. Kerr, Co-Chair

Learning Sciences & Human Development

Leslie R. Herrenkohl

Learning Sciences & Human Development

Program Authorized to Offer Degree:

College of Education

University of Washington

Abstract

Scaffolding Student Learning in the Discipline-Specific Knowledge through Contemporary
Science Practices: Developing High-School Students' Epidemiologic Reasoning through Data
Analysis

Hiroki Oura

Chairs of the Supervisory Committee:

Professor Philip Bell

Learning Sciences & Human Development

Professor Stephen T. Kerr

Learning Sciences & Human Development

Science is a disciplined practice about knowing puzzling observations and unknown phenomena. Scientific knowledge of the product is applied to develop technological artifacts and solve complex problems in society. Scientific practices are undeniably relevant to our economy, civic activity, and personal lives, and thus public education should help children acquire scientific knowledge and recognize the values in relation to their own lives and civil society. Likewise, developing scientific thinking skills is valuable not only for becoming a scientist, but also for becoming a citizen who is able to critically evaluate everyday information, select and apply only the trustworthy, and make wise judgments in their personal and cultural goals as well as for obtaining jobs that require complex problem solving and creative working in the current knowledge-based economy and rapid-changing world.

To develop students' scientific thinking, science instruction should focus not only on scientific knowledge and inquiry processes, but also on its epistemological aspects including the

forms of causal explanations and methodological choices along with epistemic aims and values under the social circumstances in focal practices. In this perspective, disciplinary knowledge involves heterogeneous elements including material, cognitive, social, and cultural ones and the formation differs across practices. Without developing such discipline-specific knowledge, students cannot enough deeply engage in scientific “practices” and understand the true values of scientific enterprises. In this interest, this dissertation explores instructional approaches to make student engagement in scientific investigations more authentic or disciplinary.

The present dissertation work is comprised of three research questions as stand-alone studies written for separate publication. All of the studies discuss different theoretical aspects related to disciplinary engagement in epidemiologic inquiry and student development in epidemiologic reasoning. The first chapter reviews literature on epistemological instruction and explores theoretical frameworks for epistemically-guided instruction. The second chapter explores methodological strategies to elicit students’ disciplinary understanding and demonstrates an approach with a case study in which students engaged in a curriculum unit for an epidemiologic investigation. The last chapter directs the focus into scientific reasoning and demonstrates how the curriculum unit and its scaffolds helped students develop epidemiologic reasoning with a focus on population-based reasoning.

TABLE OF CONTENTS

Introduction to the Dissertation	1
Context for the Dissertation: Engaging High School Students in a Contemporary Practice of Epidemiologic Investigation	2
Conceptual Framework	5
Methodology	12
Structure of the Dissertation	13
References	17
Chapter 1. Disentangling “Discipline-Specific” Epistemologies: A Conceptual Framework to Reify Epistemic Alignments in Science Instruction	25
Abstract	26
Challenges of Promoting Epistemological Development Through Science Instruction	29
Disentangling Discipline-Specific Epistemologies With Epistemic Actors and Alignments ..	36
A Case of Applying the Epistemic Design Framework.....	45
Conclusion	55
References	58
Chapter 2. Eliciting Students’ Epistemic Sense Making: A Case Study of a Value-Focused Approach.....	66
Abstract	67
Theoretical Framework: Framing Scientific Disciplines as Epistemic Alignments	70
A Case Study of an Authentic Inquiry Unit Using the Value-Focused Approach	77
Conclusion	97
References	99
Chapter 3. Developing High-School Students’ Epidemiologic Reasoning through Statistical Database Analysis	106
Abstract	107
Theoretical Framework	110
Engaging Students in Epidemiologic Data Analysis: Exploring Databases (ExDa)	117
A Case Study: Examining Students’ Engagement in the Epidemiologic Inquiry	125
Conclusion	151
References	152
Conclusion to the Dissertation	160
Limitations and Directions for Future Research	165
Final Remarks	169
References	170
Appendices.....	172
Appendix A. Pre Survey Items	172
Appendix B. Post Survey Items	178
Appendix C. Coding Schemes	185

ACKNOWLEDGEMENTS

Design-based research can only be conducted through the collaboration of multiple researchers with different types of expertise. This work was possible through the support of many at the University of Washington, particularly my faculty advisors and project team. From when I first studied at the college as an international exchange student in 2005, my co-advisor Dr. Stephen Kerr has always supported and encouraged me with his positive feedback. Dr. Philip Bell gave me a chance to join the Exploring Databases project and graciously agreed to become my co-advisor, and I was able to develop my expertise in design-based research as a result of his mentorship. Dr. Leslie Herrenkohl and Dr. John Bransford kindly gave me their advice on and insights into learning theory every time I asked for their guidance. Dr. Noel Weiss also gave me his advice on epidemiologic research and practice. My team project members, Dr. Andrew Shouse, Dr. Katie Van Horne, Dr. Maureen Munn, and Mr. Mark Gallivan have supported my research work throughout the project. The wonderful people in the Everyday Science & Technology Group (ESTG) have always been supportive and helped keep me sane as a graduate student so far from my home country. Moreover, Mr. Hiroki Terashima in the Technology Enhanced Learning in Science (TELS) center at University of California, Berkeley graciously gave me his advice on software development. Finally, all the teachers and students I worked with always reminded me of why I want to devote my career to education. I am grateful to all of you.

DEDICATION

To my parents Nobuyoshi and Asami, who showed me the greatness of
constant diligence and unconditional love

Introduction to the Dissertation

Science is a disciplined practice about knowing puzzling observations and unknown phenomena (NRC, 2012). Scientific knowledge of the product is applied to develop technological artifacts and solve complex problems in society. Scientific practices are undeniably relevant to our economy, civic activity, and personal lives, and thus public education should help children acquire scientific knowledge and recognize the values in relation to their own lives and civil society (Linn, Davis, & Bell, 2004; Roth, 2004; Kuhn, 2005). Likewise, developing scientific thinking skills is valuable not only for becoming a scientist, but also for becoming a citizen who is able to critically evaluate everyday information, select and apply only the trustworthy, and make wise judgments as well as taking actions in their personal and cultural goals (NRC, 2009). Further, such high-order thinking skills are particularly critical for today's children to obtain jobs that require complex problem solving and creative working in the current knowledge-based economy and rapid-changing world (Bereiter, 2002; Levy & Murnane, 2004).

An important goal of scientific inquiry in empirical and applied sciences is to build causal explanations for unknown phenomena and puzzling observations. Scientists apply various analytical methods to identify causal relationships and develop explanations for the mechanism through investigations and argumentation (Osborne & Patterson, 2011). Theory is a set of existing explanations that can be used as more-or-less accepted knowledge in scientific communities. Scientists attempt to construct new knowledge by coordinating existing theory and evidence from systematic investigations (Kuhn, 1991). Evidence is the product scientists derive from their analytical methods. Scientists weigh a variety of evidence from multiple sources including counter-evidence, and apply complex reasoning to develop best explanations for their questions (Nicolaidou et al., 2011). Scientists suggest their findings for new explanations as

arguments and other scientists scrutinize the validity and values to their communities. Only the accepted part contributes to the shared knowledge, whereas the unaccepted aspects may call for the need of revisions or further study. This social process is scientific argumentation, and scientific communities have developed such disciplinary norms within their disciplines to produce high-quality knowledge (Longino, 1990 & 2002). Thus, scientific inquiry is the process of both individual and community efforts, and contemporary science education emphasizes engaging students in science learning through disciplinary engagement in authentic practices of science.

Context for the Dissertation: Engaging High School Students in a Contemporary Practice of Epidemiologic Investigation

Exploring Databases (ExDa) is a collaborative project to engage high school students in a contemporary research experience that is not covered in current standard science curricula (Munn et al., 2013). The curriculum unit focuses on: 1) engaging high school students in epidemiologic analysis with an online database from a real study and 2) helping them recognize an alternative view of scientific practices in which scientists conduct observational investigations other than laboratory experiments. Specifically, it engages high school students in hypothesis development and testing for risk factors of smoking to develop their claims for what causes people becoming regular smokers in the epidemiologic discipline. Most of high school students are expected to be unfamiliar with this form of scientific practice that involves statistical analysis with existing datasets, and it was the aim of the project to help them recognize this alternative form of practice other than laboratory experiments that are predominant in the current standard curricula (NRC, 2012; Windschitl et al, 2007; Windschitl et al., 2008).

Introduction to the Dissertation

To conduct the epidemiological investigation, the curriculum unit engages students in a series of analytical tasks. First, students are given introductions to understand basic concepts that are relevant to research on smoking and epidemiologic method. Subsequently, they learn case control design—an epidemiologic method that samples “cases” who continued smoking and “controls” who did not continue smoking by matching similar backgrounds aimed to evaluate the associations in the degree of exposures between them (Koepsell & Weiss, 2003)—and the study whose data is stored in the database. To analyze the data, students next learn about statistical inference and statistical concepts. Since the goal of the epidemiologic investigation is to build population-based explanations for risk factors of smoking, students need to understand the aim of statistical inference and the meaning of statistical significance with the concepts of population and sample (Bakkar & Derry, 2011; Hirsch & O’Donnell, 2001). To guide epidemiologic thinking, students are given a set of criteria for causality used by epidemiologic practices (Bradford-Hill, 1965), and they need to critically scrutinize evidence to determine which exposure can be risk factors of smoking. In sum, students engage in this series of complex analytical tasks and thus the curriculum needed to develop various scaffolds (Wood et al., 1976; Pea, 2004) in order to make these activities feasible as well as helping students learn the epidemiologic practice.

In the first-year implementation, the project team conducted fieldwork in classrooms and observed student interactions during data analysis to evaluate the scaffolds embedded in the ExDa unit. Likewise, the team observed teacher interactions during PD workshops to evaluate whether those scaffolds were usable enough for teachers to implement the unit in their classrooms. The initial video analysis indicates that a variety of students’ argumentative reasoning emerged during hypothesis development and testing. Students applied various

interaction patterns (e.g., “if parents have close relationships, the kids would not smoke”) and functional explanations (e.g., “because parents would not want their kids to smoke”) based on their personal experience, observations, and beliefs. Sometimes, their ideas contradicted each other and argumentative discussions emerged, which led to their motivation to test competing hypotheses with the database. Most students were able to utilize the database, interpret the meaning of odds ratios, and determine the significance with the confidence intervals.

However, the analysis also identifies issues of a few students’ personal epistemology on cause and causal reasoning in the epidemiologic discipline (Lee et al., 2012). In short, the analysis indicates some students conflated the epidemiologic discipline with some other discipline (e.g., biology) and almost rejected causal indications from the evidence with a case control study. This is problematic, since students in the group consciously rejected the validity of causal reasoning with correlational evidence. One student asked a question, whether they can talk about causation from case control design and another student told her that they cannot, because it is not “controlled”. Then, she wondered why then there are the criteria for causality and asked the same question to their teacher. The teacher first appeared to confirm that they cannot discuss causation from case control studies, students showed the teacher the criteria, and the teacher also appeared to be unsure about the judgment. Likewise, similar but even more severe cases were observed in teacher discussion at a PD workshop. Teachers were unconfident or confused in the probabilistic nature of epidemiologic explanations, and a few teachers explicitly rejected causal reasoning based on correlational evidence and contended that experiments need to be conducted to develop causal claims. In sum, these cases suggest that some students and teachers may have specific epistemological views on cause and valid methods

from other disciplines (e.g., biology) and/or their personal beliefs, which may mistakenly be applied to the epidemiologic practice in the classroom activity.

Conceptual Framework

Although scientific practices share a common goal of building causal explanations, their epistemological views on cause are not uniform across scientific disciplines (Schwab, Westbury, & Wilkof, 1978). As a result, even scientists dispute what takes to make a causal claim and there is no consensus yet within/across disciplines (Woodward, 2003; Cartwright, 2004). In life sciences, for example, functional biologists pursue causes of biological processes to explain the mechanism, while epidemiologists pursue causes of the occurrence and distributions of health-related events (e.g., disease) to prevent them and promote public health (Koepsell & Weiss, 2003). Each objective is legitimate along with their professional identities in society, while the difference in their major interests lead to their different views on cause. In general, biologists pursue building functional explanations or underlying mechanisms for biological processes in how biological causes/agents interact each other and they lead to observed effects (Mayr, 1988). In contrast, epidemiologists pursue relational explanations in what exposure or characteristic can lead to a disease (e.g., epidemic) in how it occurs and distributes over populations (Gordis, 2009). Risk factors are often defined as surrogates for deeper causes and better predictions (Stampfer et al., 2004), and this form of probabilistic explanation is sometimes put under criticism by some experimentalists due to the measurement errors and potential confounding from observational (non-randomized) design in some epidemiologic studies. However, it is valid for epidemiologists to seek this form of causal explanation, because 1) they seek population-based explanations for the occurrence and distribution in populations, rather than functional explanations in how every individual contracts them; 2) the occurrence and distribution of diseases often involve multiple

causes and pathways; 3) population-based explanations account for what proportion of people with a disease can be attributed to each risk factor (Rothman, 2002); and 4) even predictive indications are useful to screen people at risk and identify deeper causes through future research (Koepsell & Weiss, 2003). In sum, these epistemic interests in epidemiologic research influence on the form of causal explanation, methodological choice, and form of causal reasoning and judgment.

Consequently, such epistemological difference influences on their views on method, evidence, and reasoning as ways to justify causal explanations including the validity of methods. For example, some biologists observe case-based phenomena to develop their hypotheses for biological agents and processes to observed effects, followed by hypothesis testing through experiments to obtain causal evidence for theory development (Mayr, 1988). In statistical theory, “randomly manipulated” experiments can control or separate the main effect(s) in interest from other effects by the power of randomization, and which is a major source of the justification that experimental evidence is causal (Montgomery, 1991). To obtain less-biased evidence, these biologists carefully design experiments by randomizing subject/object assignments and controlling irrelevant variables in study procedures and (statistical) analysis. In contrast, epidemiologists conduct not only randomized but also observational studies such as case control design (Koepsell & Weiss, 2003) to obtain correlational evidence for specific populations, such as hospitals, by which they make causal inferences for risk factors. Correlational evidence does not necessarily suggest causal relationships due to, for example, the possibility of confounding in which an observed effect can be mixed with the major effect(s) in interest and those from other sources (Montgomery, 1991). Yet, observational methods are valued in epidemiologic practices, because manipulating human subjects sometimes becomes unethical in terms of exposing them

to potential risk factors; randomizing a large sample is economically impractical to obtain evidence with a enough statistical power for population estimates; already existing medical records can be sufficient to seek causal explanations (Koepsell & Weiss, 2003). To identify causal evidence from observational studies, epidemiologists apply disciplinary criteria to seek identifying true effects from observed associations in interest, which includes elaborate study design, statistical control in analysis, and causal reasoning with multiple sources of evidence including biological explanations from experiments in medical research (Bradford-Hill, 1965).

In sum, epistemological views on cause (e.g., functional or population-based explanations), methods (e.g., randomly manipulated experiments or observational methods), and the nature of evidence and causal reasoning are shaped by heterogeneous attributes/resources including goals/problems (e.g., explaining biological mechanisms or preventing people from diseases), subjects/objects (e.g., human or non-human), constraints (e.g., ethical concerns, limited budgets), and other attributes embedded in each practice. Further, the validity of those investigative tools (e.g. methods, evidence, ways of reasoning) is mutually understood and valued within each scientific community through the history of their own practice. Consequently, the validity of scientific practice and knowledge produced from it is not determined solely by its methodology or logical reasons. Rather, it can be understood only in relation to such goal-directed social attributes and the established values in scientific communities (Longino, 1990 & 2002).

The Role of Epistemological Understanding for Disciplinary Engagement

Moving the focus to the science classroom, how do students view the nature of cause and apply causal reasoning during classroom activities? Educational researchers have reported various student problems in causal reasoning across scientific disciplines including biology (e.g.,

Peker & Wallace, 2011), chemistry (e.g., Talanquer, 2010), physics (e.g., Clement, 1982), and other disciplines (Grotzer & Baska, 2003; Resnick, 1996). These findings suggest that the problems of student causal reasoning derive from their intuitive but naïve epistemological views on cause and lack of their recognition of theory-laden reasoning. One of the typical issues, which those studies suggest, is that students tend to draw causal reasoning and judgments based on direct visual cues or variables with less attention to the theory of the underlying variables or mechanism. Peker & Wallace (2011), for example, pointed out that students tend to develop biological explanations based on their first-hand observations with less reference to related theory, rather than coordinating theory and evidence (Kuhn, 1991). Resnick (1996) revealed that students tend to have a sort of “centralized” mindset such that a leader bird guides others in the flock and try to explain their observations based on the assumption; in a scientific explanation, however, this phenomenon is de-centralized in that each bird just follows a set of simple rules and it results in forming a flock. Grotzer (2003) explains that scientists’ and students’ explanations often have a different underlying causal structure. Given her precaution that the development of causal understanding involves many different dimensions, she argued, students tend to have limited causal repertoire and it leads to related issues in their reasoning such that students simplify causal structures and even distort their observation and information along with their naïve assumptions (see also Perkins & Grotzer, 2005).

In sum, prior research suggests that students have their own epistemologies or unconsciously frame the nature of cause, and which is a source of student problems of causal understanding and reasoning. *Personal epistemology* is one’s epistemological views or beliefs on the nature of knowledge and knowing (Hofer & Pintrich, 1997), and educational psychologists—based on a substantial base of empirical evidence—argue that one’s personal epistemological

beliefs influence on his/her reasoning and judgment (Kuhn, 1991; King & Kitchener, 1994; Schommer-Aikens, 2004). Likewise, researchers in science education argue that such intuitive but naïve epistemologies are a major source of students' misconceptions of causal understanding and reasoning (Sandoval, 2003 & 2005). As an alternative view to these, cognitive psychologists argue that students' particular ways of reasoning and judgments emerge due to a series of activations of cognitive resources in the individual mind and situated context, not necessarily because of self-conscious theory or beliefs (Elby, 2009). Although this theoretical discussion is still controversial at this point (see Sandoval, 2009), developing students' epistemological knowledge on cause and analytical process is an educational strategy for improving students' scientific understanding and reasoning.

The sources of such epistemological development could be various from informal resources (e.g., television, internet, magazine, newspapers) to formal education (NSB, 2010), and yet a major source would be students' experience in the science classroom (Bell & Linn, 2002). However, researchers argue that the current standard curricula provide students with limited experience to develop authentic epistemologies on science. For example, Chinn & Malhotra (2002) reviewed science textbooks and found that the types of reasoning required in scientific practices including generating research questions, designing studies, explaining results, developing theories, and studying research reports are not covered or epistemologically different in those textbooks. The latest science education framework encourages students to learn science through engaging in authentic scientific practices in which students apply causal reasoning with the evidence they identify through their investigations to develop and support their claims for explanations (NRC, 2012). In these moments, their epistemological or unconscious views will more likely emerge and influence their causal reasoning and judgments. As result, it may appear

that students just have made mistakes or developed misconceptions on scientific concepts and practices. However, the fundamental problem could be rather their misplaced epistemologies or naïve framing of the nature on cause, reasoning, and evidence. With this rationale, I argue that it is nontrivial for researchers to focus on design principles in order to help students develop discipline-specific knowledge on cause through authentic practices.

At the same time, however, it should be noted that it is challenging—and inappropriate—to expect that K-12 students develop epistemological knowledge in the same level with scientists. Scientists develop their disciplinary knowledge through a longitudinal and intensive engagement in expert training (Nespor, 1994), whereas students spend only a limited time period on each topic and subject as a science learner, which is another major issue in public education (Linn & Hsi, 2000). Furthermore, recognizing and distinguishing epistemological norms and rules between disciplines are difficult even for scientists as they dispute the definitions of cause and criteria of the validity (Woodward, 2003; Cartwright, 2004). Consequently, it is not hard to imagine that students are more likely to be put in a situation in which they mistakenly apply one specific epistemology from one practice to another, which I call *epistemological conflation* hereafter, and they are unable to resolve some *epistemological conflicts* between disciplines. Thus, I propose that developing students' epistemological awareness through engaging in different disciplinary practices is a more realistic goal and requirement in K-12 science education.

Developing Epistemological Awareness through Engaging in Epistemic Sense Making

Then, how should curriculum developers design curricula and activities so that students develop epistemological awareness on cause and analytical process in the science classroom? Although researchers problematize students' naïve epistemologies on cause, students are not provided enough instruction or opportunities to recognize and distinguish epistemologies of

scientific knowledge under standard science instruction. Historically, traditional curricula for each subject have been designed to teach their own subject, and their epistemological assumptions are not emphasized or explicitly distinguished from other disciplines (Stevens et al., 2005; NRC, 2012). Moreover, Chinn & Malhotra (2002) point out that the current science textbooks and activities are epistemologically different from real scientific practices and even some core forms are missing (e.g., generating research questions, conducting investigations, building explanations with the evidence, engaging scientific argumentation). To this problem, the latest science education framework will guide curriculum reforms in a right direction, because it pursues reforming science curricula in a systemic approach in which each curriculum is organized in relation to other scientific disciplines with both crosscutting concepts and discipline-specific concepts (NRC, 2012).

Assuming students will engage in authentic scientific practices in the future classroom, then, how do they distinguish one discipline from another? What kind of resources are critical and how should teachers guide students to develop their epistemological awareness? This is the overarching question I will tackle through this dissertation work, and I will attempt to develop design principles for curriculum and instruction in order to solve this empirical problem. As subsequent chapters will show the series of effort, in short, I approach to the solutions by engaging students in *epistemic sense-making* in which they are encouraged to think of “why” particular ways of investigations is valid to achieve the goals of practices in addition to “what” key scientific concepts are and “how” scientists conduct research (Tabak, 2004). As discussed earlier, each scientific practice has specific views on cause, reasoning, and evidence, and such discipline-specific knowledge can be understood only in relation to goal-directed social attributes such as goals and values shared within the communities (Longino, 1990 & 2002). The

dissertation work will iteratively design and evaluate artifacts including technological tools, worksheets, and activity formats to help students develop their recognition of the relationship between scientific knowledge and epistemic attributes embedded in scientific practices. The details of the resources and scaffolds and the results of learning analyses will be developed throughout the dissertation.

Methodology

The present dissertation work is conducted as a form of design-based research (DBR) (DBR Collective, 2003). DBR is a methodological approach that applies theory-driven design of artifacts, programs, and the environment to investigate how educational innovations work in practice. Although there are different origins for this type of research with a focus on *design* (Bell, 2004), the common goal of design approach is to improve and innovate educational practices through research practices (Brown, 1992; Collins Joseph, & Bielaczyc, 2004; Hoadley, 2004). Such strong commitment to development as part of scientific research reflects on the difficulty of making applicable, accountable, sustainable, and scalable effects by incorporating expertise from one or a few schools (e.g. cognitive or social theory), due to the complex nature of learning process and the variety of learners in their cognitive, social, and cultural backgrounds (Sawyer, 2006). Thus, the design process should be approached so that these various aspects are taken into account for developing a learning environment and both learning resources and activities fit the attributes of learners to solve empirical problems and accomplish design goals (Brown, 1992). Consequently, it requires different sets of “engineering of learning” practices by incorporating multiple theoretical disciplines and drawing boundaries for what needs to be designed and not. Therefore, a design practice is iterative rather than a single design-and-evaluation trial due to the demands of selection and coordination among design components

Introduction to the Dissertation

based on the distributed expertise (Linn et al., 2004; Edelson, 2002). In other words, researchers need to iteratively solve a variety of emerging problems between theory and practice through developing artifacts, evaluating results, and designing solutions.

From a scientific viewpoint, DBR could be viewed as a form of theory application rather than scientific research, whereas such design process and feedback from the practice indeed contribute to theory refinement and advancement. For example, Edelson (2002) argued that practical demands require a theory to be sufficiently specified and the design process reveals inconsistencies between theory and practice more effectively than analytical processes. Theory-driven design requires researchers to incorporate multifaceted knowledge during development, whereas some unexpected problems will more likely emerge due to the lack of and/or inconsistency to the nature of practice and learners in the context. To solve empirical problems, researchers need to revise the designed artifacts and seek alternative theories, which informs researchers revising points of existing theory or the need for new research to close the gap between theory and practice. In addition, such problem solving in design motivates researchers to identify new design components and/or its formations built on existing design theory (diSessa & Cobb, 2004). Thus, the advantages of DBR hold in the tension between theory and practice, and the goal-directed and theory-driven design leads to a natural focus for practical theory development. In other words, DBR will guide researchers to select and advance more usable theory among the alternatives in relation to the nature of contexts and learners.

Structure of the Dissertation

The present dissertation work is comprised of three research questions as stand-alone studies written for separate publication. All of the studies discuss different theoretical aspects

related to students' disciplinary engagement in the epidemiologic investigation and development in epidemiologic reasoning.

Research Question for Chapter 1: What Epistemic Actors Shape “Discipline-Specific” Knowledge and How Can We Translate Them into Curriculum and Instruction?

This study aims to focus on issues of epistemological conflation and conflicts and discuss theoretical approaches to disentangle “discipline-specific” knowledge that differs across different practices. First, the paper reviews literature in causal reasoning (Grotzer, 2012) and epistemological instruction including the nature of science (NOS) (Abd-El-Khalick, Bell, & Lederman, 1998) to discuss students' and teachers' epistemological issues in reasoning and understanding of scientific methods as discussed earlier. In short, most NOS studies focus on developing the common or general nature of science, whereas some researchers have recently argued the shift to “discipline-specific” knowledge to engage students in learning science through authentic inquiry (Bell & Linn, 2002; Rudolph, 2003; Sandoval, 2005; Wong & Hodson, 2008). To frame and discuss the discipline-specific nature of scientific practices, this study will identify “key” epistemic actors that differ across practices by drawing from philosophical and sociological discussions in science studies. In particular, it will draw from philosophical discussions by Laudan (1984), Longino (1990/2002), Rouse (1996), and others as well as sociological discussions by Pickering (1995), Latour (1987), Knorr-Cetina (1999), and others. After identifying “key” epistemic actors from this discussion, the paper then develops a theoretical framework for curriculum and instruction to analyze authentic practices along with those key actors and translate the alignments into inquiry tasks and scaffolds.

Research Question for Chapter 2: How Can We Theorize and Reveal Students' “Discipline-Specific” Knowledge?

The foregoing conceptual framework aims to reify “discipline-specific” knowledge to overcome epistemological conflation and conflicts. The first study attempts to theorize the components and formations of such knowledge into a theoretical framework, while there still is a need to develop the methodological approaches to elicit students’ discipline-specific understanding. With what kind of evidence can we say students have developed discipline-specific knowledge? Epistemic cognition is a type of metacognitive thinking to reason about cognitive activities and make sense of why particular tools and ways of doing things matter to a specific practice along the underlying goals and values, and other cultural reasons embedded in the practice (Kitchener, 1983). As discussed earlier, the validity of causal explanations and analytical processes in a scientific practice can be understood only in relation to various epistemic actors such as goals/problems, subjects/objects, constraints, and the values shared by the community of scientists.

To approach the assessment, the present study aims to discuss a theoretical approach with the emphasis on students’ understanding in the “value” of disciplinary norms and tools and attempt to demonstrate an assessment strategy with data collected in the ExDa unit. Philosophical and sociological discussions in science studies often use this concept of value, but it still is unclear in what it exactly means. To theoretically operationalize the concept, the study first elaborates a definition of value by drawing from Dewey (1939). In short, Dewey sees value as the product of a (physical and mental) tool being used to solve problems and achieve the goals. Under this perspective, it becomes possible to discuss assessment strategies for students’ disciplinary value. For example, case control design is valued in epidemiologic practices, but why is it “valued” in their practice? It can be said that epidemiologists sometimes need to identify causes given people with the outcome (e.g., diseases), and thus it often becomes

retrospective. To such counterargument that the evidence is just correlational but not causal, it is justifiable that it is unethical to randomly assign subjects into a potentially harmful condition. In other words, “ethics” takes a role as a social actor that helps students make sense of why case control design is preferred over experimental design. Subsequently, the criteria for causality are “valued” to distinguish causation from association with the correlational evidence. In this perspective, the present study will apply this value-focused approach to the ExDa context and discuss student improvement and revising points of instruction in discipline-specific knowledge.

Research Question for Chapter 3: How Does the ExDa Unit Engage Students in the Epidemiologic Investigation and Develop Students’ Epidemiologic Reasoning?

The ExDa unit engages high school students in epidemiologic reasoning that involves both biological and statistical reasoning. Statistical reasoning is a major characteristic of contemporary sciences (Hacking, 2001; Cartwright, 2004), while it is challenging to teach high school students statistical concepts, since it generally requires a substantial understanding in formal concepts involving mathematical representations and high school students are expected to be unprepared under K-12 mathematics curricula (Watson, 2006; National Mathematics Advisory Panel, 2008). Accordingly, this study discusses instructional approaches to help high school students develop statistical knowledge and reports findings from a subsequent study to investigate how designed scaffolds guided student engagement in the epidemiologic investigation. Specifically, classroom tasks were designed to develop student understanding of *population* and *sample* drawing from informal statistical inference theory (Bakkar & Derry, 2011). Informal statistical inference (ISI) is an instructional approach that aims to develop student understanding of statistical concepts without complex mathematical representations (Paparistodemou & Meletiou-Mavrotheris, 2008). The present study focuses on the population-

based reasoning of epidemiologic investigations and discusses whether and how our scaffolds guided students' disciplinary engagement in the epidemiologic investigation and how it developed students' population-based reasoning in the epidemiologic discipline.

References

- Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The Nature of Science and Instructional Practice: Making the Unnatural Natural. *Science Education*, 82, 4, 417-36.
- Bakker, A., & Derry, J. (2011). Lessons from Inferentialism for Statistics Education. *Mathematical Thinking and Learning*, 13, 5-26.
- Bell, P. & Linn, M. C. (2002). Beliefs About Science: How Does Science Instruction Contribute? In Hofer, B. K., & Pintrich, P. R. (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, N.J: L. Erlbaum.
- Bell, P. (2004). On the Theoretical Breadth of Design-Based Research in Education. *Educational Psychologist*, 39, 4, 243-253.
- Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, N.J: L. Erlbaum Associates.
- Bradford-Hill, A. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58: 295–300.
- Brown, A. L. (1992). Design Experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2, 141–178.
- Cartwright, N. (2004). Causation: One Word, Many Things. *Philosophy of Science*, 71, 5, 805-819.

Introduction to the Dissertation

- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically Authentic Inquiry in Schools: A Theoretical Framework for Evaluating Inquiry Tasks. *Science Education*, 86, 2, 175-218.
- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50, 1.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *The Journal of the Learning Sciences*, 13, 1, 15-42.
- DBR Collective (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, 32, 1, 5-8.
- Dewey, J. (1939). *Theory of valuation*. Chicago, Ill: The University of Chicago press.
- diSessa, A. A., & Cobb, P. (2004). Ontological Innovation and the Role of Theory in Design Experiments. *The Journal of the Learning Sciences*, 13, 1, 77-103.
- Edelson, D. C. (2002). Design Research: What We Learn When We Engage in Design. *The Journal of the Learning Sciences*, 11, 1, 105-121.
- Elby, A. (2009). Defining Personal Epistemology: A Response to Hofer & Pintrich (1997) and Sandoval (2005). *Journal of the Learning Sciences*, 18, 1.)
- Gordis, L. (2009). *Epidemiology*. Philadelphia: Elsevier/Saunders.
- Grotzer, T. A. (2003). Learning to Understand the Forms of Causality Implicit in Scientifically Accepted Explanations. *Studies in Science Education*, 39, 1, 1-74.
- Grotzer, T. A., & Basca, B. B. (2003). How Does Grasping the Underlying Causal Structures of Ecosystems Impact Students' Understanding? *Journal of Biological Education*, 38, 1, 16-29.
- Grotzer, T.A. (2012). *Learning causality in a complex world: Understandings of consequence*. Lanham, MD: Rowman Littlefield.

Introduction to the Dissertation

- Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge, U.K: Cambridge University Press.
- Hirsch, L. S., & O'Donnell, A. M. (2001). Representativeness in Statistical Reasoning: Identifying and Assessing Misconceptions. *Journal of Statistics Education*, Vol. 9, 2.
- Hoadley, C. M. (2004). Methodological Alignment in Design-Based Research. *Educational Psychologist*, 39, 4, 203-212.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67, 1.
- King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass Publishers.
- Kitchener, K. S. (1983). Cognition, Metacognition, and Epistemic Cognition. *Human Development*, 26, 4, 222-232.
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, Mass: Harvard University Press.
- Koepsell, T. D., Weiss, N. S. (2003). *Epidemiologic Methods: Studying the Occurrence of Illness*, Oxford University Press.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Kuhn, D. (2005). *Education for thinking*. Cambridge, Mass: Harvard University Press.
- Latour, B. (1987). *Science in Action*. Harvard University Press.
- Laudan, L. (1984). *Science and values: The aims of science and their role in scientific debate*. Berkeley: University of California Press.

Introduction to the Dissertation

- Lee, T., Oura, H., T., Scalone, G., Shutt, K., Bransford, J., Shouse, A., Van Horne, K., Vye, N., Munn, M., Knuth, R., & Bell, P. (2012). Gathering Evidence of Scientific Argumentation Practices: From Pre-Kindergarten to High School. In J. van Aalst, K. Thompson, M. J. Jacobson & P. Reimann (Eds.), *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012)* (Vol. 2, pp. 41-48). Sydney, NSW, Australia: International Society of the Learning Sciences.
- Levy, F., & Murnane, R. J. (2004). *The new division of labor: How computers are creating the next job market*. New York: Russell Sage Foundation.
- Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*. Mahwah, N.J: L. Erlbaum Associates.
- Linn, M. C., Davis, E. A., & Bell, P. (2004). *Internet environments for science education*, Mahwah, N.J: Lawrence Erlbaum Associates.
- Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, N.J: Princeton University Press.
- Longino, H. E. (2002). *The fate of knowledge*. Princeton, N.J: Princeton University Press.
- Mayr, E. (1988). *Toward a new philosophy of biology: Observations of an evolutionist*. Cambridge, Mass: Belknap Press of Harvard University Press.
- Montgomery, D. C. (1991). *Design and analysis of experiments*. New York: John Wiley.
- Munn, M. M., Oura, H., Gallivan, M., Van, H. K., & Shouse, A. W. (2013). Engaging high school students in research on smoking behavior. *Science*, 341, 6144, 360-1.
- National Mathematics Advisory Panel. (2008). *The Final Report of the National Mathematics Advisory Panel*. U.S. Department of Education.

- National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. Washington, D.C.: National Academies Press.
- National Research Council (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C.: The National Academies Press.
- National Science Board (U.S.). (2010). *Science & engineering indicators*. Washington, D.C: National Science Board.
- Nespor, J. (1994). *Knowledge in motion: Space, time, and curriculum in undergraduate physics and management*. London: Falmer Press.
- Nicolaidou, I., Kyza, E. A., Terzian, F., Hadjichambis, A., & Kafouris, D. (2011). A Framework for Scaffolding Students' Assessment of the Credibility of Evidence. *Journal of Research in Science Teaching*, 48, 7, 711-744.
- Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, 95, 4, 627-638.
- Papariotodemou, E., & Meletiou-Mavrotheris, M. (2008). Developing Young Students Informal Inference Skills in Data Analysis. *Statistics Education Research Journal*, 7(2), 83-106.
- Pea, R. D. (2004). The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. *Journal of the Learning Sciences*, 13, 3.
- Peker, D., & Wallace, C. S. (2011). Characterizing High School Students' Written Explanations in Biology Laboratories. *Research in Science Education*, 41, 2, 169-191.

Introduction to the Dissertation

- Perkins, D. N., & Grotzer, T. A. (2005). Dimensions of Causal Understanding: The Role of Complex Causal Models in Students' Understanding of Science. *Studies in Science Education*, 41, 1, 117-165.
- Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago: University of Chicago Press.
- Resnick, M. (1996). Beyond the Centralized Mindset. *Journal of the Learning Sciences*, 5, 1, 1-22.
- Roth, W.-M., & Barton, A. C. (2004). *Rethinking scientific literacy*. New York: RoutledgeFalmer.
- Rothman, K. J. (2002). *Epidemiology: An introduction*. Oxford: Oxford University Press.
- Rouse, J. (1996). *Engaging science: How to understand its practices philosophically*. Ithaca: Cornell University Press.
- Rudolph, J. L. (2003). Portraying Epistemology: School Science in Historical Context. *Science Education*, 87, 1, 64-79.
- Sandoval, W. (2003). Conceptual and Epistemic Aspects of Students' Scientific Explanations. *Journal of the Learning Sciences*, 12, 1, 5-51.
- Sandoval, W. A. (2005). Understanding Students' Practical Epistemologies and Their Influence on Learning through Inquiry. *Science Education*, 89, 4.
- Sandoval, W. A. (2009). In Defense of Clarity in the Study of Personal Epistemology. *Journal of the Learning Sciences*, 18, 1, 150-161.
- Sawyer, R. K. (2006). *The Cambridge handbook of the learning sciences*. Cambridge: Cambridge University Press.

Introduction to the Dissertation

Schwab, J. J., Westbury, I., & Wilkof, N. J. (1978). *Science, curriculum, and liberal education: Selected essays*. Chicago: University of Chicago Press.

Schommer-Aikins, M. (2004). Explaining the Epistemological Belief System: Introducing the Embedded Systemic Model and Coordinated Research Approach. *Educational Psychologist*, 39, 1, 19-29.

Stampfer, M. J., Ridker, P. M., & Dzau, V. J. (2004). Risk factor criteria. *Circulation*, 109, 25, 3-5.

Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005). Comparative Understanding of School Subjects: Past, Present, and Future. *Review of Educational Research*, 75, 2, 125-157.

Tabak, I. (2004). Synergy: A Complement to Emerging Patterns of Distributed Scaffolding. *Journal of the Learning Sciences*, 13, 3, 305-335.

Talanquer, V. (2010). Exploring Dominant Types of Explanations Built by General Chemistry Students. *International Journal of Science Education*, 32, 18, 2393-2412.

Watson, J. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, N.J: L. Erlbaum Associates.

Windschitl, M., Dvornich, K., Ryken, A. E., Tudor, M., & Koehler, G. (2007). A Comparative Model of Field Investigations: Aligning School Science Inquiry with the Practices of Contemporary Science. *School Science & Mathematics*, 107, 1.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the Scientific Method: Model-Based Inquiry as a New Paradigm of Preference for School Science Investigations. *Science Education*, 92, 5, 941-967.

Introduction to the Dissertation

Wood, D., Bruner, J. S., & Ross, G. (1976). The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17, 89–100.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Wong, S. L., & Hodson, D. (2008). From the horse's mouth: What scientists say about scientific investigation and scientific knowledge. *Science Education*, 93, 1, 109-130.

**Chapter 1. Disentangling “Discipline-Specific” Epistemologies: A Conceptual
Framework to Reify Epistemic Alignments in Science Instruction**

Hiroki Oura

College of Education, University of Washington

Abstract

Epistemological understanding is crucial to engage students in authentic scientific practices especially under the current science education reform movements associated with the most recent vision for K-12 science education. Researchers have discussed discipline-general epistemologies and instructional approaches for K-12 science education for decades, whereas recent studies suggest that discipline-specific epistemologies also play an important role for students' disciplinary engagement. Yet, there are few studies to discuss the specifics of epistemic attributes and formations that shape each practice differently, and the theoretical reification is an urgent need to discuss scaffolding design to guide students' disciplinary engagement in authentic inquiry. This article then discusses such epistemic attributes drawing from science studies and identifies key epistemic actors and alignments. Finally, an epistemic design framework is proposed to guide researchers to identify key epistemic actors in their focal practices, followed by a demonstration of the framework in an authentic-practice-based curriculum unit.

Chapter 1

The recent framework for K-12 science education encourages learning science through engaging in authentic practices (NRC, 2012). Inquiry-based approaches emphasize engaging students in a series of analytical tasks in scientific inquiry (Linn, Davis, & Bell, 2004; Krajcik & Blumenfeld, 2006), and the educational aim is to help them understand scientific concepts along with the “process” of science—disciplinary ways of reasoning about and engaging in the world. Inquiry activities represented in commercially available textbooks often lack authentic tasks that are epistemologically unique to science including generating research questions, planning investigations, choosing variables, collecting and analyzing data, and developing their scientific argument along with legitimate scientific knowledge (Chinn & Malhotra, 2002). Researchers ascribe students’ unrealistic views in science and limited reasoning skills to the inauthenticity of classroom inquiry, and the resulting research approaches these issues by transforming it into authentic inquiry (Driver et al., 1996).

A key challenge to engage students in disciplinary tasks is the activity and scaffolding design that develop students’ epistemological understanding in scientific inquiry (Schwab, Westbury, & Wilkof, 1978). Numerous studies have reported students’ lack of familiarity with theory-laden (causal) reasoning (Grotzer, 2012), credibility-focused evaluation of evidence (Nicolaidou et al., 2011), and argument construction and scientific argumentation (Bell & Linn, 2000). Although it is unrealistic to expect students’ epistemological understanding to be the same level of scientists’, literature suggests that deliberately designed activities and scaffolds enable students’ disciplinary engagement and contribute to their epistemological development (Linn et al., 2004).

Another challenge for students’ disciplinary engagement is how to help students distinguish *discipline-specific* aspects in one practice from another and adjust their

epistemological activity to the currently engaged practice. Although there are some common characteristics in the process of inquiry, which I call discipline-general epistemologies (DGEs) hereafter, *the ways of disciplinary reasoning and analytical criteria often substantially differ across practices in the nature of sought explanations, methods adopted for investigations, the nature of evidence, and reasoning with the evidence including criteria used to make causal judgments* (Laudan, 1984; Knorr-Cetina, 1999; Cartwright, 2004). Such discipline-specific epistemologies (DSEs) are complex and hard to understand by novices or outsiders, because their formation and rationale are intertwined with various natural and social actors such as subject matter, epistemic aims and values, and social aims and values including ethics. These natural and social actors are often dealt as premises or background information (Longino, 2002) and not explicitly reflected in instructional materials. However, these resources are rather essential for students to make sense of why each practice adopts particular methods and analytical tools and help them adjust their epistemological mode to the currently engaged practice. Otherwise, they would unconsciously conflate one discipline with another from their prior learning; if the two disagreed with each other in part, students could be confused by or unable to resolve the discrepancies or conflicts (Bell & Linn, 2002; Stevens, Wineburg, Herrenkohl & Bell, 2005). I call these situations *epistemological conflation and conflicts* hereafter.

In this study, I discuss a way to disentangle DSEs in scientific practices and propose a conceptual framework to guide the instructional design in which disciplinary characteristics are salient. In the following sections, I will first review theoretical discussions in epistemological instruction, and argue that science instruction should focus on developing students' epistemic cognition along with discipline-specific knowledge through authentic inquiry. Second, I will discuss the diversity of scientific disciplines and reify key epistemic actors that serve to frame

DSEs. Then, I propose an instructional design framework aimed to guide researchers to reify key epistemic actors for scientific practices and design instructional materials and scaffolds, followed by a case study of application with the framework. In conclusion, the scopes and limitations of the framework will be discussed.

Challenges of Promoting Epistemological Development through Science Instruction

A major goal of science education is to develop students' epistemological understanding in science by having them engage in and reflect on epistemic practices in the context of investigations (NRC, 2012). Numerous studies have reported that students tend to have unrealistic images of scientists and their practices (Driver et al., 1996) and the lack of epistemological understanding seems related to their limited experience with reasoning in classroom inquiry (Sandoval, 2003) including biology (Peker & Wallace, 2011), chemistry (Talanquer, 2010), physics (Clement, 1982), and other disciplines (Grotzer & Basca, 2003; Resnick, 1996).

One of the typical problems is the lack of “theory-laden” reasoning. For example, Clement (1982) demonstrated that students tend to apply causal reasoning on the relationship between force and acceleration based on directly visible variables such as velocity with less attention to Newton's laws even after instruction. Talanquer (2010) analyzed college students' written explanations in boiling and dissolution and found that students tend to generate mechanistic explanations based on the visual cues (boiling and freezing points), rather than making sense of the underlying theory that accounts for the phenomena (e.g., osmosis). Grotzer (2003) points out that students tend to have limited causal repertoire, so that these students simplify causal structures and could even distort their interpretations of observations along the lines of their naïve assumptions. Accordingly, she argues that students should be exposed to

various causal forms (e.g., mechanistic, relational, probabilistic) through science instruction (Grotzer, 2012).

Another well-known issue is that students and even teachers tend to believe there is a single research method (classic experimental design, or the “fair test” as it is often called in curriculum materials) applied to all scientific investigations or all scientists adopt a common method in their investigations (Driver, Newton, & Osborne, 2000). Windschitl and colleagues (2008) call it the scientific method (TSM) problem and ascribe it to the current standard curricula in which experimental design is dominant and inquiry activities are structured without epistemic variations. To solve this problem, they argue for “model-based inquiries” that reflect a form of practice involving various epistemic characteristics rather than framing scientific investigations only with methodology.

Nicolaidou and colleagues (2011) pointed out that evaluating the credibility of evidence is difficult from elementary to high-school students as well as undergraduate students and pre-service teachers. They defined the credibility of evidence as “the consideration of the grounds for confidence through the use of interrogatory tools, such as the critical asking of reports about the origin of evidence, whether the evidence is simply correlational or whether there is plausible theoretical mechanism, whether results are reproducible, whether they are contested, or about the authority of the scientific source” (p. 713). Nicolaidou et al. (2011) argued that credibility assessment is an important task as part of scientific and everyday thinking, whereas there are limited studies and discussions in literature on the topic. They proposed a theoretical framework that identifies nine key components of credibility assessment including “creating the need for students to examine the credibility of evidence”, “including evidence of low, moderate, and high

credibility of evidence”, and “support collaboration and peer review” along with the design implications for learning environments.

At the curriculum level, Chinn & Malhotra (2002) reviewed commercially available science textbooks and found that the types of reasoning required in scientific practices including generating research questions, designing studies, explaining results, developing theories, and studying research reports are missing or epistemologically different from real scientific practices in those textbooks. Although students’ unrealistic images on scientific practices and limited reasoning may derive from multiple sources including teachers’ limited knowledge on scientific epistemologies (Hodson, 2009; van Dijk, 2011) and informal resources through media (Dhingra, 2003; NSB, 2010), literature suggests that a major source would be the content and structure of science instruction in the classroom.

Debates on Instructional Goals in the Nature(s) of Science

To develop students’ epistemological understanding, researchers have worked to define scientific epistemologies and what knowledge of the nature(s) of science should be taught in K-12 science education, yet their discussions also revealed that there are different views on the primary focus of K-12 science education between the generality and specificity of scientific epistemologies. In one camp, researchers pursue identifying discipline-general epistemologies (DGEs), arguably, shared by all or most scientific practices (Abd-El-Khalick, Bell, & Lederman, 1998). The nature of science (NOS) is such a set of DGEs in the nature of “scientific knowledge, method, and way of knowing”—scientific knowledge is tentative, theory-laden, and necessarily involves human inference, imagination, and creativity; and is socially and culturally embedded (Lederman, 2007). To these NOS tenets including his earlier version of the list, Alter (1997) conducted a survey with philosophers of science in the United States and argued that there are

Chapter 1

significant disagreements among philosophers and doubted the legitimacy of “the” NOS (see also Rudolph, 2003). On the other hand, Smith and colleagues (1998) contested his critique by pointing out wording issues in his survey instruments and there is rather a considerable agreement in his data.

Osborne and colleagues (2003) pointed out another issue in Alter’s study design that his claim is only based on the community of philosophers, and conducted a Delphi study to identify DGEs with experts in more diverse communities including science educators, scientists, historians, philosophers, and sociologists. The results led them to characterize “ideas about science”, an alternative set of DGEs—analysis and interpretation of data, analysis and interpretation of data, scientific method and critical testing, hypothesis and prediction, creativity, science and questioning, cooperation and collaboration in the development of scientific knowledge, science and technology, historical development of scientific knowledge, and diversity of scientific thinking. Osborne and colleagues argued that their results challenge the case by Alter (1997), and the debate is no longer be marginalized on the basis that there is little agreement among academic communities.

In another camp, researchers argue the need of discipline-specific epistemologies (DSEs) for epistemological development. For example, Elby and Hammer (2001) discussed their doubt in the productivity in teaching DGEs as descriptive knowledge. They exemplified that scientific knowledge may be ultimately tentative, but the certainty also varies by explanation and it is inefficient to doubt more certain explanations (see also Songer & Linn, 1991). Elby and Hammer pointed out that one’s epistemological understanding is only reified in disciplinary contexts, and developing epistemological tenets as generalized knowledge separated from disciplinary contexts are incorrect and unproductive. Louca and colleagues (2004) extended this discussion and argued

that one's epistemological performance is shaped by various finer-grained cognitive resources (e.g., representation, task, people) embedded in context, rather than one's (static) epistemological beliefs.

In line with this, Samarapungavan and colleagues (2006) interviewed nearly one hundred individuals from high school students to professional chemists in their beliefs in the nature of chemistry, and found that their epistemic beliefs varied with the degree of their discipline-specific knowledge (e.g., epistemic aims and heuristics) and research experience. Likewise, Wong and Hodson (2008) interviewed thirteen scientists in different disciplines and asked about their day-to-day routines in their practices, and argued that there is no single set of DGEs that are static with time and fits all disciplines, followed by their concern in the underrepresentation of discipline-specific aspects of science in instruction. In contrast, Schwartz and Lederman (2008) conducted an interview with twenty-four scientists also in different disciplines and argued that their epistemological views on DGEs are not necessarily context dependent and there is no need for teaching different natures of science in K-12 science education. Together, there are still disagreements on the conceptualization of the nature(s) of science and debates on instructional goals are still continuing among researchers.

Developing Epistemic Cognition with Discipline-Specific Knowledge

The foregoing discussion illuminates two debatable questions. First, should science instruction aim to teach DGEs, DSEs, or both? A majority of researchers have pursued reifying and teaching DGEs as an instructional goal in the last decades (Lederman, 2007), whereas the legitimacy and representativeness has been challenged and still contentious. As Wong and Hodson (2008) point out, this debate will semi-permanently continue as the knowledge work of science becomes more diverse as already seen in contemporary epistemic practices (Knorr-

Cetina, 1999). To this seemingly endless debate, Elby and Hammer (2001), Louca et al. (2004), and Samarapungavan et al. (2006) brought a new line of research on the role of DSEs for epistemological development, which focuses on developing discipline-specific knowledge through disciplinary engagement in authentic epistemic practices. I argue that more research in the DSE approach is needed particularly associated with the new vision for K-12 science education that emphasizes engaging students in authentic practices in different disciplines (NRC, 2013), and it should be a key research agenda whether developing student knowledge in DSEs in different disciplines leads to reifying DGEs in disciplinary contexts. Another agenda would be to study whether students come to understand epistemological differences between disciplines and able to adjust their epistemological mode to the currently engaged practice in the DSE approach, which would help avoiding epistemological conflation and conflicts from their prior learning.

Second, should science instruction should focus on developing “epistemological knowledge” and/or “epistemic skills?” In one camp, researchers argue the need of developing epistemological knowledge by discriminating it from epistemic “skills” to engage in scientific inquiry and emphasize explicit instruction on specific NOS elements, although there is only a small subset of studies that demonstrate significant improvements (Lederman, 2007; Hodson, 2009). In contrast, other researchers such as Elby and Hammer (2001) seem to have less focus on teaching such formal epistemologies but more focus on developing their epistemic skills so that students are able to engage in disciplinary contexts thereby developing their discipline-specific knowledge. Note that these researchers who advocate the value of DSEs do not necessarily undervalue DGEs, but their claims rather imply that it is, at least, a prerequisite step to develop some kind of metacognitive skills to recognize epistemic characteristics in each practice through developing their discipline-specific knowledge before DGEs.

This type of metacognition is often framed as *epistemic cognition* (Chinn, Buckland, & Samarapungavan, 2011), and Sandoval (2014) discussed this seemingly identical but distinctive concepts drawing from Richard Kitchener's discussion (2002). In his paper, Kitchener distinguished the epistemic versus the epistemological by explicating that epistemology is the theory of knowledge (episteme) and the epistemic is one's personal epistemology (or folk epistemology in his term) or the folk theory of the epistemic, and thus it is not equivalent with the epistemology of knowledge. Sandoval argues that educational research should respect this distinction, and researchers should study the model and process of epistemological development in the midst of students' epistemic cognition and the form of science instruction (see also Sandoval, 2005).

Although there are multiple theoretical roots on the epistemic models and perspectives such as epistemic cognition (Karen Kitchener, 1983), personal epistemology (Hofer & Pintrich, 1997), and reflective judgment (King & Kitchener, 1994), Chinn and colleagues (2011) organized these roots into a theoretical framework as epistemic cognition (EC) with five components: a) epistemic aims and epistemic values, b) the structure of knowledge and other epistemic achievements; c) the sources and justification of knowledge and other epistemic achievements, and the related epistemic stances, d) epistemic virtues and vices, and e) reliable and unreliable processes for achieving epistemic aims. Chinn et al. view that epistemic cognition is a type of metacognitive ability or skills to recognize and reason in these epistemic components and formations. Herrenkohl and Cornelius (2013) applied the framework to analyze the epistemic quality of elementary students' historical argumentation under their scaffolding design; specifically, they provided students with explicit guidelines in predicting and theorizing, summarizing results, and relating prediction and theories to results. As a result, they found that

student conversations addressed all the five components in the epistemic cognition framework and concluded that students did engage in epistemically complex argumentation practices.

Need of Theoretical Guidance to Design Epistemically-Guided Instruction

Chinn et al.'s framework provides theoretical guidance to frame epistemic cognition and key aspects for the assessment as Herrenkohl and Cornelius (2013) demonstrated, and I now want to discuss another need of theoretical guidance for instructional design. In short, how can we design such epistemically-guided instruction—what kind of epistemic characteristics should be reflected, what kind of activities should engage students to develop students' epistemic cognition and discipline-specific knowledge, and how to design such inquiry activities and learning environments? Louca et al. (2004) indicates there are various types of cognitive resources that shape students' epistemological performance and development, and several theoretical frameworks have been proposed with a focus on designing inquiry “activities” and “scaffolds” (Linn et al., 2004; Sandoval, 2005; Windschitl et al., 2008; Nicolaidou, 2011). However, we still lack our understanding in specific epistemic features and formations that compose scientific disciplines and theoretical guidance to reify the DSEs across scientific practices. I argue that this agenda is also crucial to advance scaffolding theory for authentic inquiry, and some theoretical guidance is needed to analyze the schematics of scientific epistemologies and reflect the DSEs in science curricula. In the following section, I will discuss such key epistemic categories and then propose a conceptual framework to guide reifying DSEs.

Disentangling Discipline-Specific Epistemologies with Epistemic Actors and Alignments

Scientific practices share some general aspects in the processes. Scientific communities pursue developing knowledge to unknown phenomena and puzzling observations. Scientists employ various methods to identify empirical evidence and develop new causal explanations as a

form of argument built on existing knowledge. These arguments are reviewed by other scientists and if accepted they are shared with members of the community as scientific knowledge, and arguably this form of scientific inquiry is common in (empirical) science (NRC, 2012).

On the other hand, however, the nature of epistemic attributes in each process may significantly differ across sub-disciplines of science. For example, scientific knowledge or theory ultimately corresponds to causal explanations, whereas the concept of “cause” indeed is a complex construct and its exact definition and justificatory criteria are not identical across practices (e.g., mechanistic, probabilistic) (Hacking, 2001; Woodward, 2003; Cartwright, 2004). Likewise, methods need to be determined corresponding to a particular form of causal explanations and the forms of valid evidence (Pickering, 1995; Knorr-Cetina, 1999; Longino, 1990 & 2002). This is the focus of the research and development work in this study.

Epistemic Actors That Shape the Nature of Theory, Method, and Evidence

The major challenge in the sense making of DSEs is how to reify epistemic actors that shape the nature of theory, method, and evidence (Pickering, 1995; Rouse, 1996; Knorr-Cetina, 1999). For example, *subject matter* is a key natural actor that shapes the form of causal explanation and reasoning, and Mayr (2004) demonstrates such epistemological difference by contrasting biology to physical sciences. Using the concept of “biopopulation” that every individual is unique as the product of complex interactions in genetic programs and environments, he pointed out that biological characteristics can be described with the statistical mean of a population, whereas the properties can be variable in every individual (and these characteristics even gradually change from generation to generation as result of selection and evolution). Mayr argues that such nature is fundamentally different from the nature of physical objects as well as biological reasoning from physical reasoning (e.g., essentialism, determinism).

Observation/Measurement is another key actor that shapes method and evidence and varies among practices in relation to subject matters. Knorr-Cetina (1999) demonstrates the diverse nature of observation and measurement in contemporary research by contrasting two distinctive practices – high energy physics (HEP) and molecular biology. In HEP, the subject matter is particles and they are apparently too small for eye observations and it also is unsafe to manipulate them (e.g., collisions). Therefore, experiments require a massive technological apparatus and multiple large groups of scientists need to work together in the division of labor (e.g., the CERN particle super-collider). HEP physicists collect data with various detectors (Knorr-Cetina expresses a detector “sees” and “senses” things) and reconstruct the events by transforming them into specific representations, which is used for analysis to indirectly identify evidence. Knorr-Cetina points out that the raw data from these detectors are almost meaningless in the practice, since they contain various irrelevant signals. Therefore, HEP physicists systematically isolate that noise and extract only the signals in interest by computational amplification (e.g., Monte Carlo calculation).

In molecular biology, in contrast, biologists more rely on eye and sensory observations and incorporate various strategies to enhance the quality of their experience (“experiential register” in her term). Their experiments need to cope with “blind variations”, which occurs as the result of genetic mutation and natural selection, and it makes their experiments difficult to systematically control for these effects. Therefore, molecular biologists need to discriminate important evidence by incorporating various procedures and resources including laboratory protocols for dissection, visual scripts, and their experience and observations. Knorr-Cetina points out that the HEP practice presumes that scientists’ body is unreliable as a source of evidence and thus they use sensory detectors, whereas human sensory observation is essential in

Chapter 1

the biological practice and they incorporate various resources to cope with potential biases. In sum, the nature of observation and measurement is intertwined with the nature of subject matter, method, and evidence.

Epistemic aims and values also shape the form of theory, method, and evidence. Laudan (1984) explicated a philosophical analysis to theory-method disputes and argued that the empirical legitimacy of scientific theory and method can be determined only along with (cognitive) aims and values shared by scientists. Reflecting on the disputes between philosophers and sociologists, he points out that resolving disagreements in the level of “facts” (theory) requires the agreement in methodological rules, but it often becomes difficult since they are shaped by axiological aims and values that are partly explicitly or implicitly shared by scientists in their community, which may differ in other communities. Laudan argued that the validity of theory and method should be discussed in the coordination among all three levels (theory, method, and aims).

Mayr (1988) exemplified such epistemological difference due to aims and values by contrasting functional biology and evolutionary biology. According to him, the goal of functional biology can be stated as to explain the “operation and interaction of structural elements from molecules up to organs and whole individuals”. He views functional biologists ask “how” questions and develop causal explanations through observations and experiments. In contrast, the goal of evolutionary biology asks “why” organisms form and act at present as the product of natural selection and evolution from its past, and accordingly its inquiry becomes a type of historical or retrospective analysis. For another example, Stokes (1997) exemplified the diverse nature of epistemic goals and interests among scientific research from his analysis on political debates between basic and applied research. In general, basic research is motivated by the goal of

“understanding” focal phenomena without specific goals of application, whereas applied research aims at the “use” of knowledge produced from basic research to specific real-world goals and problems. Stokes pointed out that such different goals and interests lead to different prioritized decisions in research processes (see also Latour, 1987).

From another viewpoint, Stoke’s analysis on policy debates indicates how *social aims and values* shape the direction of scientific research. He further differentiates basic research from “oriented” basic research that is motivated by the goal of understanding but “toward” present or potential economic and societal interests, which shapes decisions and its processes in research. Longino (1990) also discussed such external influence on scientific knowledge and practices. She divides values into constitutive and contextual values—the former to principles or assumptions in scientific theories and methods and the latter to moral and social values—and exemplified how both values influence on each other by contrasting different scientific practices (see also Longino, 2002). Her analysis suggests that such social aims and values interact with various social actors such as budget and ethics that enable scientific practices but also constraint methods in some ways.

Framing Scientific Practices in Epistemic Alignments

The foregoing discussion suggests that the nature of theory, method, and evidence is shaped by heterogeneous actors including not only material and cognitive actors but also natural and social ones, whereas it should be noted that it is contentious whether to include social actors into the validation of scientific processes. Some philosophers like Laudan (1984) indicate that it should be or is possible to evaluate the validity of theory and method in the coordination with (epistemic) aims separately from social actors, whereas some philosophers and sociologists such as Longino and Knorr-Cetina argue that it is essential to incorporate heterogeneous actors

Chapter 1

including the social ones into the discussion. Rouse (1996) follows the latter argument by pointing out that scientific practices have always faced “resistances” against their epistemic aims including cognitive, material, and social problems and continuously reconfigured their formations of practices to solve them. From these historical backgrounds, he argues that science and its practice should be conceived by *epistemic alignments* in which both material and social, and normative and cultural actors are aligned, so as to reflect dynamically changing disciplines in scientific practices. Moreover, Rouse also points out that narratives is the only form that can capture these heterogeneous actors and alignments into an explanatory account and argues the enactment of *epistemic narratives* as a valid form of theoretical account in scientific practices (see also Latour, 2005).

Figure 1 shows a model of epistemic alignment based on the foregoing discussion in two epistemic layers. The inner layer shape the material/cognitive aspects of scientific disciplines including the type of theory, method, evidence, and (epistemic) aims, extended from the reticulated model proposed by Laudan (1984). Other epistemic categories in the layer such as observation and measurement are also reified as Knorr-Cetina (1999) demonstrated the diverse nature of scientific practices. In addition, the outer layer is used to reify natural and social actors that influence the formation of epistemic aim, theory, method, and evidence in a practice. This layer mainly draws on Stokes’s (1997) and Longino’s (2002) discussions that various social actors may shape the epistemic aims and the material and cognitive aspects of scientific practices. Note that although the figure is intended to visualize a schematic of epistemic alignments, the exact actors that matter to practices may change from one practice to another, and thus the model does not necessarily capture comprehensive characteristics of all scientific practices.

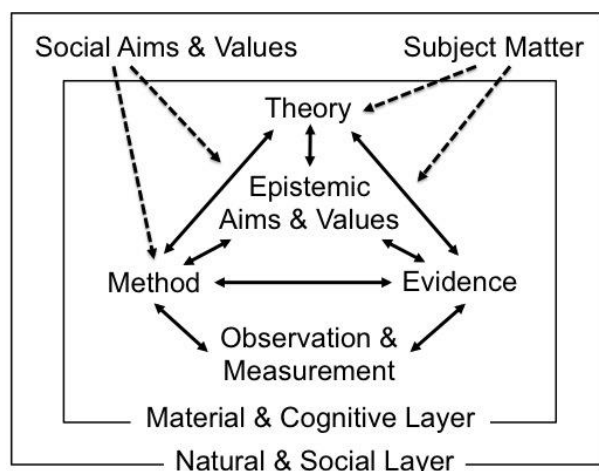


Figure 1. Epistemic actors and alignments in the material/cognitive and natural/social layers

A Theoretical Framework to Reify Epistemic Actors and Alignments

Previous sections discussed the role of epistemic cognition and discipline-specific knowledge in epistemological learning and the epistemic heterogeneity of scientific disciplines. Together, directing these discussions into a learner-centered question, how should we help students develop their epistemic cognition along with discipline-specific knowledge? The foregoing discussions suggest that students need to recognize heterogeneous actors and the epistemic alignments, yet this claim alone is still unhelpful and more specific theoretical guidance is needed to guide instructional design. The major challenge for this task, however, is that it appears unrealistic to develop a unified framework that accommodates all the diversity of scientific disciplines and practices that may consist of different epistemic actors and alignments.

To approach this challenge, I propose an *epistemic design framework* aimed to provide theoretical guidance for planning epistemically-guided science instruction. I use “epistemically-guided instruction” in such a way that those epistemic actors and alignments discussed in the previous sections are saliently reflected in inquiry activities and instruction, because it is not my goal to develop a single framework that encompasses all the variations across different practices.

Chapter 1

Rather, this framework aims to help researchers in the “design process” of reifying key epistemic actors and alignments in their focal practice through developing epistemic narratives that capture DSEs for their focal practice. I argue that these epistemic narratives are valuable for researchers to design scaffolds that develop students’ discipline-specific knowledge and epistemic cognition.

The framework consists of three major tasks with related conceptual tools: a) identifying key *epistemic actors* that represent epistemic characteristics of the focal practice, b) developing *epistemic narratives* and reifying the alignments with relevant actors, and c) *arranging the alignments* into a set of sub-alignments for classroom tasks. In the following, I will explicate the tasks and tools and demonstrate a case study as an application of the framework. Note that these tasks are not necessarily in order, and do not need to be considered concurrently.

First, it is essential to identify key epistemic actors that represent epistemic characteristics of the focal practice. Various epistemic actors and alignments could be reified in a practice, whereas too many actors and complex alignments would be unnecessary or unproductive relative to the goals and instructional capacity of K-12 science education. Accordingly, researchers should select only epistemic actors that are distinctive or contentious compared to other practices to optimize the disciplinary complexity of instruction. For example, epistemic aims and values are often key and unique in each practice and may, for example, be shaped by social aims to societal problems. Another example would be methods, since they are often contentious within and across practices, and understanding the epistemic alignment around the actor is crucial to avoid epistemological conflation and conflicts. Table 1 shows a list of epistemic actors discussed in the previous section and could be used to identify key epistemic actors in a focal practice.

Second, researchers should identify epistemic alignments through developing epistemic narratives associated with the key actor(s) in the previous task. Utilizing the list of epistemic actors in Table 1, they should ask which actors are key to make sense of the target actor, and developing the narratives would lead to identifying epistemic alignments along with the target actor as shown in Figure 1. For example, suppose a focal practice is to adopt an observational design, it would be crucial to understand why scientists in the practice adopt that method instead of the dominant method, experimental design. If the subject matter were human, ethics would be central to understanding the rationale, while other reasons might also be possible along with other analytical and social actors. Moreover, another would be to identify what analytical actor(s) is the key to overcome the limitations in causal reasoning with observational methods to develop causal theory (i.e., epistemic aim) through developing epistemic narratives.

Third, researchers should organize the alignments into a set of sub-alignments for classroom tasks so that students can develop their discipline-specific knowledge through instruction. The entire epistemic alignment often becomes complex, and it is unrealistic to develop the whole understanding at once. Rather, researchers need to organize them into sub-alignments so that each task can focus on developing a particular target actor(s) and alignment(s) at one time. For example, suppose epistemic aim and method are the target actors as instructional goals, researchers may develop a sub-alignment that explains the epistemic aim along with social aims and value and another sub-alignment that explains the rationale for the methodological choice (e.g., observational design) along with the epistemic aim and social constraints (e.g., ethics). Then they may develop two tasks in which each of the sub-alignments is addressed, and later combined as part of the whole instruction.

Alternatively, this framework could be applied to evaluate the instructional materials and scaffolds for existing curricula. In this scenario, researchers may first identify epistemic alignments for their focal practice and compare them with the instructional materials and scaffolds to evaluate the epistemological authenticity or whether these materials and scaffolds could sufficiently help students develop their discipline-specific knowledge for the focal practice.

Table 1

Key Epistemic Actors That Shape Discipline-Specific Epistemologies

Layer	Epistemic Actor	Definition
Natural/Social Layer	Subject matter	The nature of subject matter such as human and non-human subjects (e.g., physical particles, biological organisms)
	Social aim(s) & values	Societal/Communal goals and problems (e.g., curing cancer, eradicating malaria)
	Epistemic aim(s) & values	Scientific problems and theoretical interests (e.g., unknown phenomena, puzzling observations)
Material/Cognitive Layer	Theory	The nature of explanation (e.g., descriptive, mechanistic, probabilistic)
	Method	The type of method (e.g., experimental, observational)
	Evidence	The nature and criteria for evidence (e.g., descriptive, correlational, causal evidence)
	Observation & measurement	The nature of observations and measurement (e.g., visual observation, sensors/instruments, simulations, survey)

A Case of Applying the Epistemic Design Framework

Given the foregoing discussion, I now demonstrate a case of application with the design framework. First, I will describe the general background of a curriculum design project aimed to engage high school students in an authentic epidemiologic practice, and discuss epistemological

issues that emerged through the implementation. Then, I will discuss the potential solutions by applying the design framework.

Exploring Databases: Engaging Students in an Epidemiological Investigation Practice

Scientists increasingly use information communications technology (ICT) to analyze large datasets in contemporary science, and engaging students in database investigations has a great potential to promote their understanding in this line of contemporary practice. In this study, we developed “Exploring Databases (ExDa)”, a curriculum unit to engage high school students in an authentic practice of epidemiologic investigation of smoking behavior in collaboration with genome scientists and epidemiologists (Munn et al., 2013).

Epidemiology studies the occurrence and distribution of health-related events (e.g., diseases) and pursues developing population-based explanations for risk factors (Koepsell & Weiss, 2003). Risk factor is defined as a surrogate for deeper causes and better predictions (Stampfer et al., 2004), and it is a form of population-based explanation that account for which exposures (hypothesized risk factors) increase the risk (of contraction) in a study population. Nicotine addiction is the most common chemical dependence in the United States, and tobacco use still is the leading cause of illness and death in that country (U.S. Department of Health and Human Services, 2012). Smoking behavior is known as a multifactorial trait and a variety of genetic, physiological, and environmental factors may increase the risk of nicotine addiction.

Through the unit, students work in teams to engage in a database investigation to identify the risk factors of smoking, and develop and test their hypotheses by analyzing the dataset through a case control study process. Case control design is an epidemiologic method that samples “cases” who continued smoking and “controls” who did not continue smoking by matching similar backgrounds aimed to evaluate the associations in the degree of exposures

Chapter 1

between them (Koepsell & Weiss, 2003). The dataset used for student inquiry came from a previous project, and about 300 subjects answered a questionnaire consisting of nearly 100 items that ask a variety of questions related to environmental exposures (e.g., family smoking) and physiological exposures (e.g., feeling buzzed). In addition, study subjects also provided a small blood sample, which was used to genotype their DNA at three candidate gene regions suspected to be related to nicotine addiction. All of these data were stored in an online database and students could access it through a web browser.

As they engage in analysis, students explore the database and identify a few items relevant to their hypotheses, and develop statistical tests to estimate the odds ratios (OR) and 95% confidence intervals (CI) based on the database information. OR is used to evaluate the strength of association between exposure and smoking, and 95% CI is used to determine the statistical significance of the sample OR to the study population. Once some association is confirmed significant, students then apply causal reasoning to judge whether the association is causal based a set of criteria for judging causality drawing on the norms of the epidemiologic discipline (Bradford-Hill, 1965). With the statistical evidence and theoretical reasoning along with the criteria, they draw their own conclusions and present their findings to the whole classroom in a presentation or poster session.

Overview of the ExDa Unit

The first version of ExDa unit consisted of six 1- to 2- hour lessons (Table 2). Students develop their basic knowledge related to the practice through the first five lessons in a lecture format, and then engage in their own investigation in the last lesson. After the introductory lesson (Lesson 1), students develop their understanding in a scientific model with stages of smoking by reading smoking profiles that describe how people continued smoking or not (lesson

Chapter 1

2). In Lesson 3, they learn the mechanism of neurotransmission and a few genes suspected to be related to nicotine addiction. Through these two lessons, students develop their hypotheses for key exposures that might increase the risk of smoking.

In Lesson 4, students learn case control design. Students first develop understanding of its terminology and procedure in a scenario in which people became sick in a house party and the house manager investigates what caused it. They are encouraged to define which people are cases (people who got sick) and which people are controls (people who didn't get sick) and develop their hypotheses for the causes (e.g., food was bad, someone was sick). Then students are asked to think how they would investigate and develop their understanding in the procedure, followed by a few exercises with other scenarios. During the instruction, the teacher explains the difference between experimental design in the nature of procedure (e.g., retrospective vs. prospective) and evidence (correlational vs. causal).

In Lesson 5, they learn two statistical concepts—odds ratio (OR) and 95% confidence interval (CI)—in order to estimate the strength of association and determine the statistical significance in data analysis (lesson 5). Students first develop their understanding in the definition and formula and calculate ORs with a few exercises. They then learn the interpretations by the criterion for whether the value of one (null) is outside or inside the interval to determine the presence of association. During the instruction, it is emphasized that ORs indicate association and not necessarily causation, and students are introduced to the criteria for causality and engage in a few exercises with the criteria.

Table 2

The Overview of the "Exploring Databases" Unit

Lesson Index and Title	Lesson Content and Activity
------------------------	-----------------------------

1. Why and how do people do science?	• Introduction to the ExDa Unit
2. Why do some people become smokers?	• Reading smoking profiles • Learning the stages of smoking model • Mapping smoking profiles to the stages of smoking model • Developing initial hypotheses for risk factors of smoking
3. How do genes influence smoking behavior?	• Learning the mechanism of neurotransmission: dopamine receptor and <i>DRD2</i> codes • Demonstrating neurotransmission with the physical model
4. How can we study genetic and environmental influences on smoking behavior?	• Learning the terminology and procedures of case control design: outcome: cases and controls, exposure, and matching • Exploring the smoking database
5. Analysis of data in case control studies: the odds ratio	• Learning odds and odds ratio • Association and causation • The criteria for causality • 95% confidence interval
6. Database research: What can we learn from the smoking behavior data?	• Engaging in student-led investigations or hypothesis testing with the smoking database

Epistemological Conflation and Conflicts Observed

In the first stage of classroom implementation in multiple classrooms, we observed that students actively engaged in their own investigations, though we also observed some epistemologically problematic scenes. In one class, students ended up testing associations with ORs and CIs but did not try to apply causal reasoning to scrutinize whether those associations were causal or not. We later found that the teacher seemed not to have explicitly explained causal reasoning with the criteria for causality, because the main part is “data analysis” with the database. In another class, a group of students posed a question to the teacher as to whether it is possible to determine causality from case control design during their investigation. Even though all the teachers who implemented the unit had participated in our professional development

Chapter 1

workshop in advance, the teacher and students almost rejected causal reasoning with epidemiologic criteria since it is not “controlled” like with an experimental design. Then they became confused in making sense of why they had the criteria for causality. In a later professional development workshop with new teachers, some teachers appeared concerned in discussing causality from case control design and a few teachers even strongly rejected this form of causal reasoning based on their epistemic belief that it is not controlled and claimed that the “only” way to discuss causality is to conduct experiments.

Although these observations were only a few and not representative of the entire population of students and teachers who engaged in the ExDa unit, these events suggest that epistemological conflation and conflicts are likely to occur for some students and teachers given how science has been historically framed in K-12 schooling, and this motivated me to evaluate the structure of our instructional design. Along with the epistemic design framework, specifically, the first case indicates that students may not have understood the epistemic aim as identifying risk factors of smoking, not just associations. That conceptual distinction is crucial in this form of causal analysis. In the second and third cases, although their concern is valid in that risk factors do not determine the “absolute” causality, they would not have understood the epistemic aim, seeking “population-based” explanation does not ensure every individual becomes a smoker because of the risk factors. To these interpretations, I expect some counterarguments, such as; these students and teachers did not just understand experimental design itself rather than conflation and conflicts. However, this does not explain why they “rejected” the careful inquiry conducted in epidemiology, and thus I conjecture that epistemological conflation occurred due to the lack of their discipline-specific understanding in the epistemic aim and the nature of theory and method, and current materials and scaffolds may not have sufficiently reified these DSEs

during instruction. In other words, the high school students needed a similar grounding to the kind of empirical analyses customary to the discipline in a way that is parallel to post-secondary students being educated into how to conduct case control study research.

Developing Epistemic Narratives and Reifying Epistemic Alignments

Case control design (method) is often adopted in an early stage of research to identify risk factors from various environmental exposures through epidemiologic practices (Gordis, 2009). Epidemiologists often adopt observational studies due to ethical concerns in randomly assigning and exposing subjects to potentially harmful experiences or materials. In addition, they often start investigations after incidents occurred or diseases have been distributed (e.g., epidemic). Although observational studies basically produce correlational evidence and do not necessarily indicate causation, the goal of investigation is to develop causal theory (i.e., the epistemic aim) and thus they elaborate the study design (i.e., through matching) and employ analytical techniques (i.e., criteria for causality) to discriminate causation from association. OR is used to quantify and measure the strength of association, and CI is used to evaluate the statistical significance to the study population beyond the data, since the practice seeks population-based explanations. Finally, the criteria for causality are used to discriminate which association is causal rather than mere associations. Figure 2 below shows a representation of epistemic alignments based on the epistemic narratives.

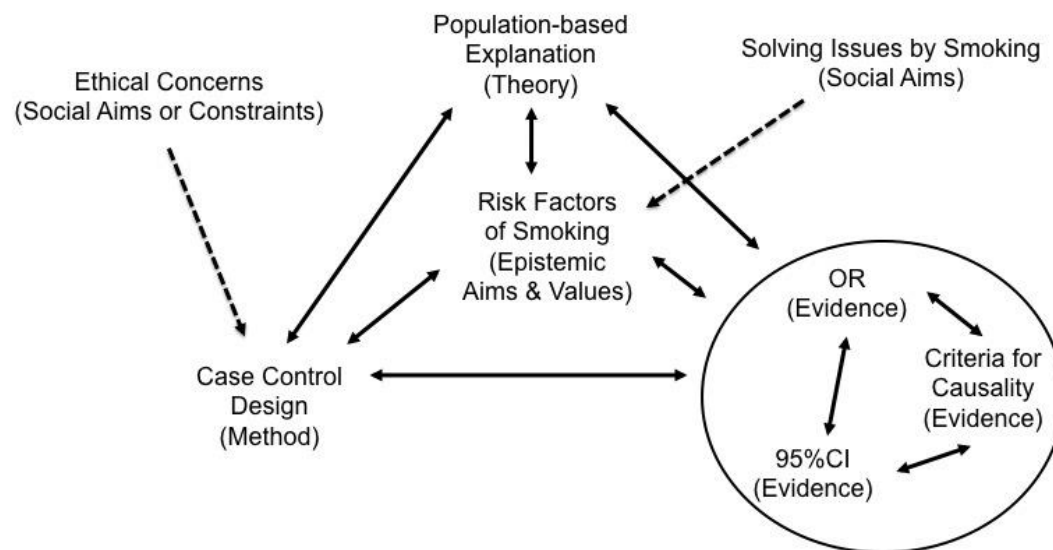


Figure 2. Epistemic alignments in the exploring databases unit

Specifically, what epistemic actors would it take to understand the rationale for adopting case control design? The narratives suggest several actors are aligned around the target actor. First, it is essential for students to understand the *epistemic aim* so as to identify risk factors of smoking. Without understanding this, it would be hard to make sense of why they need to pursue causal explanations based on correlational evidence. Second, some students may wonder why identifying risk factors matters so much and understanding the rationale requires students, in turn, to understand that the *social aim* is to keep people from starting to smoke through this population-level analysis given that smoking is the leading cause of illness and death in the country.

Third, it is also essential to recognize the nature of *theory* of risk factor as population-based, probabilistic explanation. A statement like “X is the risk factor of smoking” means people with Exposure X are more likely to become smokers, while not claiming everyone with X becomes smokers. Students without this understanding may misunderstand that X is not the risk

Chapter 1

factor because they know someone with X who does not smoke. This is partly due to the fact that most causal explanations dealt in school curricula are often mechanistic and the causal effect is rather definitive, for example, conservation of momentum and the law of physics more generally.

Fourth, students need to see why epidemiologists adopted case control design rather than experimental design. To make sense of this, it is essential to recognize social aims or *constraints* including ethics and other pragmatic reasons such that there are many smokers and non-smokers and we can learn something from observation instead of experimentation. It is relatively easy for students to understand why it is inappropriate to randomly expose subjects to a hypothesized risk factor. Rather, it would be key to understand the epistemic aim, identifying causal risk factors, even under various constraints, which would help students engage in complex methodological negotiations between the epistemic aim and social constraints.

Lastly, students need to recognize the nature of *evidence* in odds ratio as correlational evidence along with these methodological challenges. Students need to understand that OR indicates the association within the data, but the epistemic aim is to learn about the study population in relation to the nature of theory as population-based explanation. This leads to the analytical value of using 95% confidence intervals to evaluate the statistical significance of the sample OR to the study population. Together, understanding the nature of these analytical actors would help students understand why they need the criteria for causality to discriminate causation from association.

Evaluating Current Scaffolds in Contrast to the Epistemic Alignments

Given the epistemic narratives and alignments, how much do current instructional materials and scaffolds address DSEs of the epidemiologic practice? Contrasting the epistemic alignments above, I found three potential issues in the current instruction. First, current materials

Chapter 1

and scaffolds address the scientific model of smoking and its multifactorial nature, but it may not enough to explicitly address the epistemic aim as identifying *causal* risk factors and why the epistemic goal matters to students' life and society. Although some teachers mention these factors from their personal knowledge and experience with family during instruction, the current instruction may not have engaged students in any specific task to encourage them to apply reason to the epistemic aim and social aim that shapes the practice. Accordingly, some task could be added to the instruction such as showing an interview with public health researchers and engaging students in investigating and discussing why tobacco research matters to society.

Second, the current instruction engages students in developing an understanding of case control design in its terminology and procedure including ethical issues in randomization with human subjects, whereas it may not have explicitly engaged them in methodological negotiations between the epistemic aim and social constraints. Specifically, the unit engages students in a related task in which students compare the procedural difference from experimental design, but it does not explicitly engage students in making sense of why the study did not employ experimental design instead of case control design, and why they need to seek causal explanations along with the epistemic and social constraints. To provoke such reasoning, some tasks could be added to the current instruction such as engaging students in discussing why the study adopted case control design rather than experimental design or engaging them in designing a new study under a hypothetical scenario in which they have to consider ethical concerns yet need to seek causal explanations.

Lastly, instruction should more explicitly help students develop their understanding in statistical concepts along with the population-based nature of epidemiologic explanations. The current instruction provides related tasks to understand the concepts of odds ratio and 95%

confidence interval and how to interpret the numbers, but it may not enough explicitly engage students in developing their understanding along with the population-based and probabilistic nature of epidemiologic explanations. This issue might be related to the lack of scaffolding in conveying the epistemic aim of epidemiologic research. Some more explicit task might help students develop their discipline-specific knowledge along with the epistemic aim and the nature of theory and method so that students can develop their understanding of why they need to use 95% confidence intervals, instead of determining the presence of association only with odds ratios. To promote this understanding, a task could be added to the current instruction to engage students in discussing the difference between population-based and mechanistic explanations and why they need to use CIs in addition to ORs.

Conclusion

In this study, I discussed issues in epistemological learning and proposed a conceptual framework to reify discipline-specific epistemologies that can be used to guide the design of instruction. Literature suggests that students tend to develop unrealistic images of scientific practices from science instruction and the lack of epistemological understanding seems related to their limited reasoning in classroom inquiry. To improve epistemological instruction, a majority of researchers have worked to define discipline-general epistemologies (DGEs) and develop instructional strategies to develop their discipline-general knowledge. In contrast, recent studies suggest that such knowledge may not develop in a way that can be enacted when engaging in inquiry activities without developing epistemic cognition and discipline-specific knowledge through engaging in disciplinary contexts. I argued that more research in the latter approach is particularly needed given the ongoing science education reforms, and a key research agenda

would be to verify whether the development of epistemic cognition and discipline-specific knowledge leads to student improvements in epistemological performance and development.

Another key agenda is the activity design and scaffolding to develop their epistemic cognition as developing discipline-specific knowledge through authentic inquiry. However, there is little theoretical guidance to reify DSEs, and I discussed how to disentangle them drawing on science studies. Philosophical and sociological studies suggest that scientific disciplines in the nature of theory, method, and evidence are complex and intertwined with heterogeneous actors including social actors. Although it is still in debate whether to incorporate social actors into the discussion of the legitimacy of scientific practices (Laudan, 1984; Longino, 1992; Pickering, 1995; Rouse, 1996; Knorr-Cetina, 1999), I follow Rouse (1996) and argue the value of framing scientific practices in relation to epistemic alignments and using epistemic narratives to reify the formations.

From this perspective, I propose an epistemic design framework to identify discipline-specific epistemologies and reflect them into instructional materials and scaffolds. The conceptual framework aims to guide researchers to reify DSEs for their focal practice and identify the epistemic alignments through developing epistemic narratives. Specifically, it encourages researchers to: (a) identify key epistemic actors that represent epistemic characteristics of the focal practice, (b) reify epistemic alignments through developing epistemic narratives along with the key actor(s) in the previous task, and (c) organize the alignments into a set of sub-alignments that can be used to shape classroom tasks so that students can develop their epistemic cognition and discipline-specific knowledge with related conceptual tools. The preceding section demonstrated an application and the framework was used to guide potential solutions to epistemological issues in instruction. I argue that it was particularly helpful in

Chapter 1

evaluating whether the current materials and scaffolds make the epistemic alignments among relevant actors explicit enough or not, which helped thinking of instructional strategies to solve emergent issues in student learning.

As final remarks, I argue that more research is needed on the theory of *epistemic sense making*. I define epistemic sense making by students' metacognitive reasoning epistemic alignments—why scientists conduct their investigations using available analytical tools in the currently engaged practice. Students tend to pay more attention to “what to do” and less attention to “why they should do something” or “when they should do something” as they engage in instruction, and these additional dimensions of metacognitive reasoning is essential to activate their epistemic cognition and discipline-specific knowledge in a way that they can discriminate one discipline from another. Science curricula are normally designed without a specific consideration of multiple disciplines (Stevens et al., 2005), and students without epistemological awareness are more likely to conflate one discipline with another. Although epistemological conflation might not be problematic in some cases, it could lead to epistemological conflicts when the current discipline conflicts with some other from their prior learning. Or it might lead students to develop an overly narrow view of the heterogeneous epistemic practices found in the sciences. To avoid and overcome such conflict, science instruction should provide specific guidance and scaffolding to help students develop not only the disciplinary knowledge, but also “why” it is valid by incorporating heterogeneous resources, if necessary, in relation to other (dominant) disciplines.

As limitations, although the framework identifies a few key epistemic actors as shown in Table 1, there may be other actors that help reify DSEs, and the list of these actors may vary across practices and disciplines, for example, between physical, biological, and earth/space

sciences. To develop the list and verify the applicability of the framework, more research is needed through applying the framework into multiple practices and the design. In addition, the current framework aims to help researchers develop epistemic alignments through narrative development and use the reified alignments for instructional design, but it still needs improvements in guiding how to translate the alignments into specific scaffolds. There are numerous studies on scaffolding cognitive tasks such as visualization (e.g., Linn et al., 2004) and argumentation (e.g., Bell, 2004) and these findings would be useful for the translation into authentic analytical tasks. However, there are still few studies focusing on students' sense making of a discipline in relation to others with heterogeneous actors such as methodological negotiations. Thus, this line of epistemic scaffolding design will also be a key agenda in future works. If science instruction is able to support discipline-specific epistemic sense-making and learning, then students will become more scientifically literate so that they are able to deeply understand and appreciate various scientific endeavors from a broader viewpoint in relation to their personal and possibly professional interests and goals.

References

- Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The Nature of Science and Instructional Practice: Making the Unnatural Natural. *Science Education*, 82, 4, 417-36.
- Alters, B. J. (1997). Whose Nature of Science? *Journal of Research in Science Teaching*, 34, 1, 39.
- Bell, P., & Linn, M. C. (2000). Scientific Arguments as Learning Artifacts: Designing for Learning from the Web with KIE. *International Journal of Science Education*, 22, 8, 797-817.

Chapter 1

- Bell, P., & Linn, M. C. (2002). Beliefs about science: How does science instruction contribute? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing* (pp. 321-346). Mahwah, NJ: Lawrence Erlbaum.
- Bell, P. (2004). Promoting Students' Argument Construction and Collaborative Debate in the Science Classroom, In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education*. Erlbaum, Mahwah, NJ, 115-143.
- Bradford-Hill, A. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58: 295–300.
- Cartwright, N. (2004). Causation: One Word, Many Things. *Philosophy of Science*, 71, 5, 805-819.
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically Authentic Inquiry in Schools: A Theoretical Framework for Evaluating Inquiry Tasks. *Science Education*, 86, 2, 175-218.
- Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the Dimensions of Epistemic Cognition: Arguments from Philosophy and Psychology. *Educational Psychologist*, 46, 3, 141-167.
- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50, 1.
- Dhingra, K. (2003). Thinking about Television Science: How Students Understand the Nature of Science from Different Program Genres. *Journal of Research in Science Teaching*, 40, 2, 234-56.
- Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham: Open University Press.

Chapter 1

- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the Norms of Scientific Argumentation in Classrooms. *Science Education*, 84, 3.
- Elby, A., & Hammer, D. (2001). On the Substance of a Sophisticated Epistemology. *Science Education*, 85, 5, 554-67.
- Gordis, L. (2009). *Epidemiology*. Philadelphia: Elsevier/Saunders.
- Grotzer, T. A. (2003). Learning to Understand the Forms of Causality Implicit in Scientifically Accepted Explanations. *Studies in Science Education*, 39, 1, 1-74.
- Grotzer, T. A., & Basca, B. B. (2003). How Does Grasping the Underlying Causal Structures of Ecosystems Impact Students' Understanding? *Journal of Biological Education*, 38, 1, 16-29.
- Grotzer, T.A. (2012). *Learning causality in a complex world: Understandings of consequence*. Lanham, MD: Rowman Littlefield.
- Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge, U.K: Cambridge University Press.
- Herrenkohl, L. R., & Cornelius, L. (2013). Investigating Elementary Students' Scientific and Historical Argumentation. *Journal of the Learning Sciences*, 22, 3, 413-461.
- Hodson, D. (2009). *Teaching and learning about science: Language, theories, methods, history, traditions and values*. Rotterdam: Sense Publishers.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67, 1.

Chapter 1

- King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco, CA: Jossey-Bass.
- Kitchener, R. F. (2002). Folk epistemology: An introduction. *New Ideas in Psychology*, 20, 89-105.
- Kitchner, K. S. (1983). Cognition, Metacognition, and Epistemic Cognition. *Human Development*, 26, 4, 222-232.
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, Mass: Harvard University Press.
- Koepsell, T. D., Weiss, N. S. (2003). *Epidemiologic Methods: Studying the Occurrence of Illness*, Oxford University Press.
- Krajcik, J. S. & Blumenfeld, P. C. (2006). Project-based Learning, *The Cambridge Handbook of the Learning Sciences*. Cambridge University Press.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, Mass: Harvard University Press.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.
- Laudan, L. (1984). *Science and values: The aims of science and their role in scientific debate*. Berkeley: University of California Press.
- Lederman, N.G. (2007). Nature of science: Past, present, and future. In S.K. Abell, & N.G. Lederman, (Editors), *Handbook of research in science education* (pp 831-879). Mahwah, New Jersey: Lawrence Erlbaum Publishers.

Chapter 1

Linn, M. C., Davis, E. A., & Bell, P. (2004). *Internet environments for science education*.

Mahwah, N.J: Lawrence Erlbaum Associates.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*.

Princeton, N.J: Princeton University Press.

Longino, H. E. (2002). *The fate of knowledge*. Princeton, N.J: Princeton University Press.

Louca, L., Elby, A., Hammer, D., & Kagey, T. (2004). Epistemological Resources: Applying a New Epistemological Framework to Science Instruction. *Educational Psychologist*, 39, 1, 57-68.

Mayr, E. (1988). *Toward a new philosophy of biology: Observations of an evolutionist*.

Cambridge, Mass: Belknap Press of Harvard University Press.

Mayr, E. (2004). *What makes biology unique?: Considerations on the autonomy of a scientific discipline*. New York: Cambridge University Press.

Munn, M. M., Oura, H., Gallivan, M., Van, H. K., & Shouse, A. W. (2013). Engaging high school students in research on smoking behavior. *Science*, 341, 6144, 360-1.

National Science Board (U.S.). (2010). *Science & engineering indicators*. Washington, D.C: National Science Board.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C.: The National Academies Press.

National Research Council. (2013). *Next Generation Science Standards: For States, By States*. Washington, D.C.: The National Academies Press.

Chapter 1

- Nicolaidou, I., Kyza, E. A., Terzian, F., Hadjichambis, A., & Kafouris, D. (2011). A Framework for Scaffolding Students' Assessment of the Credibility of Evidence. *Journal of Research in Science Teaching*, 48, 7, 711-744.
- Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "Ideas-about-Science" Should Be Taught in School Science? A Delphi Study of the Expert Community. *Journal of Research in Science Teaching*, 40, 7, 692-720.
- Peker, D., & Wallace, C. S. (2011). Characterizing High School Students' Written Explanations in Biology Laboratories. *Research in Science Education*, 41, 2, 169-191.
- Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago: University of Chicago Press.
- Resnick, M. (1996). Beyond the Centralized Mindset. *Journal of the Learning Sciences*, 5, 1, 1-22.
- Rouse, J. (1996). *Engaging science: How to understand its practices philosophically*. Ithaca: Cornell University Press.
- Rudolph, J. L. (2003). Portraying Epistemology: School Science in Historical Context. *Science Education*, 87, 1, 64-79.
- Samarapungavan, A., Westby, E. L., & Bodner, G. M. (2006). Contextual Epistemic Development in Science: A Comparison of Chemistry Students and Research Chemists. *Science Education*, 90, 3.
- Sandoval, W. (2003). Conceptual and Epistemic Aspects of Students' Scientific Explanations. *Journal of the Learning Sciences*, 12, 1, 5-51.
- Sandoval, W. (2005). Understanding Students' Practical Epistemologies and Their Influence on Learning through Inquiry. *Science Education*, 89, 4.

Chapter 1

- Sandoval, W. (2014). Science Education's Need for a Theory of Epistemological Development. *Science Education*, 98, 3, 383-387.
- Schwab, J. J., Westbury, I., & Wilkof, N. J. (1978). *Science, curriculum, and liberal education: Selected essays*. Chicago: University of Chicago Press.
- Schwartz, R., & Lederman, N. (2008). What Scientists Say: Scientists' Views of Nature of Science and Relation to Science Context. *International Journal of Science Education*, 30, 6, 727-771.
- Smith, M. U., Lederman, N. G., Bell, R. L., McComas, W. F., & Clough, M. P. (1997). How Great is the Disagreement about the Nature of Science? A Response to Alters. *Journal of Research in Science Teaching*, 34, 10, 1101-1104.
- Songer, N. B., & Linn, M. C. (1991). How Do Students' Views of Science Influence Knowledge Integration?. *Journal of Research in Science Teaching*, 28, 9, 761-84.
- Stampfer, M. J., Ridker, P. M., & Dzau, V. J. (2004). Risk factor criteria. *Circulation*, 109, 25, 3-5.
- Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005). Comparative Understanding of School Subjects: Past, Present, and Future. *Review of Educational Research*, 75, 2, 125-157.
- Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, D.C: Brookings Institution Press.
- Talanquer, V. (2010). Exploring Dominant Types of Explanations Built by General Chemistry Students. *International Journal of Science Education*, 32, 18, 2393-2412.
- U.S. Department of Health and Human Services (2012). *Preventing Tobacco Use among Youth and Young Adults: A Report of the Surgeon General*. Atlanta, GA: U.S. Department of

Chapter 1

Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2012.

van Dijk, E. M. (2011). Portraying Real Science in Science Communication. *Science Education*, 95, 6, 1086-1100.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the Scientific Method: Model-Based Inquiry as a New Paradigm of Preference for School Science Investigations. *Science Education*, 92, 5, 941-967.

Wong, S. L., & Hodson, D. (2008). From the horse's mouth: What scientists say about scientific investigation and scientific knowledge. *Science Education*, 93, 1, 109-130.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

**Chapter 2. Eliciting Students' Epistemic Sense Making: A Case Study of a Value-
Focused Approach**

Hiroki Oura

College of Education, University of Washington

Abstract

The latest K-12 science education framework encourages science learning through engaging in authentic tasks involving scientific practices. Scientific communities apply various analytical methods and tools to identify evidence and develop explanations, whereas making sense of the rationale of scientific disciplines is a contentious topic due to the incommensurability of their epistemic foundations and assumptions. Analogous issues would also occur in the science classroom where it engages students in causal reasoning and judgment through authentic inquiry in different disciplines (Bell & Linn, 2002). However, few studies discuss the theoretical specification and ways to elicit student discipline-specific understanding, and therefore some theoretical and pedagogical guidance is needed. In this study, I first develop a theoretical discussion on the specification of discipline-specific knowledge and propose an approach to elicit student discipline-specific knowledge in scientific practices, followed by a case study in which the framework was applied to assess student discipline-specific knowledge in an epidemiologic inquiry. The findings indicate that the framework served to gauge students' disciplinary understanding and identify revision points in the instruction.

A major goal of science education is to develop students' understanding of how to engage in scientific practices (NRC, 2012). Over the last decades, researchers have reported students' unrealistic images of scientific work and limited reasoning in analytical tasks (Driver et al., 1996; Grotzer, 2012), and ascribed it to the inauthenticity of science instruction and classroom inquiry to real epistemic practices (White & Frederiksen, 1998; Chinn & Malhotra, 2002). For epistemological development, many researchers have worked to define discipline-general epistemologies arguably shared by all or most scientific practices and reflect or teach such discipline-general knowledge (DGK) by making those epistemological elements explicit in instruction (Abd-El-Khalick, Bell, & Lederman, 1998; Osborne et al., 2003; Hodson, 2009). Some researchers advocate developing explicit DGK by discriminating it from epistemic skills (Lederman, 2007), whereas other researchers question the representativeness of such generalization across different practices (Alter, 1997; Rudolph, 2003), and some doubt the educational productivity in teaching general knowledge separated from disciplinary contexts (Elby & Hammer, 2001). These researchers argue the significant role of discipline-specific aspects of scientific epistemologies (Samarapungavan, Westby, & Bodner, 2006), which I call discipline-specific knowledge (DSK) hereafter, and they emphasize epistemological development in DSK by engaging students in disciplinary contexts of authentic practices.

Although the debate on the definition and legitimacy of DGK is still continuing (Elby & Hammer, 2001; Rudolph, 2003; Sandoval, 2005; Wong & Hodson, 2008; Schwartz & Lederman, 2008; Allchin, 2011; Schwartz, Lederman, & Abd-El-Khalick, 2012), I argued elsewhere that there is an urgent need of more research on the DSK approach particularly under the current science education reforms that encourage science learning through disciplinary engagement in authentic practices including planning and carrying out investigations, analyzing and interpreting

data, constructing explanations, and engaging in argument from evidence (NRC, 2012).

Scientific disciplines are organized around specific sets of epistemic tools and norms, and they are developed and internally shared by scientific communities in the course of their own history of practice (Schwab, Westbury, & Wilkof, 1978). On the other hand, these tools and norms vary across practices, and understanding each discipline requires recognizing heterogeneous epistemic actors and alignments that shape one discipline differently from another (Laudan, 1984; Longino, 2002; Rouse, 1996; Knorr-Cetina, 1999). These actors and alignments underpin the rationale for particular analytical methods and criteria that are adopted in each practice, and thus it is crucial for students to make sense of the rationale as well as reflecting them into curriculum and instruction (Chinn & Malhotra, 2002; Samarapungavan et al., 2006).

For another rationale, especially in school settings, students often study multiple science subjects and are involuntarily put into situations in which they need to adjust their epistemological mode between subjects (Stevens et al., 2005). Some disciplines like physical and biological sciences involve substantial differences in disciplinary norms and criteria for causal reasoning and judgments, and they do not necessarily agree with each other (Woodward, 2003; Cartwright, 2004; Mayr, 2004). Therefore, students without discipline-specific awareness more likely conflate one discipline with another from their prior learning or naïve assumptions (Hofer & Pintrich, 1997). Although it may not be problematic if two disciplinary norms and criteria are consistent, if they are not consistent students become unable to resolve epistemological inconsistencies by themselves. In a worst case scenario, they could reject the rationale of one discipline based on the other without making an effort to understand the currently engaged practice. I call such situations *epistemological conflation and conflicts* (ECCs), and suspect that analogous issues would emerge in authentic-practice-based curricula associated with science

education reforms. In this respect, I argue that epistemological knowledge should be framed as DSK rather than DGK, and more research is needed to develop theoretical guidance of developing students' DSK through engaging in authentic practices (see Chapter 1).

Working from this perspective, important research questions arise regarding the theoretical specification and observational methods of one's disciplinary understanding. How can we elicit and observe students' discipline-specific understanding, and on what account can we determine that students have developed DSK for a focal practice/discipline? Although researchers argue the diversity of scientific disciplines, there are few studies to reify theoretical guidance to frame how to teach such discipline-specific epistemologies and methods to observe students' DSK learning in a scientific discipline. To answer these questions I first develop a brief theoretical discussion on the specification of disciplinary understanding from my previous work (Chapter 1). Next, I propose a value-focused approach as a way to elicit students' discipline-specific understanding with a set of open-ended questions. Then, I report the results of a case study in which the value-focused approach was adopted to assess students' discipline-specific understanding in a curriculum unit in which students engage in epidemiological inquiry. Finally, theoretical and methodological issues are discussed for future research.

Theoretical Framework: Framing Scientific Disciplines as Epistemic Alignments

Scientific practices share some general forms in knowledge development and argumentation; they also involve discipline-specific aspects in which epistemic norms and criteria differ across practices due to the significant differences in their epistemic foundations and assumptions (Laudan, 1984). Taking an example of the explanatory form of scientific knowledge, some physical sciences pursue mechanistic explanations that “reduce” target phenomena into underlying components (or variables), whereas some life sciences pursue

population-based explanations that characterize a group individuals based on statistical accounts (Mayr, 2004).

Methodology and disciplinary criteria for causal reasoning and judgments also differ across practices (Cartwright, 2004). For example, some practices adopt experimental design to identify causal evidence, whereas other practices adopt observational design to identify correlational evidence by which causal reasoning is applied. Generally, the former approach may be ideal in seeking causal explanations, but the latter is also valued in certain contexts where experimental design is difficult to implement due to ethical and/or various pragmatic reasons (Longino, 2002; Knorr-Cetina, 1999). In such context, they apply various analytical techniques (e.g., theoretical reasoning, study design, statistical control) to develop causal explanations with non-experimental methods. In sum, the nature of theory, method, and evidence is not necessarily identical across disciplinary practices and the sense making based on the specific rationale requires incorporating heterogeneous actors into a formation based on epistemic aims and values (Laudan, 1984).

Another complexity in the sense making of disciplinary norms is that they may be shaped not only by those material and cognitive actors, but also by natural and social actors, which are often considered as premise or background information (Longino, 2002). As indicated above, observational design is often chosen due to ethical concerns and pragmatic reasons, and ethics can be viewed as a social constraint from an analytical viewpoint (in other words, such value reflects our social aim or responsibility). While there are some practices solely based on theoretical interests, contemporary research is often conducted, directly or indirectly, to solve societal issues (Stokes, 1997). In sum, social actors may shape scientific practices when

applicable, and they are essential resources to make sense of the rationale and validity of theory, method, and evidence in a discipline (Longino, 2002).

To account for such heterogeneous formations of scientific disciplines, Rouse (1996) argues that scientific practices should be conceived as *epistemic alignments* that form with heterogeneous actors including cognitive, material, social, and cultural ones. Following his argument, I elsewhere propose an epistemic design framework aimed to guide researchers and curriculum developers to identify epistemic actors and alignments of a focal practice (see Chapter 1 for details). Specifically, the framework aims to guide researchers to: (a) identify key epistemic actors that represent epistemic characteristics of a focal practice (e.g., theory, method, evidence, epistemic aims and values, social aims and values), (b) identify epistemic alignments through developing epistemic narratives (see Figure 1 below), and (c) organize these alignments into a set of sub-alignments as they constitute classroom tasks.

For example, a sub-alignment of epistemic aims and social aims would help students develop an understanding in the relationship between societal and theoretical interests. Another sub-alignment in methodological choices with the nature of subject matter, epistemic aims, observational/measurement challenges, and social constraints would help students make sense of methodological choices and procedures in the tension between epistemic interests and social constraints, which is particularly critical to avoid ECCs. In addition, the sub-alignment of observational strategies and measurement tools along with the epistemic aims and methodology would also be important to help students develop their understanding in the analytical aims of each strategy and tool as part of evidence identification.

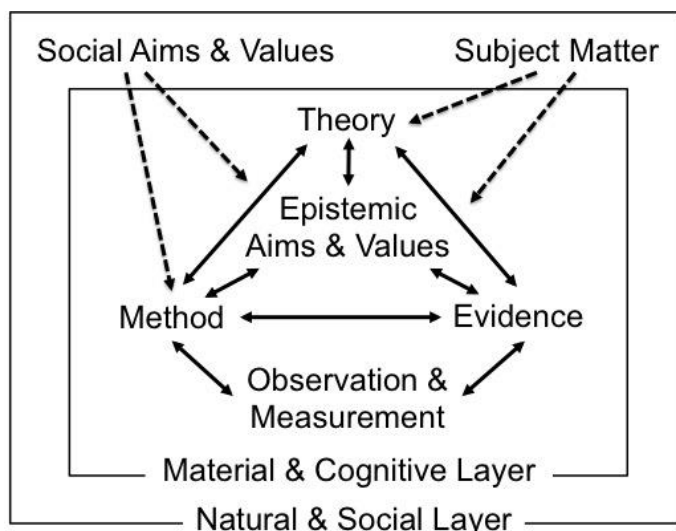


Figure 1. Epistemic actors and alignments in the material/cognitive and natural/social layers (see Chapter 1)

Value Questions: Eliciting Students' Epistemic Sense Making by Asking the Value of Epistemic Actors

In this theoretical perspective, how can we elicit students' understanding of epistemic alignments? In the present study, I approach this challenge by asking students why a particular epistemic actor(s) matters to the currently practice and eliciting student understanding in the value of the target actor(s). In a psychological sense, *value* can be seen as an emergent feeling property that emerges in mean-end relationships (Dewey, 1939). For example, a value of experimental design (a means) can be stated to estimate main effects while controlling for other effects including the unobserved through random assignment to conditions (an end). Analytical norms in scientific disciplines are organized to achieve their epistemic aims, and therefore asking "why" questions for the values of focal actors is a useful way to elicit student understanding of epistemic alignments.

Another rationale for this value-focused approach is that answering value questions requires incorporating heterogeneous actors when needed, which helps evaluate the heterogeneity of epistemic alignments in student understanding. For example, observational design (a means) is valued to identify correlational evidence (an end) when human subjects are involved in some other practice, since it could be unethical or pragmatically difficult to apply a randomized design (ethics). In such situation, scientists apply various analytical techniques (e.g., study design, data analysis) (means) to develop causal explanations with non-experimental methods (an end). Although the former example in experimental design involved only cognitive and analytical actors (e.g., confounding), this example requires incorporating ethics, a social actor, in addition to analytical actors to make sense of the value of observational design.

Together, this approach theoretically frames disciplinary understanding in these nets of epistemic alignments and aims to gauge student understanding in DSK by eliciting the cognitive structure with value questions. Note that epistemic alignments could vary across practices, and there is no unified framework that covers all types of epistemic alignments. Therefore, researchers need to identify key alignments for their focal practice along with instructional goals, and thus it would be more useful to theorize the guidance or procedures to develop those value questions. In this perspective, I identify the following three steps to develop value questions.

Identifying target actors. Before item development, researchers need to first identify key epistemic actors and alignments that characterize their focal practice. Theoretically, a variety of value questions can be composed to almost infinite combinations of epistemic actors and alignments, and therefore these questions should be limited only to those that represent epistemic characteristics of the focal practice and that are distinctive from typical practices in school science. Epistemic aim is a good example for the first criterion, since epistemic aims often are

discipline-specific and it would be informative to elicit student understanding in the current practice, in relation to social aims.

Another example would be methodology. Methodological choice is often contentious among professional scientists both within and across practices, and it is helpful to support student understanding of scientists' rationale along with the adopted analytical strategies and measurement tools that are distinctive from other practices. In particular, experimental design is dominant in standard science curricula (Windschitl, Thompson, & Braaten, 2008), and therefore value questions on observational design would be essential to evaluate and avoid the risk of ECC. Table 1 shows template actors and alignments, and may help researchers identify key actors and alignments built on the templates for their focal practice.

Table 1

Templates of Key Epistemic Alignments and Value Questions

Epistemic sub-alignment	Template value question
Epistemic aim/value	• What do scientists try to figure out?
Subject matter	• Why/How does the epistemic aim matter to scientists in the focal practice?
Social aim	• Why/How does the epistemic aim matter to our society?
Theory	• What is the nature of theory (compared to other scientific practices)?
Epistemic aim	• How does the theory serve to achieve the epistemic aim?
Method	• How do scientists try to answer their question?
Method	• What is the nature of method (compared to other practices)?
Epistemic aim	• What/How do scientists observe/measure to answer their question?
Observation/Measurement	• What social and/or pragmatic constraints do they need to cope with?
Social constraint	
Evidence	• What is the nature (criteria) of evidence (compared to other practices)?
Theory	• What is the analytical aim and limitation of each observation/measurement?
Observation/Measurement	
Analytical constraint	• How do scientists incorporate the results to develop explanations?

Developing value questions. In item development, value questions should be composed to elicit “heterogeneous” actors and alignments in student understanding. Epistemic aims are

often formed in relation to social aims, and it is useful to reveal student understanding of the relationship along with the subject matter. For example, the goal of epidemiology is to identify causes of health-related events (e.g., lung cancer) and prevent them from occurring to the degree possible. Asking questions such as “Why do epidemiologists want to identify the causes of lung cancer?” and “Why does identifying the causes of lung cancer matter to us?” would help reveal whether students understand the epistemic aims along with the social aims.

Likewise, methodological choices are often made in relation to both cognitive (or analytical) and social actors (e.g., ethical and pragmatic reasons), and therefore a value question is useful to reveal whether students have made sense of “why” a particular method is adopted in the currently engaged practice, in relation to relevant social actors. For example, such a value question like “why do epidemiologists adopt case control design rather than experimental design?” would help assess whether students incorporate ethics and pragmatic reasons into their justification statement. In these ways, it is possible to elicit students’ epistemic understanding in discipline-specific knowledge of a focal practice.

Criteria for assessment. Accordingly, the criteria for assessment are formed so as to gauge whether students (a) incorporate relevant actors to answer the value of a target actor(s) and (b) provide a reasonable justificatory statement in the relationship among the actors, including (c) the heterogeneity of epistemic alignments in their understanding. By appropriately forming “why” questions to target actors and alignments as discussed above, students would need to incorporate relevant actors and explain the alignments, and researchers could use student responses in the aim of epistemic assessment. Moreover, this assessment would also help researchers evaluate their curriculum design and identify which part of instruction should be

revised to help develop students' discipline-specific understanding in their focal practice as discussed in Chapter 1.

A Case Study of an Authentic Inquiry Unit Using the Value-Focused Approach

Rapid advances in information and communication technology (ICT) have enabled scientists to work with large-scale datasets in contemporary research. Many scientists now analyze large-scale data with computational methods including secondary data analysis with existing public databases accessible through the Web. From an educational viewpoint, this has the potential to develop low-cost and scalable curricula that engage students in large-scale data analysis and provide them with authentic experience as a form of contemporary research. As background, we developed a curriculum unit *Exploring Databases* (ExDa) aimed to engage high-school students in epidemiologic analysis of smoking behavior using a real database in collaboration with genome scientists and epidemiologists (Munn et al., 2013).

Nicotine addiction is known as the leading cause of preventable illness and deaths in the United States (U.S. Department of Health and Human Services, 2012). Smoking behavior is a “multifactorial trait” in which people start and continue smoking due to a variety of causes including genetic, physiological, and environmental factors. Identifying these factors is valuable to prevent citizens from starting and continuing smoking in the social aim of public health, and this topic is particularly relevant to high-school students given the fact that many smokers start smoking as youth.

In the ExDa unit, students use an online database and investigate the causes of smoking in the explanatory form of *risk factors*. Risk factor is defined as “surrogates for deeper causes and better predictions” (Stampfer et al., 2004), and it has two distinctive characteristics from typical casual forms dealt with in school science. First, risk factor is a population-based

Chapter 2

explanation (Rothman, 2002), and a causal statement like “X is a risk factor of smoking” means people with X are (x times) more likely to continue (or start) smoking in a study population, but it does not necessarily mean everyone with X continues smoking. Second, risk factor implies both correlation and causation due to (potential) measurement errors by methodological and social constraints in investigating populations. In epidemiology, such relationships are described as *associations*, and epidemiologists discriminate causal evidence from associations with various analytical techniques including study design (e.g., matching), statistical control, and theoretical reasoning along with disciplinary criteria for causality and literature (Koepsell & Weiss, 2003). These characteristics substantially differ from causal forms in school science, such as mechanistic explanations, in which cause and effect is rather definitive, and experimental methods that directly produce causal evidence. Accordingly, sense making of the disciplinary norms requires discipline-specific understanding or DSK along with these analytical actors in relation to the social aims and constraints of smoking research.

The database students used contains survey responses of about 300 adult smokers and nonsmokers from a case control study (Munn et al., 2011), students in teams analyzed the statistical data based on their own hypotheses. The study subjects, both cases who smoke, and controls who tried but did not continue, in the study population from a region, answered a questionnaire consisting of nearly 100 items regarding various physiological (e.g., feeling buzzed) and environmental factors (e.g., parental smoking) suspected to be associated with smoking behavior. Subjects also provided a small blood sample that was used to genotype their DNA at three candidates’ gene regions, whose data is also stored and can be queried in the database.

In analysis, students calculate the strength of association between exposures (hypothesized factors) and smoking with odds ratios (ORs) and evaluate the statistical significance with 95% confidence intervals (CIs). When a significant association is confirmed, they apply criteria for causality in the epidemiologic discipline (Bradford-Hill, 1965) and make the causal judgment (i.e., whether the association is causal or not) considering multiple sources of bias including selection bias and confounding. In short, these analytical tasks substantially differ from typical inquiry tasks in school science (e.g., “fair test” experiments), and students need to make sense of not only these analytical norms but also the rationale of these norms to adjust their epistemological mode to the epidemiologic discipline.

The project team implemented the unit in about ten high-school biology classes with teachers who had participated in our professional development workshop to develop their conceptual knowledge and instructional materials of the unit (Munn et al., 2013). Despite teachers’ and students’ active engagement in the workshops and classrooms, we observed a few problematic cases in which some students and teachers almost rejected the epidemiologic discipline particularly in causal reasoning and judgment that uses correlational evidence (Lee et al., 2012). This indicates that ECC may have occurred to these teachers and students, since they never would have been exposed to such population-based explanation and causal reasoning with correlational evidence (Grotzer, 2012).

In this background, the project team revised instructional activities and materials with an emphasis on developing students’ DSK of the epidemiologic discipline. For the present study, specifically, we (a) added a few video clips in the introductory lesson in which a public health researcher introduced effects of smoking and contributions of smoking research including the applications into policy making, (b) revised the instruction in case control design and causal

reasoning with the epidemiologic criteria to encourage students to make sense of observational design rather than experimental in relation to ethical concerns and other pragmatic reasons, and (c) revised statistical instruction in odds ratio calculation and CI interpretation to encourage students to develop their understanding of the epistemic aim (population-based explanation) and statistical inference (inferring characteristics of population from sample).

Method

The goals of this case study were to assess how well students made sense of the epidemiological research methods given the revised materials and scaffolds available through the curriculum unit and which part of instruction required further revision. In this interest, five instructional goals were set to develop students' DSK (Table 2) and corresponding assessment items were developed in the value-focused approach.

Table 2

Target Actors Epistemic Criteria, and Relevant Instruction/Materials in the ExDa Unit

Target actor and epistemic criteria	Relevant instruction and materials
DSK-1: Smoking research	
1-a) Problems caused by smoking	• Watching an interview with a public health/tobacco researcher in which she introduces the effects of smoking and contributions of tobacco research
1-b) Contributions of smoking research	
DSK-2: Case control design	
2-a) Observation in real world settings	• Watching an interview with an epidemiologist in which he introduces epistemic aims and methodological approaches in epidemiologic research • Instruction and exercises on the concepts and procedures of case control design and the nature of observational methods compared to experimental design • Instruction and exercise on the difference between association and causation and ethical concerns of randomization in human subject research
2-b) Uncontrollability of variables	
2-c) Ethical concerns	
2-d) Association, not necessarily causation	
DSK-3: Criteria for causality	

Chapter 2

3-a) Discriminating causation from association:	• Instruction and exercises on epidemiologic criteria for causality in the epidemiologic discipline
3-b) Epidemiologic criteria	• Instruction and exercises on systematic biases (e.g., selection bias)
3-c) Limitations of OR and 95% CI in causal inference	• Engaging in a student-led inquiry project in which students apply the criteria
DSK-4: Odds ratio	
4-a) Quantifying the strength of association	• Instruction and exercises on the analytical aim, calculation, and interpretation of odds ratio
4-b) Criteria for OR interpretation	• Engaging in a student-led inquiry project in which students apply ORs
4-c) Association, not necessarily causation	
DSK-5: 95% confidence interval	
5-a) Generalizability to the study population	• Instruction on the difference between population and sample and the analytical aim of statistical inference in relation to odds ratio
5-b) Meaning of 95% confidence	
5-c) Criteria for statistical significance	• Instruction and exercises on the analytical aim and interpretation of 95% confidence intervals
	• Engaging in a student-led inquiry project in which students apply 95% CIs

Twenty-one high school students in an AP Biology class at a suburban public high school in the Pacific Northwest participated in this study. The biology teacher had a background in public health and participated in our professional development workshop multiple times before taking part in the study. Data sources included: (a) video- and audio-recordings of teacher lectures and student conversations during their inquiry project (13 classes x 50 minutes), (b) student worksheets collected after the implementation ($N = 14$), and (c) pre- and post-survey responses to assess students' conceptual knowledge (CK) and discipline-specific epistemic knowledge (DSK) in the unit ($N = 20$). Video- and audio-recordings were analyzed to confirm that relevant instruction and materials were provided to develop students' DSK of the epidemiologic discipline and identify key moments of student conversation linked to their epistemological development. All the foundational lessons were videotaped from the back of the

Chapter 2

classroom to record teacher instruction, and four student groups were video- or audio- recorded during the student-led inquiry project.

In addition, 18 student worksheets were collected after the implementation and 14 decipherable ones were analyzed to evaluate the degree of students' cognitive engagement during these lessons. Worksheets contained various items to engage students in note taking and exercises on key conceptual knowledge including case control design and odds ratio calculation as well as open-ended questions on DSK of the epidemiologic inquiry. Worksheet analysis focused on these epistemological items and counted the counts of students who wrote down explicit answers to DSK criteria as an indicator of their cognitive engagement in epistemological learning (see Table 3). Specifically, student notes were counted when they included any full or partial sentence(s) including keywords that corresponded to DSK criteria, whereas they were not counted when the descriptions were not provided, were erroneous, or unclear.

Before and after the unit, students answered the same pre- and post-surveys that asked conceptual knowledge of the unit and DSK of the epidemiologic discipline. Conceptual knowledge items were developed to assess students' conceptual knowledge in case control design, causal reasoning with the epidemiologic criteria in a multiple-choice format, as well as their skills in OR calculation and 95% CI interpretations. In discipline-specific epistemic knowledge, five open-ended items were composed in the value-focused approach to address the social value of smoking research (DSK-1); advantages (values) and limitations of case control design (DSK-2); analytical value of causal reasoning with criteria for causality (DSK-3); and analytical values of using odds ratio (DSK-4) and 95% CI (DSK-5) (see Table 4). In coding, an epidemiologist and I first scored student responses individually (zero, partial, or full points) and then discussed the ratings for each criterion until both agreed with the scores. Analysis looked at

the differences in these two measures to assess student development in the conceptual knowledge and discipline-specific understanding, followed by correlational analysis to evaluate whether these two measures had some linear relationship to each other. In discussion, results of the survey analysis were discussed along with video/audio and worksheet analyses.

Results

Video/Audio analysis suggests that the instructor provided students with relevant instruction and materials for all five DSK items, but worksheet analysis suggests that the degree of student engagement and learning varied by lesson. Table 3 shows the counts of students who wrote down relevant notes to worksheet items. I start by summarizing relevant instruction and results of the worksheet analysis by specific DKS goals.

DSK-1: A scientist interview for communicating the social value of smoking research. In the first class of the unit (about 50 minutes), the teacher first asked the students who conducted tobacco research and why they did research on the topic, a few students mentioned medical profession and cancer. Then, she showed students four short video clips of an interview with a public health researcher who conducts tobacco research. In these clips, the researcher gave a summary of health issues related to smoking such as lung cancer, and the contribution of tobacco research including prevention, cessation, and protection from secondhand smoke. The researcher also introduced her own research project which her team conducted on a university campus, followed by how those findings were applied into campus policies. Students were asked to answer questions in the worksheets as they watched video clips. Worksheet analysis identified that all 14 students (100%) wrote down the problems caused by smoking and contributions of tobacco research (e.g., prevention, cessation, secondhand smoke) mentioned in the clips to worksheet (WS-1 and WS-2) (see Table 3). Closely looking at student notes for these items,

many students used expressions such as “death” and “addictive” which the researcher had mentioned during her interview.

DSK-2 & DSK-3: Making explicit the analytical values of observational design and criteria for causality. The teacher spent about 10 minutes watching an introductory video clip for epidemiologic research and two full classes (2 x 50 minutes) introducing case control design including causal reasoning with the epidemiologic criteria. Before beginning the methodological instruction, she showed a video clip of an interview with a PhD epidemiology student in which he introduces major characteristics of epidemiologic research. In the clip, he described epidemiologists as “disease detectives” and explained that epidemiologists look for associations between risk factors and diseases. To the interviewer’s question “What makes epidemiology unique from other sciences?” he answered that no single factor makes it unique, but the combination does by pointing out the observational nature of epidemiologic research; epidemiologists are often unable to control variables (uncontrollability of variables), who has risk factors or not, and they rely on what happens in reality (the nature of observational studies), and ethical concerns in randomization with human subjects in comparison to other practices where experimental design is appropriate.

In the next class, the teacher prompted students to recall the epidemiologist’s interview. For example, she asked them in what way epidemiology is unique from other sciences. A few students mentioned the uncontrollability of variables and observational nature of epidemiologic research in real-world settings in contrast to lab research. After the recall, she introduced the definitions of risk, risk/protective factor, and observational study, followed by case control design including the key terms and procedures with a hypothetical scenario in which teenagers who went to a party got sick and the house manager tried to identify the cause(s). Then, she

Chapter 2

presented another scenario of car accidents in which she introduced the concept of “matching” by pointing out that other factors could influence car accidents, such as drivers’ age (e.g., young drivers), location (e.g., busy streets), and time (e.g., commute time) for data collection. After the exercise, she introduced the difference between association and causation and explained the concept of confounding. During this instruction, she emphasized the presence of randomization in the study procedure as key in the difference between correlational and causal evidence and explained that observational studies are still valuable in certain situations with ethical concerns for experimental design and other pragmatic reasons (e.g., existing medical data can be used to identify risk factors of a disease).

In the last class of the lesson, she asked students to recall the major points from the last class including the observational nature of case control study, the difference between association and causation, and ethical issues of randomization with human subjects. Then she introduced the criteria for causality—strength of association, dose-response relationship, temporal sequence, consistence with other studies, biological plausibility, and lack of confounder—and engaged students in an exercise to apply causal reasoning with the criteria. After the exercise, she introduced systematic biases including information bias (e.g., asking about life styles in 10 years ago) and selection bias in relation to matching as a way to mitigate it. During instruction, she emphasized the analytical aim of the criteria is to determine the causality from associations as they worked on worksheet items.

Worksheet analysis identifies that 8 (57%) students noted the observational nature in epidemiologic research and 12 (86%) noted the uncontrollability of variables (WS-3); 8 (57%) students noted ethical concerns for conducting observational studies over experiments (WS-4); 10 (71%) noted the nature of correlational evidence with confounding (WS-5); 3 (21%) noted

their answers in the analytical aim of epidemiologic criteria (WS-6); and 6 (43%) described their conclusions with some epidemiologic criteria in causal inference (WS-7). These results show that only about half of the students explicitly noted ethical concerns in random assignment. Closely looking at student notes to the worksheet item, other students noted other rationale for adopting observational design (e.g., existing medical data can be used to identify risk factors) in response to instruction, which indicates that students did actually engage in worksheets. On the other hand, only 3 (21%) students explicitly described that the criteria are used to discriminate causation from association (WS-6), and closely looking at the students' notes suggest that 9 (64%) students noted that use of the criteria is to be "accurate" or "concrete", not sufficiently explicit in the difference between association and causation. Finally, about half of the students described some epidemiologic criteria, whereas other students described their conclusions without applying them.

DSK-4 & DSK-5: Making explicit analytical values of odds ratio and confidence interval. After the instruction in case control design and causal reasoning with the criteria, the teacher used two full classes to teach odds ratio and 95% confidence interval. In the first class, she presented the concepts of odds and probability using dice and compared the mathematical difference in the denominators, followed by an exercise in which students calculated both odds ratio and probability. Then she introduced the concept of odds ratio and explained the mathematical definition and conceptual meaning as the strength of association with an example. In interpretation, she told students that $OR > 1$ indicates a positive association or risk factor; $OR < 1$ indicates a negative association or protective factor; and $OR=1$ indicates no association. Afterwards, students worked on a few exercises in which they practiced odds ratio calculation and interpretation including the analytical aim of OR.

On the second day of the lesson, she asked students to recall the last class, particularly the interpretation of odds ratio when $OR > 1$, $OR < 1$, and $OR = 1$. After the recall, she showed a figure that represented a number of dots (people) in a circle (population) and introduced the concepts of population and sample. She then told students that the odds ratio they calculated last time is “sample OR” after (random) sampling but the real goal is to learn about “true OR” of the study population. During this instruction, she emphasized that it is practically impossible to study everyone in a study population to exemplify the analytical aim of statistical inference. Then she introduced the 95% confidence interval and explained how to interpret the statistical significance; it is significant when the interval is outside the one (i.e., null hypothesis implying no association), whereas it is not significant when one inside the interval. To clarify, she did not tell how to calculate the interval by saying that the database automatically calculates the numbers for them. To develop student understanding in the meaning of statistical significance, she instead showed another figure in which multiple intervals are drawn from multiple trials and explained the meaning of “95% confidence” by pointing out that the intervals will capture the true OR inside it in 95 times out of 100 trials, whereas in about 5 times out of 100 the intervals will not capture the true OR by chance. After the instruction, students worked on a few exercises for the 95% CI interpretation including the analytical aim of 95% CI.

Worksheet analysis identifies that 9 (64%) students explicitly noted the analytical aim of odds ratio (WS-8) and 12 (86%) demonstrated explicit OR interpretations (WS-9). On the other hand, only 4 (29%) students explicitly noted the analytical aim of 95% CI (WS-10), 2 (14%) did the meaning of 95% confidence (WS-11), and 9 (64%) noted explicit criteria for statistical significance (WS-12). These results suggest that students may not have understood the analytical aim of 95% CI (WS-10) and the meaning of 95% confidence (WS-11) or that part of instruction

was insufficient. Closely looking at student notes, about half of the students wrote down some notes to these items, but they tended to only mention some keywords such as “accuracy” and did not articulate the analytical aim and meaning in the relationship between population and sample in relation to random error.

Analysis of pre- and post-surveys. The statistics of pre and post results are shown in Table 5. The sample size is 20 for both pre and post surveys, as one student was absent at the pre survey another absent at the post survey. Pre results suggest that students scored 28% on average in conceptual knowledge items ($M = 4.68$, $SD = 2.83$), but students had little prior epistemic knowledge in DSK items ($M = 1.16$, $SD = 1.18$) although about half of the students seemed to know some effects of smoking before entering the unit. Closely looking at each item in conceptual knowledge, more than half of the students correctly answered multiple-choice items on study population, case and control, and sample size, and it indicates that these students may have applied their prior knowledge from some prior formal learning or guessed the answers from the item statements

A paired samples t-test was conducted to assess student improvement in conceptual knowledge. There was a significant difference in pre ($M = 4.68$, $SD = 2.83$) and post ($M = 11.89$, $SD = 2.71$) conditions; $t(18) = 7.83$, $p = .000$ and Cohen’s effect size value $d = 1.80$. The result suggests that students significantly developed their conceptual knowledge through the unit. In discipline-specific epistemic knowledge (DSK), likewise, there was a significant difference in pre ($M = 1.16$, $SD = 1.18$) and post ($M = 5.18$, $SD = 2.50$) conditions; $t(18) = 7.35$, $p = .000$ and Cohen’s effect size value $d = 1.69$. The result suggests that students also significantly developed their discipline-specific understanding. In correlational analysis, the Kendall tau-b rank

correlation between conceptual and DSK items was calculated, and the coefficient was only marginally significant; $\tau_b = .31$, $p = .08$.

Figure 2 shows the degree of improvement for two measures in percentage to the full scores. The post mean of conceptual knowledge items was 70% of the total score, whereas the post mean of epistemic scores was only 34% of the total. These suggest that, although there were significant improvements in both measures, DSK assessment still had significant room for improvement compared to their development in conceptual knowledge. Table 4 shows the counts of students who fully or partially scored by criterion in the pre and post surveys. It suggests that 15-18 out of 20 (75-90%) students fully or partially articulated the social value of smoking research, and 75% did the analytical aim of using odds ratios. However, on average only 32% fully or partially demonstrated their discipline-specific epistemic knowledge in other DSK items, after given relevant instruction.

Table 5

Results of Pre- and Post-Surveys on Conceptual Knowledge and Discipline-Specific Epistemic Knowledge Items

Category	Pre			Post		<i>t</i>	<i>df</i>	<i>P</i>	Cohen's <i>d</i>
	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Conceptual knowledge	17	4.68	2.83	11.89	2.71	7.83	18	.000	1.80
Discipline-specific epistemic knowledge	15	1.16	1.18	5.18	2.50	7.35	18	.000	1.69

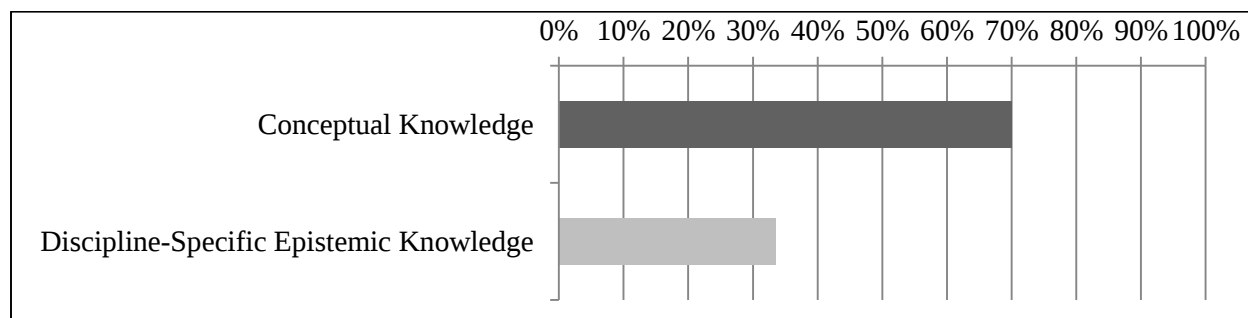


Figure 2. Percentage comparison to the full scores between conceptual knowledge and discipline-specific epistemic knowledge (DSK)

Discussion

The analytical goals of the case study were to assess students' discipline-specific epistemic knowledge in the epidemiologic discipline and identify revision points of the current instruction and scaffolds. In the following, I summarize major findings from each analysis and then discuss synthetic findings by each discipline-specific knowledge item. First, video/audio and worksheet analyses suggest that relevant instruction and materials were provided to all the five DSK items through the unit, whereas student engagement in worksheets seems to have varied by item. Worksheet analysis suggests that all students articulated the social value of smoking research, but fewer than half of the students articulated other analytical values. In particular, only 3 (21%) students articulated the analytical aim of epidemiologic criteria by discriminating causation from association, and only 2-4 (14-29%) articulated the analytical aim and meaning of confidence interval (see WS-6, WS-10, and WS-11 in Table 3).

Pre and post comparison suggests that students significantly developed both conceptual knowledge and discipline-specific epistemic knowledge (DSK) through the unit, and the result of correlation analysis suggests that these two sets of items seem to be not identical or measure

different aspects of disciplinary knowledge. Both effect sizes are large, and this indicates engaging in the unit led to a significant effect of their learning in both conceptual knowledge and DSK. However, in comparison to the full scores, DSK items still had significant room for improvement (34%) in comparison to conceptual knowledge (70%). This significant difference between two measures could be partly due to the different formats between multiple-choice and open-ended items, as the open-ended format is generally harder to answer with explicit descriptions. However, the higher standard deviation for post DSK results ($SD = 2.50$) compared to the pre ($SD = 1.18$) indicates that the degree of DSK development varied by student. Closely looking at the post results by criterion in Table 4, 15 (75%) or more students articulated the social value of smoking research and the analytical aim of odds ratio, whereas on average only 32% articulated other DSK criteria.

Together, these results suggest developing student understanding in analytical values seems more challenging than social values, even though students used these analytical actors in their inquiry project. In the following paragraphs, I analyze student learning processes related to the specific dimensions of DSK epistemic knowledge based on the triangulations among video/audio, worksheet, and survey analyses.

DSK-1: Smoking research. Given that all students articulated the social value of smoking research at the post survey, the scientist's video clips arguably helped students develop their DSK in social values, even considering that half of the students may have known the problems of smoking before entering the unit. To demonstrate an exemplar, one student wrote to the DSK item:

Tobacco/smoking is believed to cause many diseases and problems in the human body. If we can find a way to decrease these habits, people will hopefully stop smoking. If we

understood what caused people to smoke, we may be able to limit those factors. Also, by looking at the protective factors, we can increase those to stop people from smoking, such as: educating people on the damage of smoking.

His/her statement clearly mentions the problematic nature of smoking on the human body (DSK-1-a) and the contribution of smoking research in identifying the causes and applying the findings to prevent people from smoking (DSK-1-b). On the other hand, another student wrote, “It is important so people know the dangers of smoking and the risks it can bring.”, which mentions the problematic nature of smoking but it does not articulate the social aims. To account for the overall improvement, worksheet analysis suggests striking expressions such as “deaths” that the researcher expressed during her interview may have caught students’ attention and helped develop their understanding in the social value of smoking research and their interest in the topic.

DSK-2&3: Case control design & criteria for causality. Post survey results suggest that less than half of the students articulated epistemic criteria to these two DSK items. As an exemplar, one student wrote to the DSK-2 item on advantages and limitations of observational studies:

Observational studies are used over experimental studies for a few reasons, they cost less money and are more socially acceptable (it is difficult to get permission to run an experiment where we are trying to give people diseases), and they also tell us how the real world works. Just because something happens in a lab with no external variables, does not at all mean it will happen the same way outside of the lab. Using observational studies is more difficult to get firm proof of causation from, and is much less controllable, but we want the lack of control because that's how the real world is.

Chapter 2

His/Her statement clearly mentions the nature of observational studies (DSK-2-a), uncontrollability of variables (DSK-2-b), ethical and pragmatic concerns (DSK-2-c), and difficulty of confirming causation (DSK-2-d). However, other students tended to make less clear statements; for example, one student wrote, “Observational studies are cheaper and easier to do, but still are valid. They're also quicker, while randomly manipulated ones take more time. It's better to use them when the outcome has already happened.” This points to some advantages of observational studies but does not articulate the limitations. From worksheet analysis, about half of the students had taken notes to the DSK criteria in worksheets during instruction (WS-3 to WS-5), and about half articulated the criteria at the post survey except that only 4 (20%) articulated the difference between association and causation (DSK-2-d). To account for this inconsistency, one possible explanation would be that the item statement mentioned that the evidence with case control design is associative rather than causal, and therefore students may have not articulated it again in their answer. In addition, it would be also reasonable to interpret that the item statement did not explicitly ask students to articulate the four criteria and therefore they just articulated only some of them.

As for criteria for causality (DSK-3), many students made less clear statements and some students appeared to conflate the analytical value with odds ratio and confidence interval. For example, one student wrote, “The criteria for causality is used to make sure that the CI and odds ratio calculated from a study are not being influenced by untested or unknown factors,” and his/her statement addresses one of the epidemiologic criteria (lack of confounder) (DSK-3-b) but it does not articulate other DSK criteria. Another student wrote, “The criteria for causality is used in order to evaluate the data a little further, and if it follows all the criteria, it can be considered valid and reliable”, and many students used expressions such as “correct”, “valid”,

and “reliable” like the second statement, but their statements were unclear to identify what those expressions refer to. Although this may be also due to the lack of clarity of the item statement, I conjecture that some students may not have clearly understood the analytical value of criteria for causality by discriminating it from odds ratio and confidence interval.

DSK-4&5: Odds ratio and 95% confidence interval. The post survey results on these DSK items show that 15 (75%) students scored on 4-a) the analytical aim of odds ratio, whereas only 5-7 (25-35%) students scored on the other criteria. For example, one student wrote to the DSK item for odds ratio (DSK-4) as follows:

Odds ratio explains the likelihood of a case becoming a factor of a study. Scientists use odds ratio to determine what factors may be connected to a study or they can use it to compare factors that are more accurate and associated with the case. If an odds ratio number was less than or more than one, not containing the number one, then there is an association and significance. If the number is less than one, it is a protective factor. If the number is more than one, it is a risk factor. Odds ratio tells a scientist is there is an association of a factor to a study. It tells a scientist the chances of the factor being accurate.

His/her statement articulates the analytical aim of odds ratio (DSK-4-a), the interpretation when OR is higher or lower than one (DSK-4-b), and the discrimination from causation (DSK-4-c), although he/she could conflate it with confidence interval (“the chances of the factor being accurate”). Worksheet analysis suggests that only half of the students articulated the analytical value of odds ratio during foundational instruction (WS-8), and this indicates that the later inquiry project may have helped develop their epistemic understanding of it.

On the other hand, student statements to the DSK item for confidence interval (DSK-5) tended to be less articulated. For example, one student wrote, “It shows that 95% of the time, the odds ratio will be in that range. If the range includes the number 1, it does not show association.” This statement articulates the judgment criteria for statistical significance (DSK-5-c), but it does not mention the analytical aim for generalizability assessment (DSK-5-a). Although multiple explanations could be drawn—including the clarity of the item statements as discussed above—worksheet analysis suggests that students may not have understood the concept of 95% confidence interval from prior instruction (see WS-10, WS-11, and WS-12 in Table 3). These indicate that instruction may have been insufficient to develop their epistemic understanding of the idea of a confidence interval only for a short time period (one full class with 50 minutes). Yet, this still makes sense given that statistics education research suggests learning statistics is challenging to many students including college and adult learners through full coursework (Johnson & Kuennen, 2006; Green et al., 2009; Lunsford & Poplin, 2011).

Rare epistemological conversations during student-led inquiry project. There is evidence that students need to engage in sense-making discussions, reflecting on scientific practices, in order to develop knowledge of them (NRC, 2012). In this study, video/audio analysis was conducted for four video- and audio-recorded groups during their inquiry project and it was found that students discussed epidemiologic criteria (DSK-3-b), OR interpretation when $OR > 1$, $OR = 1$, $OR < 1$ (DSK-4-b), and criteria for statistical significance with CIs (DSK-5-c). However, students’ spontaneous conversations rarely occurred on the other items, particularly on the analytical aims of epistemic actors (tools); analysis identified at least one moment in which the teacher and I encouraged students to think why they use odds ratio, 95% confidence interval, and criteria for causality, and one student group started the following

Chapter 2

conversation on analytical aims of odds ratio, 95% confidence interval, and the epidemiologic criteria:

Karen: Why do you use odds ratio? ((Linda: Odds ratio.)) Isn't that to determine if there's a correlation between...

Linda: Okay. And then the 95%. That's to be confident in your results, right?

Karen: Yeah.

Linda: And what was the third question?

Karen: I don't remember. I'm sure we can...

Linda: Yeah, that's right. Yesterday, she wasn't clear but that's (inaudible).

This conversation shows that Karen and Linda briefly discussed analytical aims of the three epistemic tools, and at least the analytical aim of odds ratio was specifically stated. However, analysis did not identify this type of “why” talk from other groups, and most of their conversations were devoted to “what” these tools mean and “how” to use or interpret them. This reflects that the epidemiologic inquiry engaged students in choosing which exposure (variable) to investigate in their own decision but the method and measures (case control design, odds ratio, and confidence interval) were under the given condition, and therefore students did not have to ask for “why” they need to use them (Chinn & Malhotra, 2002). Given that it is unrealistic to leave every aspect of inquiry up to students' decision, I argue that one potential approach under such conditions would be to promote such “why” talk about the epistemic practices and activate students' epistemic sense making (Sandoval, 2003). In the epidemiologic inquiry unit, some specific tasks could be added to the unit such as having students justify the methodological choice for case control design over experimental design in a hypothetical context during instruction. However, adding these tasks would mean that it would require more time spent on the unit, and we could not spend more time for such activity due to constraints of regular class

schedules. To promote authentic inquiry in the science classroom, such time constraints should also be considered from a system viewpoint (Clark & Linn, 2003).

Conclusion

In this paper, I discussed the theoretical specification of epistemological understanding and proposed an observational approach to elicit students' disciplinary understanding. From the discussion, I argued that students' epistemological understanding should be framed as discipline-specific knowledge (DSK) rather than discipline-general knowledge (DGK) especially with an emphasis on the diversity of scientific disciplines under the current curriculum reforms in which the latest science education framework (NRC, 2012) encourages science learning through engaging in authentic practices. Engaging in authentic analytical tasks requires a deep disciplinary understanding of each scientific practice, while their epistemic norms and criteria do not necessarily agree each other. Therefore, students need to adjust their epistemological mode to the currently engaged practice from other practices or their prior learning. Although some researchers argue the significant role of DSK in disciplinary engagement, there are few studies on ways in which to theoretically frame DSK for observational research studies.

To this challenge, I proposed a value-focused approach to reveal students' DSK. The core idea of the value-focused approach is to frame DSK in the networks of epistemic alignments of heterogeneous epistemic actors (cognitive, social, material, and cultural). To reveal students' discipline-specific understanding in the epistemic relationships, I propose asking reflective "why" questions to encourage justifying the epistemic value of target actors and such value questions as to help elicit students' discipline-specific understanding in their justificatory statements. In this perspective, the proposed framework provides theoretical guidance for researchers to identify

and develop value questions for a focal practice, rather than pursuing a list of generalized static value questions that can be applied to all practices and disciplines.

In this study, five assessment items were developed based on the framework along with instructional goals. Pre- and post-survey results suggest that some students developed aspects of DSK of the epidemiologic discipline particularly in the alignment among epistemic and social aims of the epidemiologic practice and the analytical aim of odds ratio, whereas it seems more challenging to understand other analytical alignments such as the value of case control design, 95% confidence interval, and epidemiologic criteria. I argue that these value questions served to gauge students' DSK of the epidemiologic discipline and helped identify which parts of the lesson should be revised.

As for limitations, the theoretical discussion is a first step to reify DSK assessment and further research is needed. The foregoing discussion indicates that the clarity of assessment items is crucial to reveal student knowledge of focal epistemic alignments, and further research is needed to discuss how this value-focused approach could be used in other practice-focused curricula. An alternative empirical approach would be semi-structured interviews in which researchers could dynamically adjust their line of questioning to reveal students' discipline-specific understanding and identify their misconception.

Finally, the findings indicate that some specific scaffolding in such curriculum units is needed to promote students' epistemological conversations. One way to approach this challenge would be to engage students in "why" talk such as methodological negotiations or encourage students to justify why they need to use epistemic tools adopted in their investigation. Further research is needed to investigate how this type of epistemic talk serves to develop students' DSK in a range of science disciplines. The findings of the present study would contribute to the

theorization of discipline-specific knowledge, assessment methods, and scaffolding design for epistemic sense making toward epistemically-guided instruction so that students be able to deeply understand and appreciate the epistemic and social values of scientific practices with a broad viewpoint.

References

- Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The Nature of Science and Instructional Practice: Making the Unnatural Natural. *Science Education*, 82, 4, 417-36.
- Allchin, D. (2011). Evaluating Knowledge of the Nature of (Whole) Science. *Science Education*, 95, 3, 518-542.
- Alters, B. J. (1997). Whose Nature of Science? *Journal of Research in Science Teaching*, 34, 1, 39.
- Bell, P., & Linn, M. C. (2002). Beliefs about science: How does science instruction contribute? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing* (pp. 321-346). Mahwah, NJ: Lawrence Erlbaum.
- Bradford-Hill, A. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58: 295–300.
- Cartwright, N. (2004). Causation: One Word, Many Things. *Philosophy of Science*, 71, 5, 805-819.
- Clark, D., & Linn, M. C. (2003). Designing for Knowledge Integration: The Impact of Instructional Time. *Journal of the Learning Sciences*, 12, 4, 451-493.
- Dewey, J. (1939). *Theory of Valuation*. Chicago, Ill: The University of Chicago press.
- Driver, R., & Driver, R. (1996). *Young people's images of science*. Buckingham: Open University Press.

Chapter 2

- Elby, A., & Hammer, D. (2001). On the Substance of a Sophisticated Epistemology. *Science Education*, 85, 5, 554-67.
- Grotzer, T.A. (2012). *Learning causality in a complex world: Understandings of consequence*. Lanham, MD: Rowman Littlefield.
- Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge, U.K: Cambridge University Press.
- Hodson, D. (2009). *Teaching and learning about science: Language, theories, methods, history, traditions and values*. Rotterdam: Sense Publishers.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67, 1.
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, Mass: Harvard University Press.
- Koepsell, T. D., Weiss, N. S. (2003). *Epidemiologic Methods: Studying the Occurrence of Illness*, Oxford University Press.
- Laudan, L. (1984). *Science and values: The aims of science and their role in scientific debate*. Berkeley: University of California Press.
- Lederman, N.G. (2007). Nature of science: Past, present, and future. In S.K. Abell, & N.G. Lederman, (Editors), *Handbook of research in science education* (pp 831-879). Mahwah, New Jersey: Lawrence Erlbaum Publishers.
- Longino, H. E. (2002). *The fate of knowledge*. Princeton, N.J: Princeton University Press.
- Mayr, E. (2004). *What makes biology unique?: Considerations on the autonomy of a scientific discipline*. New York: Cambridge University Press.

Chapter 2

- Munn, M. M., Oura, H., Gallivan, M., Van, H. K., & Shouse, A. W. (2013). Engaging high school students in research on smoking behavior. *Science*, 341, 6144, 360-1.
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C.: The National Academies Press.
- Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "Ideas-about-Science" Should Be Taught in School Science? A Delphi Study of the Expert Community. *Journal of Research in Science Teaching*, 40, 7, 692-720.
- Lee, T., Oura, H., T., Scalone, G., Shutt, K., Bransford, J., Shouse, A., Van Horne, K., Vye, N., Munn, M., Knuth, R., & Bell, P. (2012). Gathering Evidence of Scientific Argumentation Practices: From Pre-Kindergarten to High School. In J. van Aalst, K. Thompson, M. J. Jacobson & P. Reimann (Eds.), *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012)* (Vol. 2, pp. 41-48). Sydney, NSW, Australia: International Society of the Learning Sciences.
- Rothman, K. J. (2002). *Epidemiology: An introduction*. Oxford: Oxford University Press.
- Rouse, J. (1996). *Engaging science: How to understand its practices philosophically*. Ithaca: Cornell University Press.
- Rudolph, J. L. (2003). Portraying Epistemology: School Science in Historical Context. *Science Education*, 87, 1, 64-79.
- Samarapungavan, A., Westby, E. L., & Bodner, G. M. (2006). Contextual Epistemic Development in Science: A Comparison of Chemistry Students and Research Chemists. *Science Education*, 90, 3.

Chapter 2

Schwab, J. J., Westbury, I., & Wilkof, N. J. (1978). *Science, curriculum, and liberal education: Selected essays*. Chicago: University of Chicago Press.

Schwartz, R., & Lederman, N. (2008). What Scientists Say: Scientists' Views of Nature of Science and Relation to Science Context. *International Journal of Science Education, 30*, 6, 727-771.

Schwartz, R. S., Lederman, N. G., & Abd-El-Khalick, F. (2012). A series of misrepresentations: A response to Allchin's whole approach to assessing nature of science understandings. *Science Education, 96*, 4, 685-692.

Stampfer, M. J., Ridker, P. M., & Dzau, V. J. (2004). Risk factor criteria. *Circulation, 109*, 25, 3-5.

Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005). Comparative understanding of school subjects: Past, present and future. *Review of Educational Research, 75*, 2, 125-157.

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, D.C: Brookings Institution Press.

Sandoval, W. A. (2005). Understanding Students' Practical Epistemologies and Their Influence on Learning through Inquiry. *Science Education, 89*, 4.

U.S. Department of Health and Human Services (2012). *Preventing Tobacco Use among Youth and Young Adults: A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction, 16*, 1, 3-118.

Chapter 2

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the Scientific Method: Model-Based Inquiry as a New Paradigm of Preference for School Science Investigations.

Science Education, 92, 5, 941-967.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Wong, S. L., & Hodson, D. (2008). From the horse's mouth: What scientists say about scientific investigation and scientific knowledge. *Science Education*, 93, 1, 109-130.

Chapter 2

Table 3

Results of Worksheet Analysis for Student Engagement During Instruction (N = 14)

ID	Worksheet Item	Relevant DSK Criteria	# Count (%)
WS-1	What are the problems with smoking?	1-a) Problems caused by smoking	14 (100)
WS-2	What are the three major goals on which tobacco research focuses?	1-b) Contributions of smoking research	14 (100)
WS-3	What makes epidemiology unique from other sciences?	2-a) Observing in real world settings	8 (57)
		2-b) Uncontrollability of variables	12 (86)
WS-4	Why do scientists conduct observational studies (e.g., case control design) other than experiments?	2-c) Ethical concerns	8 (57)
WS-5	Why does association NOT necessarily indicate causation?	2-d) & 4-c) Association, not necessarily causation	10 (71)
WS-6	What do we use the criteria for causality for in analysis of case control studies?	3-a) Discriminating causation from association	3 (21)
WS-7	Draw your conclusion using the criteria for causality (as response to an exercise in causal inference).	3-b) Epidemiologic criteria	6 (43)
WS-8	What do we use odds ratio (OR) for in analysis?	4-a) Quantifying the strength of association	9 (64)
WS-9	What would the odds ratio tell us in the following scenarios? Explain. When OR = 1 & When OR < 1.	4-b) Criteria for OR interpretation	12 (86)
WS-10	What do we use 95% confidence intervals and evaluate the statistical significance for in analysis of case control studies?	5-a) Generalizability to the study population	4 (29)
WS-11	What does “statistically significant” exactly mean?	5-b) Meaning of 95% confidence	2 (14)
WS-12	How do we determine whether or not an odds ratio is significantly bigger/smaller than one with the confidence interval?	5-c) Criteria for statistical significance	9 (64)

Note. WS-5, WS-6, WS-8, and WS-10 are also related to 3-c) Limitations of OR and CI in causal inference in the DSK criteria

Chapter 2

Table 4

By-Criterion Results of the Pre and Post Surveys on Discipline-Specific Epistemic Knowledge Items (N = 20)

Assessment Item	Assessment Criteria	Pre	Post	Diff.
DSK-1: Smoking research				
Chloe wonders why tobacco/smoking research matters to society. How would you help her understand the value of this research? Explain problems caused by tobacco/smoking and the contributions of tobacco/smoking research.	1-a) Problems caused by smoking	11	18	7
	1-b) Contributions of smoking research	11	15	4
DSK-2: Case control design				
Knowing that the evidence from observational case control studies is associative rather than causal, Aaron wondered why epidemiologists often use observational studies rather than randomly manipulated experiments that can produce causal evidence. How would you help him understand the value of observational studies? In your answer, include the advantages and limitations of observational studies and under what conditions it is better to use them.	2-a) Observation in real world settings	3	9	6
	2-b) Uncontrollability of variables	2	8	6
	2-c) Ethical concerns	1	6	5
	2-d) Association, not necessarily causation	0	4	4
DSK-3: Criteria for causality				
Steve has never used the criteria for causality to analyze data from observational case control studies. He wonders why scientists need it in addition to odds ratios and 95% confidence intervals. How would you help him understand the value of using the criteria for causality? Explain its purpose and the limitations of odds ratio and 95% CI.	3-a) Discriminating causation from association:	0	6	6
	3-b) Epidemiologic criteria	0	7	7
	3-c) Limitations of OR and CI in causal inference	0	5	5
DSK-4: Odds ratio				
Tyler has never used odds ratio calculation to analyze data from case control studies. He wonders why scientists use odds ratios to analyze data. How would you help him understand the value of using odds ratio to analyze data? Explain its purpose and what the number indicates.	4-a) Quantifying the strength of association	0	15	15
	4-b) Criteria for OR Interpretation	0	7	7
	4-c) Association, not necessarily causation	0	6	6
DSK-5: 95% confidence interval				
Sandra has never used 95% confidence intervals (CIs) to analyze data from case control studies. She wonders why scientists need it in addition to odds ratio to analyze data. How would you help her understand the value of using 95% CIs? Explain its purpose and what the interval indicates.	5-a) Generalizability to the study population	0	7	7
	5-b) Meaning of 95% confidence	0	5	5
	5-c) Criteria for statistical significance	0	7	7

Note. Each count is the number of students who gained full (= 1) or partial (= 0.5) scores to the DSK item

**Chapter 3. Developing High-School Students' Epidemiologic Reasoning through
Statistical Database Analysis**

Hiroki Oura

College of Education, University of Washington

Abstract

Developing disciplinary habits of reasoning in students is a major goal of science education. Psychological studies suggest that children have a sense of causal relationships from an early age, although their causal reasoning is limited to relatively simple relationships in which cause and effect are close to each other in time and space. Probabilistic explanation and thinking is one of the causal forms that is challenging for students to learn and researchers have discussed instructional approaches to develop students' statistical reasoning. This article reports such an effort to engage high school students in an epidemiologic practice involving statistical reasoning and discusses design principles to develop their epistemological knowledge and reasoning on population-based explanation. Findings from a case study suggest that our scaffolding design guided students' disciplinary engagement in the epidemiologic inquiry and developed their causal reasoning about populations, while some points for refinement were also identified on their generalizability assessment with statistical evidence.

Developing disciplinary habits of causal reasoning in students is a major goal of science education (AAAS, 1993; NRC, 1996). Psychological studies suggest that children have a sense of causal relationships from an early age, while their causal reasoning is limited to relatively simple relationships in which cause and effect are close each other in time and space (Carey, 2011). Researchers argue that science education should serve to develop students' complex reasoning, whereas research suggests that it is formidable to develop their reasoning skills only by traditional instruction (Clement, 1982). Some researchers ascribe it to students' lack of knowledge in various forms of causal explanations and argue the need of epistemic variations in science instruction to expand students' epistemological understanding on cause (Grotzer, 2012). Probabilistic explanation and thinking is one of the causal forms that is challenging for students to learn and researchers have discussed instructional approaches to develop students' informal statistical reasoning (Makar, Bakker, & Ben-Zvi, 2011).

In recent years, researchers have come to a consensus that engaging students in authentic epistemic practices is a promising approach to develop students' disciplinary reasoning (NRC, 2012), and some researchers have argued that students' disciplinary engagement requires developing not only their conceptual knowledge but also epistemological knowledge in focal practices (Bell & Linn, 2002). Such epistemological knowledge includes forms of explanation (e.g., descriptive, mechanistic, probabilistic) and methodological choices (e.g., experimental, observational) associated with studying specific phenomena along with epistemic aims. Such discipline-specific epistemic knowledge is essential to make sense of why scientists pursue particular forms of explanation and adopt particular methods and why it matters to pursue particular epistemic aims. Accordingly, discipline-specific scaffolding is crucial to ensure

Chapter 3

students' disciplinary engagement in an appropriate epistemological mode to currently engaged practice (see Chapter 1).

In this study, I report such an effort of discipline-specific design to engage high school students in the authentic practice of epidemiological analysis. Epidemiology is an area of contemporary research that studies the occurrence and distribution of health-related events (e.g., disease), and epidemiologic theory contributes to identifying causes of illness and preventing people from it in public health (Koepsell & Weiss, 2003). Epidemiologic inquiry is epistemologically distinctive from typical inquiry tasks in standard curricula: (a) epidemiologic explanations are population-based, (b) epidemiologic methods are often observational rather than experimental, and (c) epidemiologic reasoning requires both statistical and biological reasoning (Rothman, 2002). The present study examines student development in epidemiologic reasoning by analyzing their engagement in designed inquiry tasks and their written responses to argumentative tasks.

In the following sections, I first review literature in causal reasoning and instructional approaches to develop students' disciplinary reasoning with a focus on the need for discipline-specific scaffolding to develop student reasoning through engaging in authentic practices including characteristics in epidemiologic reasoning. Second, I introduce our curriculum unit Exploring Databases (ExDa) (Munn et al., 2013) and discuss the scaffolding design to encourage students' disciplinary engagement in epidemiological data analysis. Last, I report results of a case study and discuss how students engaged in the epidemiological investigation and developed epidemiologic reasoning.

Theoretical Framework

Becoming able to reason about complex relationships in the world benefits children's reasonable decision-making ability (Grotzer, 2012). Psychological studies show that children have a sense of causal reasoning in their observations from an early age (Carey, 2011), whereas their causal cognition is limited to relatively simple and direct relationships in which cause and effect is visible and nearby in time and space. There are more complex phenomena in which causal relationships are hardly visible by size, distanced in time and space, and/or involve multiple causal pathways (Hacking, 2001; Woodward, 2003; Cartwright, 2004), and scientific practices have established various analytical norms and tools to account for such phenomena (Knorr-Cetina, 1999). Therefore, researchers believe science education should serve to develop students' disciplinary habits of mind through the learning of science (NRC, 2012).

Yet literature suggests that developing causal reasoning is a formidable goal with traditional instruction, as researchers have reported various student problems in causal reasoning across disciplines including biology (e.g., Peker & Wallace, 2011), chemistry (e.g., Talanquer, 2010), physics (e.g., Clement, 1982), and other disciplines (Grotzer & Baska, 2003; Resnick, 1996). For example, Clement (1982) demonstrated that students tend to apply causal reasoning about the relationship between force and acceleration based on directly visible variables such as velocity and with less attention to Newton's laws, even after instruction. Talanquer (2010) analyzed college students' written explanations in boiling and dissolution and found that students tend to generate explanations based on visual cues (boiling and freezing points) rather than making sense of the underlying theory that accounts for the phenomena (e.g., osmosis). One representative issue across these studies is that students tend to draw causal reasoning based on direct visual cues or variables with a lack of "theory-laden" reasoning.

In order to develop students' theoretical reasoning, researchers identify the need of developing epistemological understanding in different forms of causal relationships. Grotzer (2003) points out that students tend to have a limited causal repertoire and this leads to simplifying causal structures and distorting their observations along with their naïve frames. Perkins and Grotzer (2005) argue that science education should expose students to various forms of causal explanations (i.e., mechanism, interaction pattern, probability, and agency) and expand their epistemological understanding in different forms of causal relationships. In probabilistic thinking, psychologists pursue modeling human contingent reasoning (Allan, 1993; Gopnik et al., 2004; Pineño & Miller, 2007), and accumulative studies have demonstrated that the human mind seems not naturally built to reason decision making in stochastic phenomena in the same ways with formal theory (e.g., Kahneman and Tversky, 1972; Hirsch & O'Donnell, 2001; Hjalmarson, Moore, & delMas, 2011).

Engaging Students in Probabilistic Reasoning through Statistical Inference

Scientific practices often pursue probabilistic explanations through *statistical inference*. In general, the goal of descriptive statistics is to organize, summarize, and present data to understand and communicate aggregate information, whereas the goal of inferential statistics is intended to generalize data or draw conclusions “beyond data” (Moore, 1993; Watson, 2006). Although the first goal may be achieved through mathematics courses, achieving the second goal requires a substantial understanding in related concepts specific to statistics.

Regardless of the fact that statistical inference is one of the core methodologies in various scientific practices (Hacking, 2001), teaching statistics is challenging in K-12 education due to students' inadequate preparation regarding mathematical skills covered by K-12 curricula (Watson, 2008). To learn statistics, students usually need to know mathematical concepts such as

fraction, proportion, and probability, whereas mathematics education research suggests that these topics are challenging to K-12 students (National Mathematics Advisory Panel, 2008). Statistics education research reports that learning statistics can be challenging to college-level students and adults, even those whose academic success seems to be associated with their knowledge in mathematics (Johnson and Kuennen 2006; Green, Stone, Zegeye, & Charles, 2009; Lunsford & Poplin, 2011). Given these reports, statistical analysis in a problem-solving context would be even more challenging especially to primary and middle school students (Mix, Levin, & Huttenlocher, 1999; Sophian & Wood, 1997).

Related to these challenges, mathematics and statistics education communities have been actively discussing the ways of developing students' knowledge and skills in *informal statistical inference* (ISI). ISI aims to develop students' informal statistical reasoning "as the ways in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples" (Pratt & Ainley, 2008). Paparistodemou and Meletiou-Mavrotheris (2008) also define it by "informal statistical reasoning as a continuum of experience from the point when students start to pose questions about datasets to the point when they are about to meet formal inferential statistics." The core ideas are to develop student understanding on statistical concepts without or with less formal symbols and evolve their formal knowledge through inquiry or problem-solving activities.

Specifically, Paparistodemou and Meletiou-Mavrotheris identify four "big ideas" as core conceptual categories for ISI: (a) properties of aggregates rather than properties of individual cases, (b) sample size and its effect on the accuracy of population estimates or on process signals, (c) controlling bias, and (d) tendency to distinguish between claims that are always true and those that are often or sometimes true. The properties of aggregates refer to population

parameters such as mean and variance. Students should develop an understanding that sample size affects the accuracy of population estimates, but not the ratio of sample size to the number of population (Smith, 2004). Signals refer to some patterns seen in data, and students should interpret the patterns by distinguishing them from noise expected by chance (Konold & Pollatsek, 2002). Students should be aware of potential systematic biases such as confounding and develop and interpret conclusions with the degree of certainty using probabilistic language. Moreover, Paparistodemou and Meletiou-Mavrotheris also emphasize developing student understanding in the whole process of data analysis from sampling to drawing conclusions in order to convey these principles along with the goal of statistical inference in scientific practices (see also Watson, 2006). In instruction, ISI researchers often suggest using graphing and modeling tools for statistical learning (Wild, Pfannkuch, Regan, & Horton, 2011; Konold & Miller, 2005, Makar et al., 2011).

Engaging Students in Authentic Data Analysis with Discipline-Specific Scaffolding

In recent years, science education researchers have come to a consensus that engaging students in *authentic* epistemic practices is a promising approach to develop students' disciplinary habits of mind. Although standard curricula already engage students in “inquiry” activities, researchers have pointed out that these activities are often inauthentic compared to professional scientific inquiry. As an example of this, Chinn and Malhotra (2002) analyzed commercially available textbooks compared to those developed by researchers. They found that inquiry tasks in textbooks are often simplified and lack epistemic tasks that are unique to scientific inquiry including generating research questions, designing studies, explaining results, developing theories, and studying research reports. Researchers argue that such epistemic features should be made explicit and salient in classroom inquiry in order to develop students'

epistemological understanding in scientific inquiry (Chinn & Malhotra, 2002; Sandoval, 2003; Hodson, 2009). Accordingly, the latest science education framework emphasizes science learning through engaging in authentic “practices” (NRC, 2012).

To engage students in authentic practices, scaffolding design is thought to be crucial to make novice students’ disciplinary engagement possible. In the last two decades, researchers have discussed design principles to engage students in disciplinary tasks including hypothesis development and prediction (Lawson et al., 2000), experiments and observations (Chinn & Brewer, 1993; de Jong & Van Joolingen, 1998; Windschitl et al., 2007; White & Frederiksen, 1998), data analysis (Lehrer & Schauble, 2002; Hug & McNeill, 2008), evidence evaluation (Nicolaidou et al., 2011), and explanation construction and scientific argumentation (Bell, 2004; Bell & Linn, 2000; Sandoval, 2003; McNeill, 2011). Some of these studies have actually developed curriculum modules and demonstrated that students’ disciplinary engagement in authentic inquiry is feasible under deliberate scaffolding (White & Frederiksen, 1998; Bell & Linn, 2000; Lehrer & Schauble, 2002; Sandoval, 2003).

While these studies mainly focus on modeling focal processes associated with scientific practices and theorizing design principles to guide students’ disciplinary engagement in each focal process of inquiry, some researchers have recently argued the need of a more explicit focus on epistemological development, particularly in discipline-specific aspects of scientific practices (Elby & Hammer, 2001; Bell & Linn, 2002; Sandoval, 2003; Samarapungavan, Westby, & Bonder, 2006). Sandoval (2003) suggests engaging students in disciplinary contexts related not only to scientific concepts in the subject matter but also epistemological knowledge in what should be explained and what data should be counted as evidence to support their claim in the currently engaged practice (see also Hammer & Elby, 2001). I elsewhere pointed out that

scientific practices involve various epistemological attributes that are peculiar to each practice such as epistemic aim, form of causal explanation, methodology, and criteria for causality, and argued that knowing these attributes are important for students to make sense of why scientists pursue a particular form of explanation and adopt a particular method in focal practices (see Chapter 1). To achieve this goal, I proposed an epistemic design framework that helps researchers reify discipline-specific epistemologies for a focal practice with those key epistemic actors and epistemic narratives to account for the relationships among the actors (see Chapter 1).

A Contemporary Practice: Epidemiology and Population-based Reasoning

In this study, I report on an effort of discipline-specific scaffolding of population-based reasoning that engaged students in a contemporary practice of epidemiology. Epidemiology is an area of research that studies the occurrence and distribution of health-related events (e.g., disease), and epidemiologic inquiry has distinctive characteristics from the forms of reasoning students typically experience through school inquiry in standard curricula. First, epidemiological inquiry often pursues a form of explanation called “risk factor.” Risk factor is defined as “surrogates for deeper causes and better predictions” (Stampfer et al., 2004), and the epistemic aim is to account for probabilistic or *population-based* explanations rather than single-case events due to its major interest in the occurrence and distribution of health-related events. Another reason for pursuing such probabilistic form of explanation is that many diseases have multifactorial traits and there can be multiple causal pathways to the single outcomes, and therefore it is often difficult or impossible to explain a disease with one single cause among people in a definitive manner. Therefore, epidemiologists often pursue probabilistic risk estimates based on statistical evidence associated with theoretically-informed variables of interest. This form of reasoning is arguably very rare in standard science curricula in which

Chapter 3

functional and mechanistic explanations are dominant. Note that this does not say epidemiology disregards all single cases and mechanistic explanations; it says that the epistemic aim is population-based explanations and it actually incorporates mechanistic explanation to account for causal relationships between risk factors and health-related outcomes when experimental evidence is available (Koepsell & Weiss, 2003).

Another distinctive difference between epidemiologic inquiry and school inquiry is *causal reasoning with correlational evidence*. Epidemiologists often adopt case-control design whose method aims to compare the degree of exposures, hypothesized factors, between “cases” who have the outcome (disease) and “controls” without the disease in their studies. Case-control design is basically *observational* without randomization, and therefore the evidence is correlational, not necessarily causal. Yet, the ultimate goal is to develop causal explanations and epidemiologists apply causal reasoning and judgments with various analytical techniques and disciplinary criteria to discriminate causation from association. This form of causal reasoning in curricula is arguably rare, given that classical controlled experiments are dominant in school inquiry (Windschitl et al., 2007).

Last, epidemiologic reasoning involves *explicit assessments on the generalizability of explanations* between statistical and biological representativeness. Statistical evidence generally provides estimates of the generalizability to study populations (i.e., statistical representativeness), while epidemiologic inquiry also takes *biological representativeness* into account for the generalizability assessment. Biological representativeness could be claimed based on biological similarities, for example, between mice and human in medical research, although their biological characteristics are not perfectly identical (Rothman, 2002). In epidemiology, an explanation derived from male subjects could be applied to female unless there is a specific biological

mechanism that differs between them in the literature. In these ways, epidemiologic reasoning involves complex assessments on the generalizability between statistical and biological representativeness.

Engaging Students in Epidemiologic Data Analysis: Exploring Databases (ExDa)

Exploring Databases (ExDa) is a curriculum unit designed to engage high-school students in an epidemiologic practice to identify risk factors of smoking in collaboration with genome scientists and epidemiologists (Munn et al., 2013). Nicotine addiction is known as a leading cause of preventable illness and death in the United States and smoking behavior is known as a “multifactorial trait” in which people start and continue smoking due to a variety of causes and pathways including physiological/genetic and environmental factors. Identifying risk factors of smoking is valued to prevent citizens from starting and continuing smoking in the social aim of public health, and this topic is particularly relevant to high-school students given the fact that many smokers start smoking in their youth.

The unit consists of foundational lessons about the topic and analytical methods and a student-led inquiry project where they work to develop a model for the health behavior. First lessons are designed to develop student knowledge in key concepts related to smoking by combining neurobiology, genetics, and epidemiology. Specifically, students read smoking profiles that describe how people start, continue, and/or quit smoking in a story format and develop their understanding in the scientific model of these transitions. Next, they learn about the biological/genetic mechanism of nicotine addiction with models of neurotransmission and a few gene regions suspected to be associated with the addiction. The subsequent lessons are devoted to introducing epidemiology and students develop their understanding in epidemiologic research. Specifically, students learn case control study design and two statistical concepts of odds ratio

(OR) and 95% confidence interval (95% CI) that are used in their inquiry projects. After the foundational learning, students explore an online database of survey responses from approximately 300 residents in the Puget Sound area. The survey was designed by a collaboration of scientists, teachers, and students in a previous project. In sum, students explore possible factors that are associated with smoking addiction and conduct hypothesis development and testing to develop population-based explanations for the risk factors.

Model of Epidemiologic Inquiry

The goal of inquiry in the ExDa unit is to identify risk factors of smoking in the Puget Sound area, and students engage in four basic procedures: learning to use the smoking database, hypothesis development, statistical test development & interpretation, and causal inference (Table 1). After foundational learning, students first learn about the smoking database. The database contains about nearly 100 questions related to a variety of topics including environmental exposures (e.g., parental smoking) and physiological exposures (e.g., arousal after smoking) for about 300 survey participants. In addition, these participants provided a small blood sample to genotype their DNA at three candidate gene regions, whose data also are stored in the database. Students develop an understanding of the dataset by reading the survey/genotype items and study procedure for the original case control study including the definitions of cases and controls and demographic imbalance between cases and controls using descriptive statistics. Meanwhile, they identify a research topic and develop their overarching hypothesis (e.g., “family smoking increases the risk of smoking”).

After hypothesis development, they work to test their hypothesis through queries of the database. Students search for up to five items related to their overarching hypothesis and develop statistical tests to estimate the odds ratios (ORs) and 95% confidence intervals (CIs). OR is

calculated by the odds of cases with exposure to those without it over the odds of controls with exposure to those without it, and the numerical value shows the strength of association between exposures and smoking. Specifically, when $OR > 1$ the exposure could be a risk factor of smoking, whereas $OR < 1$ shows that the exposure could be a protective factor that actually keeps people away from smoking. In addition, CI is the interval estimate of the population OR and used to determine whether the sample OR is statistically significant or the population OR is expected to be bigger/smaller than one (= no association) with the 95% confidence. Specifically, a sample OR is confirmed significant when the value of one (null) is outside the interval, otherwise it does not have enough evidence to claim the significant association.

When some association(s) is confirmed significant, students make a causal inference to judge whether the association(s) is causal rather than just associative with epidemiologic criteria: strength of association, temporal sequence, dose-response relationship, consistent with other studies, biological plausibility, and lack of confounder drawing from Bradford-Hill (1965). The strength of association is assessed with the value of OR as discussed above and dose-response (proportional) relationship is assessed under specific conditions when it can check the proportional change of OR by changing the degree of exposure with the database. In addition, students search Google Scholar for related studies and read the abstracts to reason whether their results are consistent with prior studies. They also apply theoretical reasoning to judge whether the relationship between the exposure and smoking makes sense by some mechanistic explanation or biological sense and whether some other (third) factor could influence both exposure and smoking (confounding). With the statistical and theoretical reasoning, students finally draw their own conclusions and report them with their peers and teacher.

Table 1

Model of Epidemiological Inquiry in the ExDa Unit

Inquiry phase	Task
Learning the Smoking Database	• Reading study procedures for the original case control study • Understanding case and control definitions • Checking selection bias between cases and controls
Hypothesis Development	• Reading smoking profiles • Exploring the smoking database • Developing an overarching hypothesis
Statistical Test Development & Interpretation	• Identifying up to five relevant items to the overarching hypothesis • Develop statistical tests and estimate the ORs and CIs • Interpreting the statistics
Causal Inference	• Making a causal inference with epidemiologic criteria • Drawing conclusions for their presentation

Scaffolding design for epidemiological reasoning

A major challenge for engaging in the epidemiological inquiry is statistical analysis and causal reasoning on population-based explanation. To engage in this practice, students need to understand not only the statistics of ORs and CIs, but also the concept of statistical inference and discipline-specific attributes including the epistemic focus on population-based explanation rather than individual cases. It is expected that students are unfamiliar with this form of reasoning since mechanistic or case-based reasoning is dominant in school inquiry unless they have taken related AP statistics courses. Given the assumption, we designed three different sets of scaffolds to help students develop statistical and discipline-specific knowledge and guide their disciplinary engagement.

Communicating the epistemic aim of population-based explanation. First, students need to understand the epistemic aim of *risk factors* and why it matters along with the social aim of public health. Figure 1 shows the epistemic alignments with key epistemic actors for the

epidemiologic inquiry (see Chapter 1). Specifically, the epistemic aim of developing population-based explanation is valued due to health-related problems caused by smoking, and epidemiologists work to identify risk factors of smoking that explain patterns in populations rather than individual cases for the sake of public health. In addition, they adopt case control design because of ethical concerns in randomly assigning subjects into the conditions that may lead to smoking.

To help students understand the epistemic alignment, we created two sets of short video clips in which two scientists talk about the epistemic aim of identifying risk factors of smoking and the epistemic value of observational methods in epidemiology. In the first set, a public health researcher explains in an interview format the effects of tobacco on human health and a few key findings from tobacco research (e.g., addiction, cessation, secondhand smoke) including her own research on characteristics of youth smoking. These clips were intended to help students understand problems related to smoking, the social aim of tobacco research, and the epistemic aim of identifying risk factors.

In the other set of video clips, a PhD epidemiology student explains general ideas of epidemiologic research in which he emphasizes its analytical approaches with observational methods. Specifically, he expresses epidemiologists as “disease detectives” and explains that epidemiologists look for associations between risk factors and diseases. To the interviewer’s question “What makes epidemiology unique from other sciences?”, he answers that no single factor makes it unique but the combination does by pointing out the observational nature of epidemiologic research; epidemiologists are often unable to control variables (i.e., uncontrollability) for who has specific risk factors or not, and therefore they rely on what

happens in reality (i.e., an observational design) in relation to ethical concerns in randomization with human subjects compared to experimental studies.

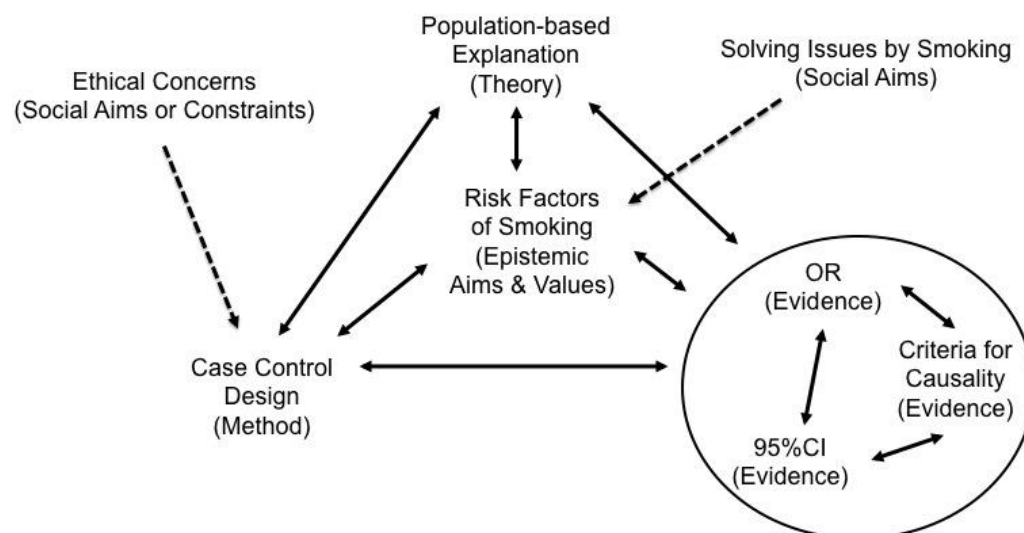


Figure 1. Epistemic alignments for the epidemiologic inquiry

Developing student understanding in statistical inference. Second, students need to understand the methodology of statistical inference in order to make sense of why they need to use statistics to achieve the epistemic goal. Statistical inference is an essential concept to understand analytical procedures to seek population-based explanation in the epidemiologic inquiry. To develop student understanding of the concept, we designed expository materials for teachers and students to understand the concepts of *population* and (random) *sample* and the analytical role of 95% confidence interval as a way to assess the statistical significance. Specifically, the teacher shows Figure 2 and explains that individual studies are only able to collect and calculate the (sample) odds ratios from a subset of people in a population (small circle), whereas the true interest is to learn about the entire population (large circle). Then, the teacher introduces the concept of 95% confidence interval and explains the analytical role as a

way to assess the statistical significance. Meanwhile, the teacher also explains that random sampling is assumed in the analytical logic, whereas the case control study whose data students analyze was voluntary and some *systematic bias* may exist. To develop student understanding in the concept, the unit engages students in one worksheet task of comparing the numbers of cases (those who smoke) and controls (those who do not smoke) in the database by demographic category including age, gender, and ethnicity to evaluate whether there is any imbalance between cases and controls in the dataset when they learn about the database.

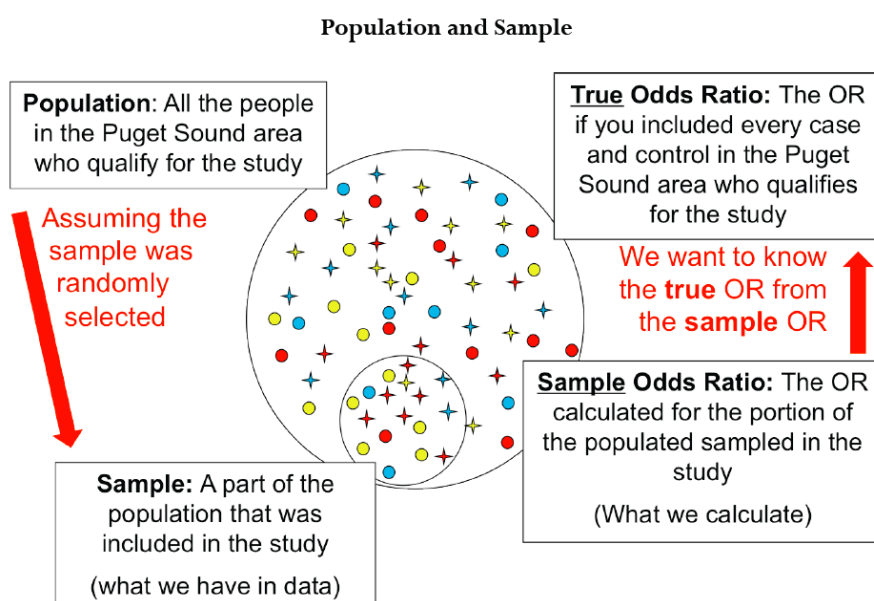


Figure 2. Population and sample on worksheets. Colored stars and circles represent cases and controls, and the colors represent different characteristics. Students are encouraged to notice that there are relatively more red stars and yellow circles in the sample than in the whole population.

Process-focused scaffolds to guide statistical analysis and causal inference. To guide disciplinary engagement in epidemiological analysis, we designed paper worksheets and technological scaffolds in the Web-based Inquiry Science Environment (WISE) (Slotta & Linn,

2009), in which each step of the data analysis (searching the database, developing statistical tests, and interpreting the results) were guided by corresponding scaffolds. Since the central goal of the inquiry is to engage students in epidemiologic reasoning, not develop mathematical or calculative skills, we developed an interface on the online database to make the mathematical procedures *transparent* and allow students to focus on analytical reasoning (Figure 3).

Specifically, once students choose an item in the database, the interface asks them to write down their hypothesis for the item or exposure and then choose which answer key (e.g., yes, no) is defined as exposed or not-exposed according to their hypothesis (left image). With the pair defined, students click the get-odds-ratio button and results with the OR and 95% CI are calculated with a figure that shows the point and interval estimates in relation to the null hypothesis, the value of one, which indicates no association (right image).

To engage students in disciplinary reasoning, we designed a set of worksheet tasks to guide causal inference with epidemiologic criteria. Specifically, these worksheet items ask students to write down their overarching hypothesis, relevant database items, and which answer key is defined as exposed or not-exposed with a table in which students record their test results of relevant statistics. In causal inference, worksheet items ask students to search Google Scholar for a few related studies and summarize the major findings and critically think of any potential systematic bias along with the epidemiologic criteria.

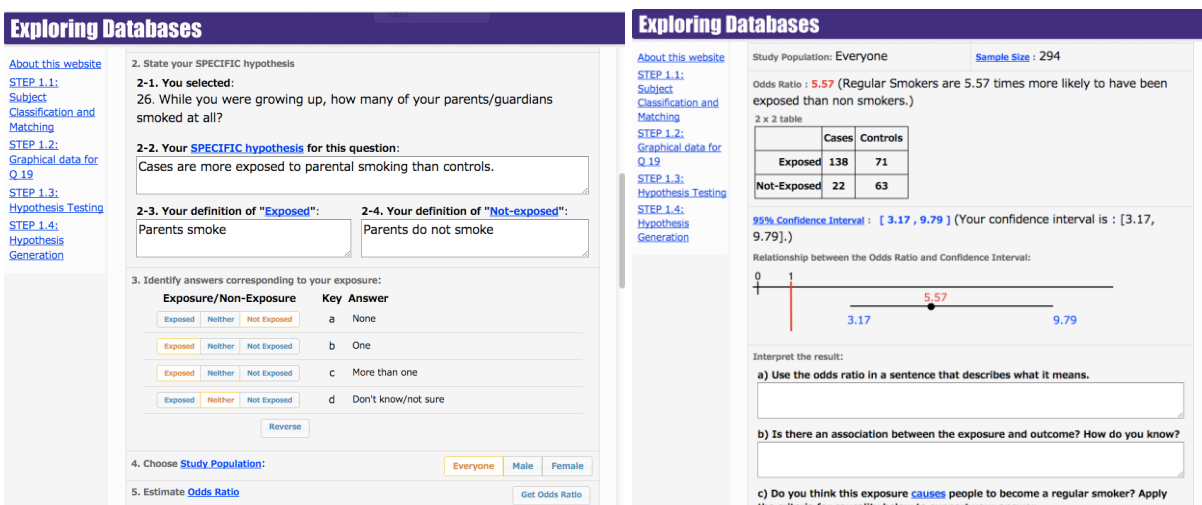


Figure 3: Scaffolding for statistical test development (left) and interpretation with visual statistics (right).

A Case Study: Examining Students' Engagement in the Epidemiologic Inquiry

The foregoing sections discussed the educational value of disciplinary engagement in authentic practices as a way to develop reasoning and introduced the ExDa unit along with its scaffolding design in the WISE learning environment and instructional materials including both discipline-focused and process-focused scaffolds. In the following, I report results of a case study conducted to demonstrate students' disciplinary engagement and development in epidemiologic reasoning. The present study specifically examines: (a) whether and how our scaffolding design with research worksheets and the database interface guided students' disciplinary engagement in epidemiologic analysis, and (b) whether and how well students developed their epidemiologic reasoning with the focus on their population-based reasoning.

Method

Twenty one high school students in an AP Biology class at a suburban public high school in the Pacific Northwest participated in the study. The biology teacher has a background in

public health and had a deep understanding of the unit as she had participated in our professional development workshop prior to the study. Students engaged in a total of 13 sessions (50 minutes per class) for the unit, and they engaged in foundational learning to develop conceptual knowledge in smoking and addiction, and epidemiologic inquiry including case control design and statistical analysis through the first nine sessions. The following three sessions were spent on students' conducting their own investigation with each group presenting their study and findings at the last session.

Video/Audio and worksheet analysis to examine student engagement in epidemiologic inquiry. To examine student engagement in the epidemiologic inquiry, one video-recorded group and three audio-recorded groups were analyzed. Analysis focuses on the video-recorded group (three male students) and presents transcripts for research-related conversations during the three sessions for students' own investigation to discuss how our scaffolding guided student engagement in the inquiry, followed by similar or distinctive observations from other groups. Specifically, the analysis discusses four segments when students were working on: (a) learning the smoking database in which they learn the nature of data including the study procedure, (b) hypothesis development in which they identified their research topic, (c) statistical test development and interpretation in which they developed a statistical test to estimate the strength of association for an exposure and its statistical significance, and interpret the results, and (d) causal inference in which they discussed their findings to draw their conclusions with epidemiologic criteria. In addition, worksheets were collected from 15 students and these were used to evaluate my interpretations to these recorded observations.

Analysis of written assessments to examine student development in epidemiologic reasoning. Students answered pre- and post-surveys with three open-ended tasks designed to assess the quality of their epidemiologic reasoning (Table 2). All the open-ended tasks took a form of “position taking” that presented two different positions in a decision-making context with relevant data presented from the same database, and it asked students to support either of the positions with some warrant and rationale. In assessment, either of the positions is not necessarily correct and therefore students’ warrant and rationale was assessed to examine the quality of their epidemiologic reasoning based on criteria designed along with learning goals in the unit (Table 3).

Specifically, the first open-ended task (EPI-1) presents two positions; Sarah concludes that “feeling buzzed” (when smoking) can be a risk factor of smoking based on the significant association, whereas Tom disagrees based on his rationale that everyone does *not* necessarily smoke if he/she felt a buzz. Sarah’s position emphasizes population-based reasoning, whereas Tom’s position does case-based reasoning (Table 2). This task lays weight on the population-based position along with the epistemic aim, and student response was assessed along with seven criteria: significant association, strength of association, probabilistic reasoning (using probabilistic expressions such as “more likely”), biological plausibility (e.g., physiology, genetics, psychology), indirect cause (one or more intervening factors), and (lack of) confounder (i.e., confounding variable). The second task (EPI-2) is an analogous item to assess students’ weight between population-based and psychological (one’s own will) positions with the same criteria (Table 2).

The third task (EPI-3), on the other hand, presents three different positions and aims to assess student reasoning on generalizability assessment (Table 2). Specifically, Nancy claims

Chapter 3

that the finding can be applied to the study population, Fraser claims that it can be applied only to the people who participated in the study, and Jose claims that it can be applied to other populations rather than the study population only. Fraser's position is the most conservative with respect to the generalizability, and his claim could be justified by pointing out that the sampling strategy was voluntary and it may be biased with some rationale in the study procedure. Nancy's position is based on statistical representativeness, and it could be justifiable if the study sample were not so biased with some rationale to the study population. Jose's position is the most extensive among the three, and it could be justifiable if both statistical and biological representativeness were taken into account with some rationale, for example, by pointing out some similarity of the study population with other populations. Note that there is no single correct answer in the problem context, and the analysis rather looks at the rationale and warrant in student justification.

In coding of student responses to these tasks, two evaluators, an epidemiologist and I first scored student responses individually. In scoring, we assigned one point to each criterion if a statement sufficiently satisfied the criterion (Table 3); assigned half point (= 0.5) if it partially satisfied; and zero if it was erroneous, insufficient, or no description. After individual coding, we compared the scores for each criterion and discussed until all agreed with the same scores when they were different.

In addition to these tasks, students also answered a set of multiple-choice and open-ended items to assess their declarative knowledge in the nature of data in the database, case-control design, and statistical analysis with OR calculation and interpretation, and CI interpretation tasks. Respectively, six multiple-choice items were presented to assess student understanding in the nature of data in the database—the study population, definitions of cases and controls, the target

phase in stages of smoking (maintenance), and the sampling procedure. 10 multiple-choice items were presented to assess student understanding in the declarative knowledge of case control design—study population, outcome, case, control, exposure, the retrospective nature, epidemiologic criteria, the purpose of matching, and systematic biases (two items). Finally, 2 multiple-choice items were presented to assess student understanding in the declarative knowledge of sample size and the nature of odds ratio (likelihood, not probability), and 4 open-ended items were presented to assess student skills in OR calculation and interpretation, CI interpretation, and the conceptual meaning of “statistically significant.” All the scores for correct answers were valued as one point except that two points were assigned to the last item on statistical significance.

Analysis first compares pre and post scores to examine student improvement in the quality of their epidemiologic reasoning and declarative/conceptual understanding in the epidemiologic inquiry. Then it discusses how our scaffolding design and student engagement led to the development.

Table 2

Open-ended Items to Reveal Epidemiologic Reasoning

Assessment Item
EPI-1: Population-based Explanation Task against a Counterargument Based on Exception Sarah came up with a hypothesis that feeling a buzz when one tried smoking can be a risk factor for regular smoking, and tested her hypothesis with the same database. She identified a question “During your experimental smoking phase, did you experience any of the following regularly? A buzz”, and defined “Yes” as exposed and “No” as not-exposed. The odds ratio was 6.33 and the confidence interval was [3.33, 12.02]. She applied the criteria for causality and concluded that feeling a buzz can be a risk factor for regular smoking. However, her teammate Tom disagreed and claimed that everyone does NOT necessarily smoke even if he/she felt a buzz. Whose position do you support? Explain why and refer to each position in your explanation.

EPI-2: Population-based Explanation Task against a Counterargument for Personal Decision

Andrew came up with his hypothesis that having close friends who smoke can be a risk factor for regular smoking, and tested his hypothesis with the same database. He identified a question “During your experimental smoking phase, did any of your closest friends smoke cigarettes?”, and defined “Most or all” as exposed, and “Some” and “None” as not-exposed. The odds ratio was 3.61 and the confidence interval was [2.07, 6.3]. He applied the criteria for causality and concluded that having close friends who smoke can be a risk factor for regular smoking. However, his teammate Kari disagreed and claimed that the cause is one’s will, since people can make their own choice for smoking. Whose position do you support? Explain why and refer to each position in your explanation.

EPI-3: Representativeness Task

Three students (Nancy, Fraser, and Jose) tested their hypothesis that having parents who smoke can be a risk factor for regular smoking. They identified a question “While you were growing up, did any of your parents/guardians smoke inside your home?” and defined “Yes” as exposed, and “No” as not-exposed. The odds ratio was 5.2 and the confidence interval was [3.12, 8.67]. They applied the criteria for causality and concluded that their hypothesis is valid, but they wondered to whom they could apply their results. Knowing that the people who participated in the study were all from the Puget Sound area, Nancy claimed that their results can apply to the people who live in this area, while Fraser claimed that their results should only apply to the people who participated in the survey. Jose claimed that their results can apply to other populations as well. Whose position do you support? Explain why and refer to each position in your explanation.

Table 3

Assessment Criteria for Epidemiologic Reasoning Tasks

Assessment Criteria	Description
EPI-1 & EPI-2: Population-based Reasoning Task 1 & 2	
a) Significant association	Articulating the condition for statistical significance (e.g., one is outside the interval)
b) Strength of association	Taking into account the size (e.g., large or small) of association
c) Probabilistic reasoning	Articulating some statistical/probabilistic statement(s) (e.g., more likely, higher probability) rather than deterministic judgment (e.g., all or nothing)
d) Discriminating causation from association	Articulating that the OR indicates association, not directly causation, but also seeking some causal explanation (e.g., risk)
e) Biological plausibility	Articulating some biological explanation(s) for the risk factor
f) Indirect cause	Taking into account some indirect cause/effect(s)
g) (Lack of) confounding	Articulating some confounder(s) or why it is not the case

EPI-3: Representativeness Task

a) Study population	Articulating the study population (= the Puget Sound area)
b) Sampling procedure	Articulating data representativeness based on random or voluntary sampling
c) Statistical representativeness	Articulating statistical representativeness based on the study population
d) Biological representativeness	Articulating biological representativeness based on the similarity or dissimilarity with other populations

Analysis of Students' Interactional Sense-Making

In the following, I will present four segments of classroom interaction coordinated with the model of epidemiologic inquiry (learning the smoking database, hypothesis development, statistical test development & interpretation, and causal inference) and discuss how our scaffolding design guided students' disciplinary engagement with a focus on population-based reasoning. Analysis reports on interactional segments of three male students (Jeff, Ted, and Dave) whose group worked to investigate whether lack of education can be a risk factor of smoking. Note that most of the utterances are from Jeff and Ted and Dave did not participate in the talk as frequently. In short, these four segments suggest that a set of paper "worksheets" and the database "interface" successfully guided student engagement in the epidemiologic inquiry and it appears that these can also apply to other audio-recorded groups with some notable observations.

a) Learning the smoking database. Before entering their own project, students in teams worked on a few worksheet tasks to learn about the smoking database. First, they read study procedures for the original case control study and confirmed the epistemic aim of the study and key conditions including the study population and case and control definitions. Next, students explored the database and checked the demographic imbalance between cases and controls by counting the numbers separating by gender, age, and ethnical backgrounds. Then, students

Chapter 3

worked on an exercise of estimating odds ratio and confidence interval with the database. It is an important question to confirm whether and how students were able to conduct statistical tests given the scaffolds, and the following transcripts show a segment in which Jeff and Ted are talking about the decision making for statistical significance.

- J: Odds ratio is 4.08.
T: So it's bracket. So, the odds ratio is significant. So yes, because it doesn't include, CI doesn't include one.
J: Yeah, I just... so do you conclude that the exposure can be a risk factor for smoking?
T: Most probably, yes.
J: So I think yes, yeah? You say?
T: Oh, yeah, because I mean the odds ratio is a powerful one showing that there is a correlation and then with the CI. It's still showing a positive correlation towards the effects of that exposure.
J: Surely showing.

Looking at the results of odds ratio and confidence interval on his monitor, Ted correctly interpreted the odds ratio was significant based on the criterion that “it” (= the confidence interval) does not include one, null hypothesis suggesting no association. Subsequently, Jeff asked if the exposure could be a risk factor of smoking and it appeared Ted agreed with it. Although this finding is actually a little premature to conclude based on only one significant association. Ted's statement “the odds ratio is a powerful one” reflects the high odds ratio (=4.08), which corresponds to the strength of association in epidemiologic criteria. On the other hand, it is rather uncertain whether they understood the conceptual meaning of statistical significance correctly based on their interactions.

Other groups also demonstrated similar conversations in the judgment of statistical significance. Although some students showed their lack of confidence or confusion in the statistical criterion at the beginning, they became able to draw correct judgments on statistical significance based on the location of confidence interval to one (or the location of one to the

interval) by consulting their worksheets and/or through interactions with their peers in their small groups and occasionally other groups next to them. In addition, analysis also identifies a moment in which two female students in another group were talking about case and control definitions on worksheets and this conversation led them talk about the epistemic aim (“to find factors of who is persistent smokers”) and the social aim (“and then how they can stop it early”). On the other hand, video/audio analysis indicates that some students seemed to mistakenly interpret that the sampling procedure for the case control study was random, though the worksheets explicitly describes that it was voluntary. These students’ worksheets confirmed this concern and the scaffolding might not explicitly allow some students pay attention to the sampling procedure.

b) Hypothesis development. After the exercise, the focal group started discussing their own project and developed their overarching hypothesis to guide their inquiry. The following transcripts show a relevant segment, and it demonstrates that the worksheets guided their hypothesis development.

- T: And there is how are, you are in your teen. So I mean like part of interesting what you were as a teen. Kinda... questions? There is a lot of... so we could maybe signal (inaudible). What?
- J: We could do like out of this education and like this we could also talk about are... like in high school we talked about dangers of smoking and how you in school and what you call about side effects. So just kind of talk about how education affects likeliness to smoke.
- T: Maybe
- J: So, I think we are gonna do for... I think how education affects likeliness of somebody with smoke, so we should... think of... so.
- T: That’s a question riding on hypothesis.
- J: What?
- T: (Inaudible) Hypothesis. That’s a question.
- J: Yes, it’s kind of a less education the more likely somebody smoke.
- T: Yeah.
- T: Because they are not aware of the dangers of smoking because they may not be fully aware of the dangers of smoking?
- J: Yeah.

Chapter 3

Ted points out that there are several questions asking what survey respondents were like in their teens. Looking at database items, Jeff suggested education could be related to smoking with his hypothesis that lack of education could be a risk factor of smoking by drawing on his experience (“like in high school we talked about dangers of smoking”). He then wrote down something assumed to be research questions on his worksheets, while Ted pointed out that Jeff’s description is a form of question, not hypothesis, as worksheets instruct them to write their overarching hypothesis. To his comment, Jeff rephrased it as “a less education the more likely somebody smoke,” followed by the rationale that people might not know the dangers of smoking without education. Starting from his initial conjecture in the effect of education on smoking, his description developed into a more precise hypothesis that could be tested with the database. Subsequently, they explored the database and identified five relevant items to test their hypothesis (Figure 4). The transcripts and the figure suggest that worksheets guided their hypothesis development and they could identify relevant items on the online database.

- b. Identify relevant 4-5 questions from the questionnaire and name the specific exposures in relation to your overarching hypothesis

Exposure	Related Question
Example: (your answer to j.)	No: 50 Exposed: b Not-exposed: c, d
Not taught about dangers of smoking	No. 41 Exposed B Not-exposed A
Not discussing why people smoke	No. 42 Exposed B Not-exposed A
Not taught about side effects of smoking	No. 43 Exposed B Not-exposed A
Not receiving adequate info about smoking and risks	No. 72 Exposed B Not-exposed A
Not reaching college	No. 80 Exposed A, B Not-exposed C, D, E

Figure 4. Database items selected to test students' (Jeff, Ted, and Dave's) hypothesis (lack of education) in their research worksheets

As for other groups, one group was interested in how stress affects smoking and discussed how stress can make people smoke by applying biological explanations (biological plausibility) with the function of dopamine from previous foundational learning. On the other hand, another audio-recorded group had to change their hypothesis, as they could not identify relevant variables in the database to test it. Specifically, two female students of the group developed their interest in socioeconomic status and searched the database for relevant items. However, they could not identify relevant items to investigate their hypothesis and changed their research topic into life style. They discussed that people with a healthy lifestyle (e.g., healthy food, regular exercise) would be less likely to smoke and they developed their hypothesis that a healthy lifestyle can be a “protective” factor instead of risk factor of smoking.

c) Statistical test development & interpretation. Having determined education as their research topic, the three focal students identified relevant items to their hypothesis and developed statistical tests to estimate the odds ratios for specific exposures. The following transcript shows an interaction in which they were developing a statistical test for the exposure to an item “When you were in school, were you taught about the dangers of smoking (for example, lung cancer, heart disease, fire hazards)?” on the database.

T: Okay.

J: So what are we gonna do. So starting with 41 [= When you were in school, were you taught about the dangers of smoking (for example, lung cancer, heart disease, fire hazards)?]??

T: Dave, we were thinking about questions 41 through 43 [= When you were in school, were you taught about the side effects of smoking, such as making teeth yellow, causing wrinkles, or making smokers smell bad?] and then...

Chapter 3

- J: Eighty [=What is the highest grade or year of school you have completed?] and 72 [= When you were a teen, do you think you received adequate information about smoking and its risks?]?
T: Yeah.
J: So we need to know... so for 41, we will stay with that. When you're in school...
T: When you're in school you talk about the dangers of smoking, so...
J: So 'no' would be exposed and not exposed would be 'yes.' So, not exposed... and so exposure would be... not [as he writes on worksheet]...
T: What?
J: Our, we... do... would be getting smoking education and not getting smoking education. And we are on 41.

Having decided to test the exposure in absence of smoking education with Item 41 (“When you were in school, were you taught about the dangers of smoking (for example, lung cancer, heart disease, fire hazards)?”), Jeff defined “No” as exposed and “Yes” as not-exposed on the database. These are reasonable definitions given the item statement – the item asks whether respondents had some education in the dangers of smoking, and those who answered with “No” would not have such education and be thusly exposed; those who responded with “Yes” would be not-exposed. This suggests that the database interface helped students develop their statistical test focusing on theoretical reasoning.

This applies to other groups and they could develop tests without technical problems. On the other hand, some students expressed their confusion in which answer key (e.g., yes or no) should be defined as exposed or not-exposed for their hypothesis and how to interpret the result particularly when the ORs are less than one. For example, the audio-recorded group of two female students with the hypothesis on healthy life style defined exposure and non-exposure in the reverse manner and later found some inconsistency with their hypothesis. Specifically, they developed a statistical test along with their hypothesis that a healthy life style can be a protective factor of smoking (reversely, an unhealthy life style can be a risk factor). However, since they had mistakenly flipped the definitions for exposure and not-exposure, the odds ratio suggested a

contradictive result against their hypothesis and they became confused in the interpretation. Later, they called me and we figured out why it did not make sense and they re-estimated the odds ratio with a correct definition. In addition, another group of students were confused with their result with $OR < 1$ at one point and their subsequent discussion by flipping exposed and not-exposed definitions led them deepen their understanding in the concept of exposure and odds ratio. These observations indicate that the data analysis activity promoted student's sense making as they worked on analytical tasks, which led them to develop their conceptual understanding of (non-) exposure and relevant statistics.

d) Causal inference with epidemiologic criteria. After estimating odds ratios and confidence intervals for a few exposures, they finally worked on summarizing these results to prepare slides for their presentation. The following transcript shows an interaction in which students engaged in causal reasoning with epidemiologic criteria in the worksheets.

- T: So that it?
- J: Okay. So association with causality.
- T: So the larger the odds ratio the stronger the association.
- T: The association is extremely strong with this.
- J: Definitely yes, especially with question 80 [= What is the highest grade or year of school you have completed?].
- T: Oh definitely question.
- J: And then...
- T: Eighty, 41 [= When you were in school, were you taught about the dangers of smoking (for example, lung cancer, heart disease, fire hazards)?], and 72 [=When you were a teen, do you think you received adequate information about smoking and its risks?], yeah.
- J: I just put a note at the bottom that all these exposed men not educated.
- J: I mean we can kind of build one that (inaudible). So, strength of association, dose response, we can either show this temporal sequence.
- T: The exposure must occur before the outcome?
- J: I think that's really.
- T: No.
- J: We don't really know that's situation dependent.

- T: Because even like you saw people, people that were probably... like people that have a bachelor's degree and probably smoking (inaudible).
- J: Do you know, I think it situation dependent. It's what we were saying with that. You don't really... because it said sometimes in case control study, this can be difficult to verify, only...
- T: But like someone may drop out of school before he starts smoking.
- J: Or somebody can start smoking like elementary school before they get any smoking education. So, let's just say...
- J: Let's just say it's situation dependent. I think so consistent with other studies we can just show this.
- J: Biological plausibility, I think it's plausible and then.
- T: It's biological.
- J: Like on confounder, significant bias, didn't we mention in here? But there was like it could have something to do with like... wealth.
- T: Yup.

In this scene, Jeff mentioned epidemiologic criteria as he checked worksheets and the subsequent conversation was devoted to their sequential evaluation of their results. Ted restated the criterion that exposure must occur before the outcome, and Jeff claimed that it is situation dependent or difficult to confirm it, followed by some hypothetical scenarios that may not satisfy the temporal sequence criterion. Then, they mentioned the consistency with other studies by saying “we can just show this” and “this” refers to the prior studies they had found on Google Scholar. Jeff also mentioned biological plausibility and Ted appeared to agree with it, although the specific explanation was not explicitly stated. Finally, Jeff mentioned lack of confounder (“confounder, significant bias”) and raised wealth as a potential confounder and Ted agreed with it. The transcripts suggest that they engaged in causal reasoning along with epidemiologic criteria and the worksheets guided students' disciplinary reasoning in drawing their conclusion. Another group also demonstrated similar conversations along with epidemiologic criteria including literature hunting with Google Scholar, and these indicate that worksheets guided these students' causal reasoning in the epidemiologic discipline.

In sum, these observations suggest that the set of paper worksheets and the database interface guided students' disciplinary engagement in epidemiologic reasoning including statistical analysis. Specifically, the transcripts of all segments indicate that worksheets guided the flow of epidemiologic investigation and reminded students of analytical guidelines/criteria in each phase (learning the smoking database, hypothesis development, statistical test & interpretation, and causal inference). In the statistical analysis phase, the database interface in the WISE environment provided scaffolds for statistical test development and interpretation with its test-setup guidance and figures of statistical data (odds ratio and confidence interval) as shown in Figure 3, and students could analyze statistical data focusing on theoretical reasoning without technical problems. In addition, the analysis also identifies some notable observations in which some students leveraged their personal experience into a scientific question and hypothesis; some developed their conceptual understanding of exposure and statistics through sense-making confusing or contradictory results. On the other hand, it also highlighted that some students had to change their hypothesis, because they could not find relevant items on the database.

Assessment Results

Twenty students answered pre and post surveys, since one student was absent at pre and another student absent at post. The overall statistics for those who answered both surveys ($N = 19$) are shown in Table 4 below, and only the post result is presented in terms of database items that asked about the nature of data (e.g., voluntary sampling, study population), as these are specific to the unit and therefore students were asked to answer them only at post.

By-student analysis ($N = 19$). In conceptual knowledge items, at pre, students had little knowledge in the unit before entering the unit except that students had gained a few points ($M = 3.63$, $SD = 2.45$) in case control design. Closely looking at these items, more than half of the

Chapter 3

students correctly answered the study population, cases, and controls, which implies that students may have known some related concepts from their prior learning or inferred the definitions from the item statements. At post, students correctly answered 86% (5.16 out of 6 points) in database items about the nature of data ($M = 5.16$, $SD = 0.90$), 74% (7.37 out of 10 points) in case control design ($M = 7.37$, $SD = 1.83$), and 65% (4.53 out of 7 points) in statistical analysis ($M = 4.53$, $SD = 1.27$). Paired samples t-tests found significant differences in both case control design ($t(18) = 4.99$, $p = .000$) and statistical analysis ($t(18) = 10.46$, $p = .000$) between pre and post. Both effect sizes in case control design (Cohen's $d = 1.15$) and statistical analysis ($d = 2.40$) are bigger than one, and these results suggest that students significantly developed their basic knowledge in the nature of data in the unit, case control design, and statistical analysis (i.e., using odds ratio and confidence intervals) through the curriculum experience.

In epidemiologic reasoning tasks, student responses met few criteria ($M = 0.87$, $SD = 1.01$) at pre, and in fact most of the students did not provide any description. At post, students gained on average 4.66 points ($M = 4.66$, $SD = 1.86$) out of three tasks, and there was a significant difference between pre and post ($t(18) = 7.34$, $p = .000$) with a large effect size ($d = 1.68$). Closely looking at each task, the post results show that student responses on average met one partial and one full points per task (EPI-1: $Mdn_{partial} = 1$, $Mdn_{full} = 1$; EPI-2: $Mdn_{partial} = 1$, $Mdn_{full} = 1$; EPI-3: $Mdn_{partial} = 1$, $Mdn_{full} = 1$). These indicate that students' epidemiologic reasoning significantly improved, and they provided on average two reasonable warrants or rationale in their response per task. In all of these cases, the strong gains resulting from a relatively modest educational intervention (13 classes x 50 minutes of instruction and the student-led investigation) implies that the kinds of conceptual learning and epistemic practices associated with this form of complex reasoning are possible with students within this context and

with their backgrounds. These results open up possible new work to bring this approach to a broader variety of settings and perhaps age groups.

Table 4
Comparison of Pre- and Post- Results (N = 19)

Category	Pre			Post		<i>t</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>
	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Database	6	-	-	5.16	0.90	-	-	-	-
Case Control Design	10	3.63	2.45	7.37	1.83	4.99	18	.000	1.15
Statistical Analysis	7	1.05	0.60	4.53	1.27	10.46	18	.000	2.40
Epi. Reasoning	-	0.87	1.01	4.66	1.86	7.34	18	.000	1.68

By-criterion analysis of epidemiologic tasks (N = 20). Closely looking at position for the EPI-1 task in pre results, six students explicitly supported the case-based position (Tom), and only one student supported the population-based position (Sarah), although few students provided any justificatory statement. At post, 15 (75%) students supported the population-based position, two (10%) supported the case-based position, two (10%) supported both positions or support neither, and one student did not have an explicit position. By criterion (Figure 5), 70% students applied the statistical significance as part of their warrant; 55% applied the strength of association; 45% demonstrated probabilistic reasoning with some probabilistic expressions (e.g., more likely, higher probability); 25% explicitly discriminated causation from association; and 15% mentioned biological plausibility in their justification. These results suggest that most students applied their causal reasoning with statistical evidence after the unit, and about half of the students articulated probabilistic statements to support the population-based position. In contrast, they also show that none of the students applied indirect cause and (lack of) confounding in their justification (Figure 5), which can be interpreted by the object of counterargument for this task. Specifically, the counterargument points to the case-based claim

Chapter 3

(“everyone does not necessarily smoke”) and a rebuttal should be made based on statistical evidence with the significant high odds ratio ($= 6.33$), rather than the matter of indirect cause or lack of confounding.

In the EPI-2 task, at pre, 6 students supported the statistical position (Andrew), 4 students supported the case-based position (Kari), and one student supported both positions. At post, 19 out of 20 students supported the population-based position, and one student supported both positions. By criterion (Figure 6), 55% students applied the statistical significance; 40% students applied the strength of association; 40% discriminated causation from association; 35% demonstrated population-based reasoning; 20% mentioned biological plausibility, 10% reported indirect cause, and 5% used lack of confounder in their justification. Although the percentage for applying the statistical evidence is lower than the percentage in EPI-1 (Figure 5), this difference would also derive from the different object of counterargument. That is, while the counterargument for EPI-1 was a case-based claim (e.g., supporting a position based on a few observations) and therefore it made sense to make a rebuttal based on statistical evidence, the counter argument for this task is rather biological or psychological (“one’s will”), and therefore they would lay weight not only on statistical evidence but also on biological plausibility, indirect cause, and lack of confounder. Yet, only a few student responses still applied the indirect cause and lack of confounder, and this is likely due to the fact that no specific information was provided in the task statements. These highlight additional areas for future research and development work.

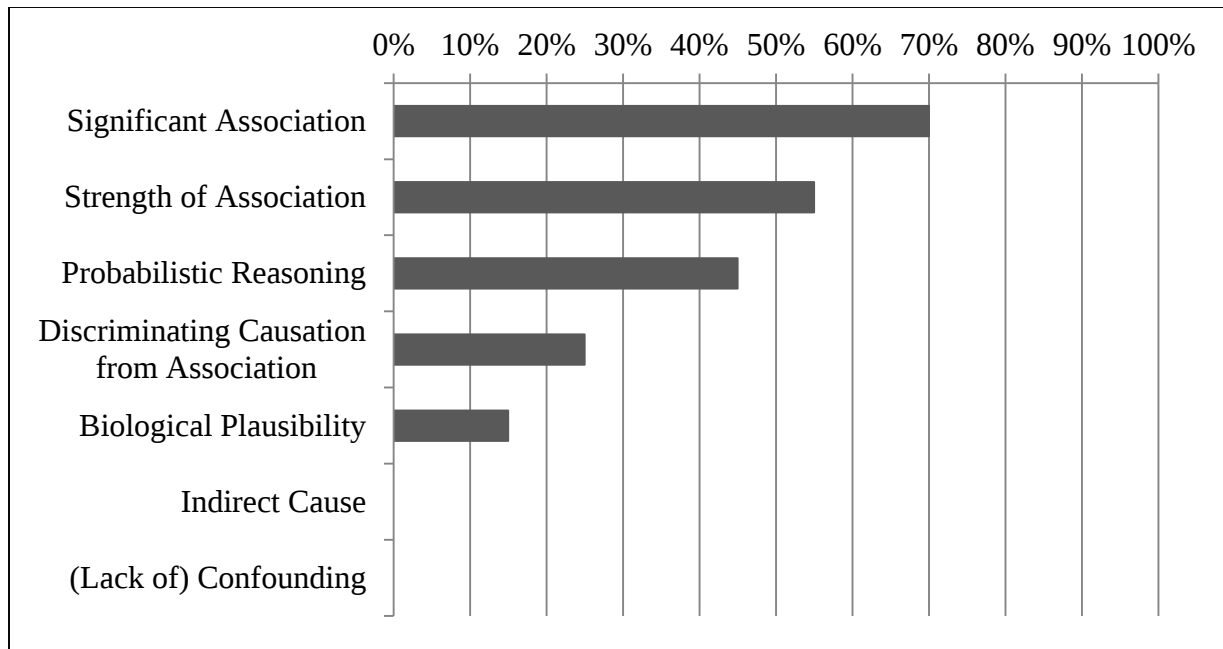


Figure 5. By-criterion results for the first population-based reasoning task (EPI-1)

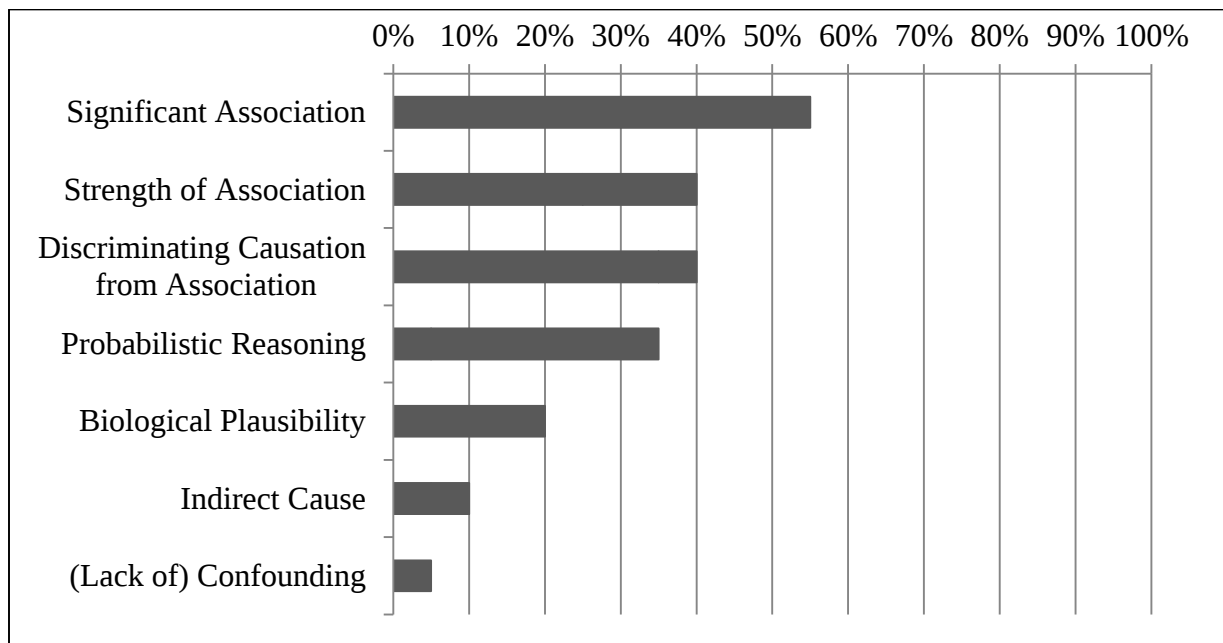


Figure 6. By-criterion results for the second population-based reasoning task (EPI-2)

As for the EPI-3 task, the pre result shows that 3 students supported the sample-only (Fraser's) position, one student did the study-population (Nancy's) position, and 4 students used

the extensive (Jose's) position. At post, 2 students supported the only-sample position, 6 students did the study-population position, and 9 students supported the extensive position. For others, one student supported both Nancy's and Jose's positions, one student supported neither of the positions, and the others were indecisive.

While there is no single correct answer for the generalizability assessment task, how did students reason on the generalizability assessment and which position of students demonstrated complex reasoning between statistical and biological representativeness? The by-criterion results for those that gained partial or full scores with either of explicit positions ($N = 17$) are shown in Figure 7 below to examine the questions. The figure shows that all students who supported the only-sample or study-population position articulated the Puget Sound area as the study population, whereas 44% students who supported the extensive position did the study population. Next, all the students who supported the only-sample position pointed out that the sample procedure for the case control study was not random but voluntary as their rationale. 67% of the students in the study-population position and 44% in the extensive position incorporated statistical representativeness in their rationale, whereas no one in the only-sample position incorporated the rationale in their justification. Finally, 50% of students in the only-sample position and 83% in the study-population position incorporated biological representativeness by pointing out (potential) dissimilarities of the study population with other populations. In contrast, only 22% of the students in the extensive position incorporated the rationale by pointing out (potential) similarities with other populations.

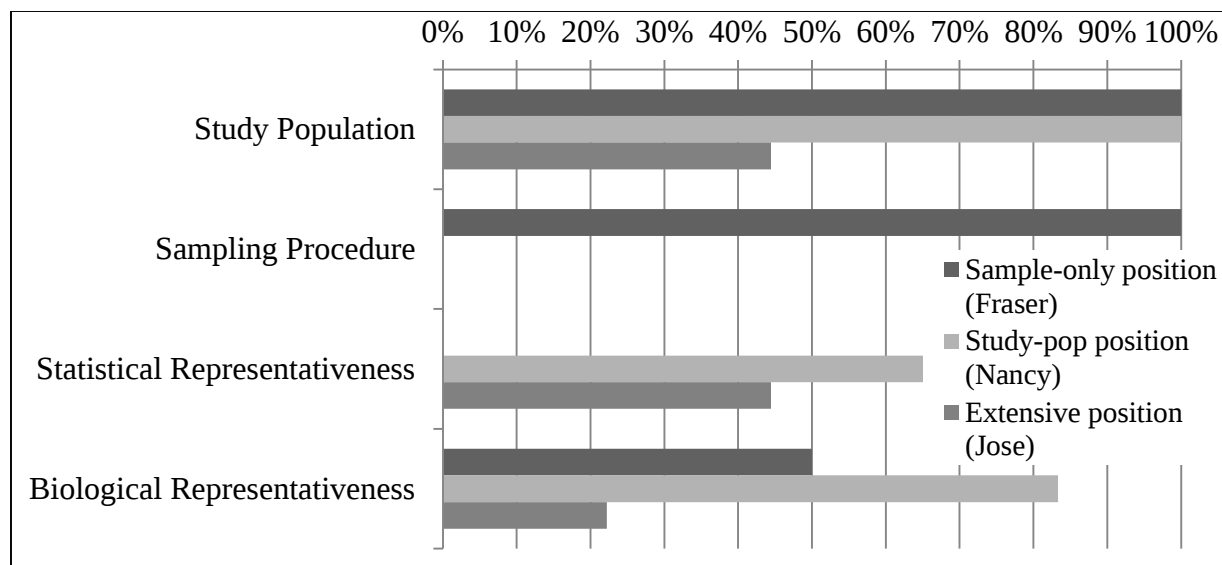


Figure 7. By-criterion results for the generalizability assessment task (EPI-3)

These post results suggest that although their patterns more varied than other tasks, students applied some warrants and rationale to support their position (Figure 7). Overall, students in the only-sample position were more likely to articulate the study population and point out that the sampling procedure was voluntary in their justification to restrict the generalizability to the study participants. Students in the study-population position were also more likely to articulate the study population and apply statistical representativeness to support their position and half of these students pointed out (potential) demographic dissimilarities of the study population with other populations as their rationale to restrict the generalizability to the study population. In contrast, those who supported the extensive position likewise applied statistical representativeness in their rationale, whereas only a few articulated biological representativeness by pointing out (potential) similarities with other populations. This is counterintuitive since biological representativeness is essential to justify the generalizability to broader populations,

and this indicates that students in the position were less engaged in theoretical reasoning on the generalizability assessment.

In sum, written assessment analysis suggests that students significantly developed both conceptual knowledge in the epidemiologic investigation as well as epidemiologic reasoning on population-based reasoning, whereas some more deliberate scaffolds may be needed to develop their critical reasoning on generalizability assessment. For the two population-based reasoning tasks (EPI-1 & EPI-2), some students supported case-based positions at pre, while most students supported population-based positions at post by applying statistical evidence as warrant. Moreover, about half students demonstrated some explicit expressions (e.g., more likely) for probabilistic reasoning to the first task (EPI-1). As an exemplar, one student wrote:

I support Sarah because her CI did not include one and showed that there was a 6.33 times more chance of the people becoming a maintenance smoker. While Tom may be right with his idea that everyone doesn't feel a buzz, but her studies show that there is a more likely chance of continuing to smoke when feeling some sort of buzz of comfort feeling. Because her numbers were high, this shows there is a correlation between feeling a buzz and maintenance smoking.

His/Her statement admits that not everyone necessarily continues smoking but still makes a rebuttal that the high odds ratio ($= 6.33$) suggests a statistically significant correlation between feeling buzzed and maintenance of smoking by using probabilistic expressions such as “more likely.”

On the other hand, results for the generalizability assessment task (EPI-3) suggest that student positions more varied in their written responses and 6 students who supported the population-based position appeared to demonstrate more complex reasoning than those in the

other positions. In contrast, students who supported the sample-only position appeared to more incur case-based reasoning, and those who supported the extensive position tended not to incorporate a rationale pointing to the similarity of the study population with other populations. These results suggest that some more deliberate scaffolding may be needed to develop student reasoning on generalizability assessment.

Discussion

The analytical goal of this case study was to examine: (a) how students engaged in the epidemiological inquiry given designed scaffolds, and (b) how well their epidemiological reasoning developed through disciplinary engagement in that inquiry. The following section discusses the results of video/audio and assessment analyses with respect to these two goals.

Students' disciplinary engagement in the epidemiological inquiry. Were students able to engage in some reasonable form of epidemiological inquiry? Video/Audio analysis suggests that designed scaffolds with research worksheets and the smoking database guided students' disciplinary engagement in the epidemiologic inquiry. The analyzed interactions demonstrated that students were able to develop their hypothesis, develop statistical tests and interpret the results, and make causal inferences as they consulted with research worksheets. Although some students were confused by the procedures at the beginning of the instruction, the combination of scaffolding in worksheets, the database interface in the WISE environment, and interactions with peers and instructors about the tasks led them to recognize their misunderstanding and revise their conceptual understanding in the course of inquiry. With respect to engaging in statistical analysis, students appeared to develop statistical tests without any major problems and this indicates that the database interface made the technical procedure transparent and allowed students to focus their queries on theoretical reasoning. Although our observations are limited to

four groups, about half the students in the entire class, we did not observe any case in which students struggled in figuring out the technical procedures, and I argue that our scaffolding design overall guided students' disciplinary engagement in the epidemiologic inquiry with some revision points. On the other hand, however, students engaged in only one cycle of the inquiry because of the limited time for which the class could spend to engage in the unit, and there was no explicit "fading" phase in the instruction (Pea, 2004). Therefore, the present analysis could not confirm whether students were able to internalize the process of inquiry. Student development in conceptual understanding and epidemiological reasoning as gauged through the assessment is discussed next.

Student development in epidemiological reasoning. What did students learn about the conceptual and epistemological aspects of epidemiological inquiry? Survey analysis suggests that students developed conceptual knowledge related to the epidemiologic inquiry and about half of the students demonstrated a sense of population-based reasoning through the unit. In their post responses, a majority of students incorporated statistical evidence for association in their justification. Although this would appear self-evident as they did the same activity through the unit, about half of the students explicitly used probabilistic expressions such as "more likely" to justify their position to the counterargument for case-based reasoning in the task about taking either position of supporting population-based or case-based conclusion (EPI-1). With respect to another analogous task taking either position of supporting population-based or psychological conclusion (EPI-2), student justification slightly shifted to pointing out psychological or mechanistic explanations, and this seemed to reflect on the corresponding counterargument for personal (one's will) decision in addition to the population-focused rebuttal. These suggest that students developed their claim along with epidemiologic criteria compared to the pre in which

few students provided any description for their justification, and they arguably engaged in complex reasoning by weighing both statistical warrant and biological rationale corresponding to the nature of counterarguments.

Student responses to the generalizability assessment task about taking either of sample-only, study population, or extensive position for data representativeness (EPI-3) suggest that students demonstrated some critical thinking in the generalizability of their findings, although the patterns varied. Closely looking at individual statements, students who supported the sample-only position justified their position based on the voluntary sampling procedure, whereas their justification indicates weighing case-based reasoning. Those who supported the study-population position were more likely to incorporate statistical and biological representativeness in their justification, and most of them explicated the diversity of population characteristics as their rationale of limiting the generalizability to the study population. Although the study-population position demonstrated critical reasoning in the generalizability, on the other hand, only a few students who supported extensive position explicated the biological similarity (biological representativeness) between populations as their rationale for their argument in the generalizability. Two students in the position indicated that statistical significance means it can apply to “other” populations, these students might have misunderstood the concept of confidence interval. In this viewpoint, students in the study-population position demonstrated more complex reasoning in their justification than the other two groups.

Student development in statistical reasoning. From the informal statistical inference (ISI) viewpoint, finally, I argue that the ExDa unit served to develop students’ informal statistical reasoning. The survey analysis showed most students applied statistical evidence and about of the half students used probabilistic language to support the population-based position to the first

task (EPI-1), although their responses to EPI-2 slightly shifted to a biological/psychological rebuttal (biological plausibility, indirect cause, and lack of confounding) because of the counterargument based on “one’s will.” These results indicate that students focused on properties of aggregates for the study population rather than a few single cases (Paparistodemou & Meletiou-Mavrotheris, 2008). Although student evaluation of generalizability needs more effort with better scaffolding, the present study arguably developed some sense of informal statistical reasoning through engagement in the epidemiologic inquiry. Note that students who participated in the study were in AP Biology and might have had some prior knowledge in statistics, whereas the pre survey results suggest that few students demonstrated their reasoning in the epidemiologic context before entering the unit. From a general viewpoint, together with all the results, I argue that it is feasible to develop students’ informal statistical reasoning through disciplinary engagement in authentic scientific practices with deliberate scaffolding.

Limitations with the generalizability assessment task. Together with all three tasks, I argued that the unit served to develop students’ sense of epidemiologic reasoning particularly on the population-based or probabilistic nature, whereas some revision would be needed to promote their critical reasoning on the generalizability assessment task. Only 6 students (30%) supported the study-population position and demonstrated complex reasoning on the generalizability weighing between statistical and biological representativeness, whereas other students did not demonstrate such reasoning from multiple viewpoints. That said, being able to think deeply about statistical/biological representativeness is a most complex aspect of epidemiologic reasoning, and it is quite challenging to develop such reasoning with students with a few days of instruction. I conjecture that this would be due to the limited time this class could spend on the unit, whereas this short instruction time was a necessary trade off in the unit development in

relation to regular class schedules. These suggest that it requires more explicit analytical activities or scaffolding to develop critical reasoning on the generalizability assessment. In other words, ensuring sufficient time for the unit may be a key factor to promote meaningful engagement in disciplinary practices (Clark & Linn, 2003).

Conclusion

The present study discussed the educational value of disciplinary engagement in authentic practices as a way to develop students' complex causal reasoning and reported the results of classroom enactment of a curriculum aimed at developing students' epidemiologic reasoning. Literature suggests that intentional conceptual and process-focused scaffolding are key to make novice students' disciplinary engagement possible, and I additionally argued the need of discipline-focused scaffolds that attend to specific epistemological features of inquiry to develop students' disciplinary understanding. I then discussed the Exploring Databases unit designed to engage students in an authentic practice of epidemiological analysis and I argued that our designed scaffolds served students' disciplinary engagement and development in epidemiologic reasoning in some important ways, although we also found that more explicit scaffolding is needed to develop students' critical thinking in sampling and the generalizability of their explanations. Most importantly, this study demonstrated that disciplinary engagement in complex reasoning involving both statistical and biological reasoning is feasible under deliberate scaffolding, and it will also contribute to theoretical development of instructional design in probabilistic causal reasoning (Grotzer, 2012).

For design implications, the present study demonstrated the empirical value of intentional design on both process- and discipline-focused scaffolds. Specifically, instructional materials in the ExDa unit scaffolded not only “how” to conduct the focal investigation (e.g., paper

worksheets, the online database in the WISE environment), but also “why” the epistemic aim (identifying risk factors of smoking) matters by relating it to the social aim (preventing people from smoking for public health) and social constraints (e.g., ethical concerns) to help students develop their disciplinary understanding in the nature of population-based explanation and the value of observational design (case control design) that are peculiar to the epidemiologic discipline and distinctive from standard science curricula (Windschitl et al., 2007). Such epistemic or discipline-specific aspects of scaffolding are arguably a crucial research agenda to help students recognize the epistemic values of scientific analytical practices and find the relevance to their communities and personal interests (see Chapter 2). On the other hand, my arguments made here are based on a case study in an AP biology class, and therefore further research is needed to test the unit in different student populations (e.g., general biology classes) and develop more robust design principles associated with discipline-specific inquiry.

References

- AAAS (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.
- Allan, L. G. (1993). Human contingency judgments: rule based or associative?. *Psychological Bulletin*, 114, 3, 435-48.
- Bell, P., & Linn, M. C. (2000). Scientific Arguments as Learning Artifacts: Designing for Learning from the Web with KIE. *International Journal of Science Education*, 22, 8, 797-817.
- Bell, P. & Linn, M. C. (2002). Beliefs About Science: How Does Science Instruction Contribute?. In Hofer, B. K., & Pintrich, P. R. (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, N.J: L. Erlbaum.

Chapter 3

- Bell, P. (2004). Promoting students' argument construction and collaborative debate in the science classroom. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 115-143). Mahwah, NJ: Erlbaum.
- Bradford-Hill, A. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58: 295–300.
- Carey, S. (2011). Précis of 'The Origin of Concepts'. *The Behavioral and Brain Sciences*, 34, 3, 113-24.
- Cartwright, N. (2004). Causation: One Word, Many Things. *Philosophy of Science*, 71, 5, 805-819.
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction, *Review of Educational Research*, Vol. 63, No. 1, 1-49.
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically Authentic Inquiry in Schools: A Theoretical Framework for Evaluating Inquiry Tasks. *Science Education*, 86, 2, 175-218.
- Clark, D., & Linn, M. C. (2003). Designing for Knowledge Integration: The Impact of Instructional Time. *Journal of the Learning Sciences*, 12, 4, 451-493.
- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50, 1.
- de Jong, T., & van Joolingen, W. R. (1998). Scientific Discovery Learning with Computer Simulations of Conceptual Domains. *Review of Educational Research*, 68, 2, 179-201.
- Elby, A., & Hammer, D. (2001). On the Substance of a Sophisticated Epistemology. *Science Education*, 85, 5, 554-67.

Chapter 3

Gopnik, Alison, Glymour, Clark, Sobel, David, Schulz, Laura, Kushnir, Tamar, & Danks, David.

(2004). *A theory of causal learning in children: Causal maps and Bayes nets*,
Psychological Review, Vol. 111, No. 1, 3-32.

Green, J. J., Stone, C. C., Zegeye, A., & Charles, T. A. (2009). How Much Math Do Students
Need to Succeed in Business and Economics Statistics? An Ordered Probit Analysis.
Journal of Statistics Education, 17(3), 1-22.

Grotzer, T. A. (2003). Learning to Understand the Forms of Causality Implicit in Scientifically
Accepted Explanations. *Studies in Science Education*, 39, 1, 1-74.

Grotzer, T. A., & Basca, B. B. (2003). How Does Grasping the Underlying Causal Structures of
Ecosystems Impact Students' Understanding?. *Journal of Biological Education*, 38, 1, 16-
29.

Grotzer, T. A. (2012). *Learning causality in a complex world: Understandings of
consequence*. Lanham, MD: Rowman Littlefield.

Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge, U.K:
Cambridge University Press.

Hirsch, L. S. & O'Donnell, A. M. (2001). Representativeness in Statistical Reasoning:
Identifying and Assessing Misconceptions, *Journal of Statistics Education*, Vol. 9, No. 2.

Hjalmarson, M. A., Moore, T. J., & delMas, R. (2011). Statistical Analysis when the Data Is an
Image: Eliciting Student Thinking about Sampling and Variability. *Statistics Education
Research Journal*, 10, 1, 15-34.

Hodson, D. (2009). *Teaching and learning about science: Language, theories, methods, history,
traditions and values*. Rotterdam: Sense Publishers.

Chapter 3

- Hug, B., & McNeill, K. L. (2008). Use of First-Hand and Second-Hand Data in Science: Does Data Type Influence Classroom Conversations?. *International Journal of Science Education*, 30, 13, 1725-1751.
- Johnson, M., & Kuennen, E. (2006). Basic math skills and performance in an introductory statistics course. *Journal of Statistics Education*, 14(2), 1-14.
- Kahneman, D. & Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness, *Cognitive Psychology*, 3, 430-454.
- Koepsell, T. D., Weiss, N. S. (2003). *Epidemiologic Methods: Studying the Occurrence of Illness*, Oxford University Press.
- Lawson, A. E., Clark, B., Cramer-Meldrum, E., Falconer, K. A., Sequist, J. M., & Kwon, Y.-J. (2000). Development of Scientific Reasoning in College Biology: Do Two Levels of General Hypothesis-Testing Skills Exist? *Journal of Research in Science Teaching*, 37, 1, 81-101.
- Lehrer, R., & Schauble, L. (2002). *Investigating real data in the classroom: Expanding children's understanding of math and science*. New York: Teachers College Press.
- Lunsford, M. L., & Poplin, P. (2011). From Research to Practice: Basic Mathematics Skills and Success in Introductory Statistics. *Journal of Statistics Education*, 19(1), 1-22.
- Makar, K., Bakker, A. & Ben-Zvi, D. (2011). The Reasoning Behind Informal Statistical Inference, *Mathematical Thinking and Learning*, 13:1-2, 152-173.
- McNeill, K. L. (2011). Elementary Students' Views of Explanation, Argumentation, and Evidence, and Their Abilities to Construct Arguments over the School Year. *Journal of Research in Science Teaching*, 48, 7, 793-823.

Chapter 3

- Mix, K. S., Levin, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology*, 35(5), 164 - 174.
- Moore, D. S., & McCabe, G. P. (1993). *Introduction to the practice of statistics*. New York: Freeman.
- Munn, M. M., Oura, H., Gallivan, M., Van, H. K., & Shouse, A. W. (2013). Engaging high school students in research on smoking behavior. *Science*, 341, 6144, 360-1.
- National Mathematics Advisory Panel. (2008). *The Final Report of the National Mathematics Advisory Panel*. U.S. Department of Education.
- National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C.: The National Academies Press.
- Nicolaidou, I., Kyza, E. A., Terzian, F., Hadjichambis, A., & Kafouris, D. (2011). A Framework for Scaffolding Students' Assessment of the Credibility of Evidence. *Journal of Research in Science Teaching*, 48, 7, 711-744.
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, Mass: Harvard University Press.
- Konold, C., & Miller, C. (2005). Tinkerplots. *Data analysis software for middle school curricula*. Emeryville, CA: Key Curriculum.
- Konold, C., & Pollatsek, A. (2002). Data Analysis as the Search for Signals in Noisy Processes. *Journal for Research in Mathematics Education*, 33(4), 259- 289.

- Papariotodemou, E., & Meletiou-Mavrotheris, M. (2008). Developing Young Students Informal Inference Skills in Data Analysis. *Statistics Education Research Journal*, 7(2), 83-106.
- Pea, R. D. (2004). The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. *Journal of the Learning Sciences*, 13, 3.
- Peker, D., & Wallace, C. S. (2011). Characterizing High School Students' Written Explanations in Biology Laboratories. *Research in Science Education*, 41, 2, 169-191.
- Perkins, D. N., & Grotzer, T. A. (2005). Dimensions of Causal Understanding: The Role of Complex Causal Models in Students' Understanding of Science. *Studies in Science Education*, 41, 1, 117-165.
- Pineño, O., & Miller, R. R. (2007). Comparing associative, statistical, and inferential reasoning accounts of human contingency learning. *Quarterly Journal of Experimental Psychology* (2006), 60, 3, 310-29.
- Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 3-4.
- Resnick, M. (1996). Beyond the Centralized Mindset. *Journal of the Learning Sciences*, 5, 1, 1-22.
- Rothman, K. J. (2002). *Epidemiology: An introduction*. Oxford: Oxford University Press.
- Samarapungavan, A., Westby, E. L., & Bodner, G. M. (2006). Contextual Epistemic Development in Science: A Comparison of Chemistry Students and Research Chemists. *Science Education*, 90, 3.
- Sandoval, W. (2003). Conceptual and Epistemic Aspects of Students' Scientific Explanations. *Journal of the Learning Sciences*, 12, 1, 5-51.

Chapter 3

- Slotta, J. D., & Linn, M. C. (2009). *WISE science: Web-based inquiry in the classroom*. New York: Teachers College Press.
- Smith, M. H. (2004). A Sample / Population Size Activity: Is it the sample size of the sample as a fraction of the population that matters? *Journal of Statistics Education*, 12, 10.
- Sophian, C., & Wood, A. (1997). Proportional reasoning in young children: The parts and whole of it. *Journal of Educational Psychology*, 89(2), 309 - 317.
- Stampfer, M. J., Ridker, P. M., & Dzau, V. J. (2004). Risk factor criteria. *Circulation*, 109, 25, 3-5.
- Talanquer, V. (2010). Exploring Dominant Types of Explanations Built by General Chemistry Students. *International Journal of Science Education*, 32, 18, 2393-2412.
- Watson, J. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, N.J: L. Erlbaum Associates.
- Watson, J. M. (2008). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal*, 7(2), 59-82.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction*, 16, 1, 3-118.
- Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society: Series a (statistics in Society)*, 174, 2.)
- Windschitl, M., Dvornich, K., Ryken, A. E., Tudor, M., & Koehler, G. (2007). A Comparative Model of Field Investigations: Aligning School Science Inquiry with the Practices of Contemporary Science. *School Science & Mathematics*, 107, 1.

Chapter 3

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Conclusion to the Dissertation

The foregoing chapters have discussed epistemological issues of science learning and instruction and explored a theoretical framework, assessment strategy, and instructional design approach to promote students' disciplinary engagement in an authentic practice of epidemiologic investigation. In the following, I first summarize major findings and implications for curriculum and instruction, and then discuss limitations of the dissertation findings and directions for future research.

Translating Discipline-Specific Knowledge into Scaffolding Design with the Epistemic Design Framework

The first chapter discussed existing literature in epistemological learning and instruction and pointed out that a majority of research has mainly focused on discipline-general aspects of scientific epistemologies in last decades. In this research, some researchers have recently emphasized the role of discipline-specific epistemologies, and I argued that researchers should pay more attention to discipline-specific aspects of epistemological knowledge in science instruction. A rationale for this argument was the risk of epistemological conflation and conflicts (ECCs); although the latest science education framework encourages science learning through engaging in authentic science practices, epistemologically naïve students (and teachers) more likely conflate one discipline with another. Moreover, if the disciplines differed (even in part) with each other, students (and teachers) would be unable to resolve the epistemic inconsistencies or conflicts by themselves as demonstrated by a few contexts of the ExDa project in the chapter. To avoid ECCs and to guide students' disciplinary engagement, I presented a theoretical model to reify key epistemic actors that characterize discipline-specific knowledge and proposed an epistemic design framework. This framework aims to guide researchers to identify epistemic

Conclusion to the Dissertation

alignments for their focal science practice with heterogeneous actors including subject matter, epistemic aims and values, observation and measurement, and social aims and values.

Subsequently, I applied the framework into the ExDa unit to evaluate the current instructional design and discussed the utility by demonstrating that the framework helped identifying revision points in the current instruction.

In terms of my argument to emphasize discipline-specific knowledge, some may counter that those students and teachers with ECCs just had not understood the ideas of experimental design correctly or that epistemological conflation is not problematic within K-12 science education. However, such counterarguments cannot explain or should not overlook why those students and teachers “rejected” the epidemiologic discipline or causal reasoning with observational design (i.e., correlational evidence). The origin of this problem, I conjecture, is not just that they did not know experimental design correctly or alternative methods, but rather due to their lack of metacognition to reason the rationale for why one particular method (i.e., case control design) is/was adopted over experimental design in the current inquiry practice along with its epistemic aims (e.g., identifying risk factors of smoking) under various social constraints (e.g., ethical concerns). With this rationale, I argue that science instruction should not only engage students in various practices with epistemic variations but also develop their metacognitive skills or epistemic cognition (Chinn et al., 2011) to help students become able to reason a series of epistemic (methodological) choices in the currently engaged practice. Discipline-specific epistemologies are the reification of those characteristics and formations, and the epistemic design framework was proposed to guide researchers to reify epistemic characteristics for their focal practice with heterogeneous actors including material, cognitive, social, and cultural ones and translate the epistemic alignments and narratives into curriculum

and instruction. Although the proposed framework is just an initial step and more epistemic actors and forms of alignments should be identified in other science disciplines, I still argue that this work would contribute to the design theory for “epistemically-guided” instruction to make student engagement in practice-focused learning more meaningful along with authentic scientific values.

Assessing Students’ Disciplinary Understanding and the Design of Instruction

Extending the theoretical discussion in the first chapter, Chapter 2 proposed an assessment strategy to reveal students’ disciplinary understanding and discussed the utility in a curriculum enactment design research study of the ExDa implementation. Chapter 1 discussed a way to disentangle discipline-specific epistemologies for science practices, and another important question was how to reveal and assess student’s disciplinary understanding and its development. To approach this, I proposed a value-focused approach to elicit student understanding of the “values” of epistemic actors and alignments as an indicator of students’ disciplinary understanding. Specifically, this value-focused approach aims to elicit student understanding in a means-and-end relationship of a target epistemic/social actor(s) by asking students “why” the actor(s) matters to the currently engaged practice with relevant actors in the epistemic alignments. After articulating components and procedures of the value-focused approach, I applied the strategy to the ExDa context in a case study and the findings suggested that students significantly developed their disciplinary knowledge of the epidemiological practice over the course of the curriculum unit. However, at the same time, it also identified that it seems more challenging to develop student understanding in the values of analytical actors than social actors, which helped identifying revision points in the current instruction.

A central argument for the chapter was to point out the need of specific assessment strategies to reveal students' disciplinary knowledge, because such assessment tools guide researchers to reify learning goals, design scaffolds, and evaluate and revise the design based on concrete evidence of student engagement and performance. The value-focused approach defines key procedures to develop assessment items with a set of open-ended questions so that it can be applied to various science practices in different disciplines. Although the curriculum enactment study also identified some methodological issues in a caution for the “clarity” of problem statements for open-ended questions, I still argue that the value-focused approach would contribute to advancing the theorization of epistemic assessment and identifying strong and weak points of instructional design.

Using Epistemic Sense-Making to Develop Students' Disciplinary Knowledge

Another major finding from Chapter 2 was that the curriculum enactment study illuminated the need of specific scaffolding to promote *epistemic conversation*. This refers to a sort of metacognitive conversation in which students talk about “why” they need to use particular epistemic (analytical) tools—case control design, odds ratio, and 95% confidence interval in the ExDa unit—in the currently engaged practice. The curriculum enactment demonstrated that lecture-based instruction seemed insufficient to develop students' disciplinary knowledge, and I conjecture that it would be partly because most epistemic characteristics (e.g., epistemic aims, methods, and analytical tools) are often in the “given” condition in inquiry-based curricula and students do not have to engage in the decision-makings that would actually promote students' epistemic sense making and knowledge. Since it is unrealistic to leave all the choices to students in science curricula, I argue that some epistemic task or activity would be crucial to provoke

epistemic conversations such as engaging students in choosing and justifying methodological strategies and analytical tools under some hypothetical context.

High-School Students Can Engage in Statistical Data Analysis with Process- and Epistemic-focused Scaffolds

Chapter 3 shifted the theoretical focus to causal reasoning and investigated how the ExDa unit engaged students in an epidemiologic investigation. The focal investigation pursues the explanatory form of risk factor and it requires statistical reasoning to develop population-based explanations rather than focusing on a few single cases. However, students are expected to be unfamiliar with statistical ideas in K-12 curricula, and literature suggests that learning statistics is challenging even for college and adult students. To engage this challenge, I drew on informal statistical inference theory and applied it to the design of scaffolding in the ExDa unit to support and develop students' epidemiologic reasoning with both epistemic- and process-focused scaffolds. Epistemic-focused scaffolds included instructional materials to develop student understanding in the epistemic aim of the epidemiologic investigation, and statistical inference with the concepts of population and sample. Process-focused scaffolds included worksheets to guide the flow of the investigation and technological artifacts to allow students to estimate statistics without deep technical knowledge. The case study demonstrated that the ExDa unit guided students' disciplinary engagement in the epidemiologic discipline and the findings suggested that students significantly developed their conceptual knowledge and population-based reasoning. On the other hand, the analysis also indicated that more deliberate scaffolding is needed to develop student understanding and reasoning in generalizability associated with statistical and biological representativeness. I conjecture that this was partly because understanding statistical inference is challenging to high school students but the unit could spend

Conclusion to the Dissertation

only one 50-minute class for the topic given time constraints. Accordingly, I argued that it is still feasible to engage students in statistical data analysis with deliberate scaffolding by developing both process-focused (i.e., how to analyze data) and epistemic-focused scaffolds (e.g., epistemic aims and values, analytical values, social aims and values) with sufficient time for students to engage in the analytical practices.

Limitations and Directions for Future Research

In the preceding discussion, I argued that the present thesis would contribute to the theoretical development in scaffolding design for students' epistemological development, whereas all the dissertation findings are based on the enactment of a curriculum unit and two student case studies. Therefore, the applicability of proposed frameworks or theoretical tools and the generalizability of the findings should have certain limitations. Also, other lines of research are needed to promote practice-focused science curricula in different approaches. In the following sections, I discuss two limitations of the dissertation findings and three potential directions for future research.

Need of Applying Proposed Theoretical Frameworks to Other Science Curricula

First, the epistemic design framework (Chapter 1) and value-focused approach (Chapter 2) were developed drawing from a broad literature, whereas the application of these ideas was conducted only in the ExDa curriculum unit. Therefore, more research is needed to evaluate the applicability and advance the theoretical details by applying them into different science practices and disciplines. Specifically, although Chapter 1 derived the epistemic design framework based on a broad literature including science studies, I expect that other key epistemic actors and alignments may exist in other science disciplines and sub-fields and the applicability should also be tested in those practices to advance the theoretical details in future research. Related to this, I

Conclusion to the Dissertation

also propose the need of research on epistemological conflation and conflicts in school contexts. As Stevens et al. (2005) pointed out, each science subject (e.g., physics, biology, chemistry) has developed epistemic practices aligned to inquiry in that subject, and therefore it would be informative to investigate whether ECCs occur in other science subjects in relation to the current science education reforms that promote practice-focused science learning.

Using the Epidemiologic Reasoning Approach across Diverse Student Populations

Second, all the case studies in the present dissertation were conducted in an AP Biology class at a suburban high school and student reactions to the Exda unit may differ in general biology classes and different schools, which might influence on the findings in Chapter 3. The particular student background was selected because of the cooperating teacher's decision based on her experiential conjecture that the unit could be too advanced for her general biology students. This assumption does not necessarily say that general biology students cannot learn important ideas from the unit as the matter of fact that the unit in the broader project was implemented at more than ten different high schools including general biology classes. Yet, I expect that it would require more deliberate scaffolds and dedicative support to engage students in epidemiologic reasoning in general biology classes, because it involves complex reasoning with statistical concepts that are not covered in standard curricula. Therefore, further research is needed to test applicability of the unit and improve its scaffolding design to engage more diverse students in epidemiologic reasoning.

Developing Teacher Education and Professional Development to Promote Epistemological Instruction

As potential new directions for future research, first, I argue there is a need for research on teacher education and professional development focused on epistemological instruction.

Conclusion to the Dissertation

Chapter 1 demonstrated that a few teachers rejected the epidemiologic discipline or causal reasoning with correlational evidence. This would be partly because teachers generally do not have ample research experience (White & Frederiksen, 1998), and they need some specific training or experience to develop their epistemological understanding in science(s). However, literature on discipline-general knowledge (DGK) approaches suggests that teacher education and professional development on epistemological learning are also a formidable goal (Hodson, 2009), and thus specific research is needed to develop teachers' disciplinary understanding in the goal of promoting epistemological instruction along with the current science education reforms (NRC, 2012). Although the present thesis did not focus on teacher education and professional development, I conjecture that the epistemic design framework and value-focused assessment might also contribute to this line of research.

Investigating Scientists' Understanding in Epistemic Alignment with Their Practices

Second, the value-focused research on scientists' disciplinary knowledge and its development might contribute to advancing instructional and scaffolding theory for students/teachers' disciplinary engagement and development. Chapter 1 and Chapter 2 introduced a few studies that investigated professional scientists' viewpoints on scientific epistemologies (e.g., Osborne, Collins, Ratcliffe, Millar, & Duschl, 2003; Samarapungavan, Westby, & Bodner, 2006), and their findings revealed some common aspects (Osborne et al., 2003) of epistemological knowledge on science in general as well as some divergent aspects (Samarapungavan et al., 2006) associated with their experience in disciplinary engagement. However, the main focuses of these studies are on scientists' "established" perspectives on scientific epistemologies, and it may be informative for educational research to investigate "what constitutes or influence on professional scientists' epistemological knowledge and its

development” (e.g., Nespor, 1994). I conjecture the value-focused approach proposed in Chapter 2 might be useful to reveal scientists’ understanding in epistemic alignments of their own practices and the findings could contribute to advancing scaffolding theory for students/teachers’ epistemological learning.

Technological Infrastructures to Provide Students with More Opportunities for Scientific Investigations with Statistical Data

The last proposal is the need of technological tools or infrastructure to provide students with more opportunities for authentic inquiry with statistical data. Informal statistical inference (ISI) theory and Chapter 3 findings suggest that it is feasible to engage students in authentic data analysis with statistical data under deliberate scaffolding, whereas one practical issue to promote such effort is that it requires to develop technological/visualization tools for each curriculum unit. Although ISI studies utilize existing software such as Excel (Duffy, 2010) and TinkerPlots (Konold & Miller, 2005), it has some limitations in the customizability and types of data analysis that the software can handle.

To respond to this challenge, I argue that it is possible to optimize time and cost to build technological tools by utilizing open source projects. For example, a server-side module *rApache* (Horner, 2013) and a web application framework *Shiny* (<http://shiny.rstudio.com/>) allow for running the R language—a programming language for statistical analysis—on a web server and communicating with web browsers on client computers. Today, a majority of statisticians utilize R to analyze their data and many of statistical analysis procedures as well as visualization tools can be processed with R libraries through online repositories such as Comprehensive R Archive Network (CRAN). I argue that by utilizing these open source projects it is feasible to develop and integrate various statistical representations into science curricula so that students are able to

Conclusion to the Dissertation

analyze authentic data in the science classroom. This type of mash-up projects would also contribute to developing learning environments and infrastructures to promote authentic data analysis experience in science instruction.

Final Remarks

Throughout the dissertation work, I have tackled theoretical and empirical challenges of epistemological learning and instruction in K-12 science education. Even though it is just a small step toward the solution, I believe that this work will contribute to theoretical advancement of scaffolding design theory in learning sciences and the refinement of classroom instruction with respect to epistemic aspects of student learning. In his article entitled “Epistemology for the Masses: The Origins of “The Scientific Method” in American Schools” (Rudolph, 2005), John L. Rudolph vividly demonstrated the historical moves of science education in the United States in which Dewey’s theory and vision of education were mistranslated into a way that he would never wish in the vortex of popularization of science, its integration into education, rapid expansion of student enrollment, and demands of teaching the “Masses.” Alongside the recent review on comparative understanding of school subjects by Stevens, Wineburg, Herrenkohl, and Bell (2005), his article gave me an important lesson in the risk of “oversimplification” and “overgeneralization”, and their insights were reflected into a theoretical focus on the “design process” of epistemic framing (Chapter 1), assessment (Chapter 2), and instructional design (Chapter 3) rather than pursuing a static generalized set of theoretical components and frameworks. I argue that educational researchers should take the historical lesson seriously and carefully engage in the translational work of theory into practice with respect on the diverse and dynamic nature of contemporary science. Although the present thesis is just a small step in this direction, I strongly hope that this work will stimulate further discussions and theoretical

Conclusion to the Dissertation

development toward epistemically-rich instruction and contribute to helping children become critical consumers of science in their personal interests with an authentic understanding in the value of science, and hopefully encouraging some of them to pursue a STEM career in their own choice.

References

- Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the Dimensions of Epistemic Cognition: Arguments from Philosophy and Psychology. *Educational Psychologist*, 46, 3, 141-167.
- Duffy, S. (2010). Random Numbers Demonstrate the Frequency of Type I Errors: Three Spreadsheets for Class Instruction, *Journal of Statistics Education*, 18, 2.
- Hodson, D. (2009). *Teaching and learning about science: Language, theories, methods, history, traditions and values*. Rotterdam: Sense Publishers.
- Horner, J. (2013). rApache: Web application development with R and Apache.
<http://www.rapache.net/>
- Konold, C., & Miller, C. (2005). Tinkerplots. *Data analysis software for middle school curricula*. Emeryville, CA: Key Curriculum.
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C.: The National Academies Press.
- Nespor, J. (1994). *Knowledge in motion: Space, time, and curriculum in undergraduate physics and management*. London: Falmer Press.

- Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "Ideas-about-Science" Should Be Taught in School Science? A Delphi Study of the Expert Community. *Journal of Research in Science Teaching*, 40, 7, 692-720.
- Rudolph, J. L. (2005). Epistemology for the Masses: The Origins of "The Scientific Method" in American Schools. *History of Education Quarterly*, 45, 3.
- Samarapungavan, A., Westby, E. L., & Bodner, G. M. (2006). Contextual Epistemic Development in Science: A Comparison of Chemistry Students and Research Chemists. *Science Education*, 90, 3.
- Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005). Comparative Understanding of School Subjects: Past, Present, and Future. *Review of Educational Research*, 75, 2, 125-157.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction*, 16, 1, 3-118.

Appendix A. Pre Survey Items

Question 1.

Your Name

Read the scenarios below and answer the following questions.

Question 2.

Sarah came up with a hypothesis that feeling a buzz when one tried smoking can be a risk factor for regular smoking, and tested her hypothesis with a database. She identified a question “During your experimental smoking phase, did you experience any of the following regularly? A buzz”, and defined “Yes” as exposed and “No” as not–exposed. The odds ratio was 6.33 and the confidence interval was [3.33, 12.02]. She applied the criteria for causality and concluded that feeling a buzz can be a risk factor for regular smoking. However, her teammate Tom disagreed and claimed that everyone does NOT necessarily smoke even if he/she felt a buzz. Whose position do you support? Explain why and refer to each position in your explanation.

Question 3.

Andrew came up with his hypothesis that having close friends who smoke can be a risk factor for regular smoking, and tested his hypothesis with the same database. He identified a question “During your experimental smoking phase, did any of your closest friends smoke cigarettes?”, and defined “Most or all” as exposed, and “Some” and “None” as not–exposed. The odds ratio was 3.61 and the confidence interval was [2.07, 6.3]. He applied the criteria for causality and concluded that having close friends who smoke can be a risk factor for regular smoking. However, his teammate Kari disagreed and claimed that the cause is one’s will, since people can make their own choice for smoking. Whose position do you support? Explain why and refer to each position in your explanation.

Question 4.

Three students (Nancy, Fraser, and Jose) tested their hypothesis that having parents who smoke can be a risk factor for regular smoking. They identified a question “While you were growing up, did any of your parents/guardians smoke inside your home?” and defined “Yes” as exposed, and “No” as not–exposed. The odds ratio was 5.2 and the confidence interval was [3.12, 8.67]. They applied the criteria for causality and concluded that their hypothesis is valid, but they wondered to whom they could apply their results. Knowing that the people who participated in the study were all from the Puget Sound area, Nancy claimed that their results can apply to the people who live in this area, while Fraser claimed that their results should only apply to the people who

Appendices

participated in the survey. Jose claimed that their results can apply to other populations as well. Whose position do you support? Explain why and refer to each position in your explanation.

Question 5.

Knowing that the evidence from observational case control studies is associative rather than causal, Aaron wondered why epidemiologists often use observational studies rather than randomly manipulated experiments that can produce causal evidence. How would you help him understand the value of observational studies? In your answer, include the advantages and limitations of observational studies and under what conditions it is better to use them.

Question 6.

Tyler has never used odds ratio calculation to analyze data from case control studies. He wonders why scientists use odds ratios to analyze data. How would you help him understand the value of using odds ratio to analyze data? Explain its purpose and what the number indicates.

Question 7.

Sandra has never used 95% confidence intervals (CIs) to analyze data from case control studies. She wonders why scientists need it in addition to odds ratio to analyze data. How would you help her understand the value of using 95% CIs? Explain its purpose and what the interval indicates.

Question 8.

Steve has never used the criteria for causality to analyze data from observational case control studies. He wonders why scientists need it in addition to odds ratios and 95% confidence intervals. How would you help him understand the value of using the criteria for causality? Explain its purpose and the limitations of odds ratio and 95% CI.

Question 9.

Chloe wonders why tobacco/smoking research matters to society. How would you help her understand the value of this research? Explain problems caused by tobacco/smoking and the contributions of tobacco/smoking research.

Read the story below and answer the following questions.

Mr. Limon notices that each day a number of students fall asleep in his World History class. He thinks that students are falling asleep because they don't get enough sleep at night. He decides to do a case control study to test his idea. He will ask students who fall asleep in class and those who don't whether or not they slept for 7 hours or more the night before and then compare these two figures.

The next day Mr. Limon passes out anonymous questionnaires to his four World History classes and collects the following data. Not everyone, but 120 students responded to the survey,

Appendices

and 28 reported falling asleep in class yesterday. Of these students, 20 had slept less than 7 hours the night before. Among students who had stayed awake, 39 had slept less than 7 hours.

Data Summary

	Fell asleep	Stayed awake
Slept LESS than 7 hours	20	39
Slept MORE than 7 hours	8	53

Question 10.

What is the study population?

- Students who attended Mr. Limon's History classes
- Students who fell asleep in class
- Students who stayed awake in class
- Students who slept more than 7 hours
- Students who slept less than 7 hours

Question 11.

What is the outcome?

- Attending Mr. Limon's History classes
- Falling asleep in class
- Staying awake in class
- Sleeping more than 7 hours
- Sleeping less than 7 hours

Question 12.

Who are the cases?

- Students who attended Mr. Limon's History classes
- Students who fell asleep in class
- Students who stayed awake in class
- Students who slept more than 7 hours
- Students who slept less than 7 hours

Question 13.

Who are the controls?

- Students who attended Mr. Limon's History classes
- Students who fell asleep in class
- Students who stayed awake in class
- Students who slept more than 7 hours
- Students who slept less than 7 hours

Question 14.

Appendices

What is the exposure?

- Attending Mr. Limon's History classes
- Falling asleep in class
- Staying awake in class
- Sleeping more than 7 hours
- Sleeping less than 7 hours

Question 15.

Is this study prospective or retrospective?

- Prospective
- Retrospective
- Question 16.
- What is the sample size?
- 20
- 28
- 92
- 120
- Question 17.
- Calculate the odds ratio (OR) between cases and controls.

Question 16.

What is the sample size?

- 20
- 28
- 92
- 120

Question 17.

Calculate the odds ratio (OR) between cases and controls.

Question 18.

Explain what the OR tells us.

Question 19.

Is odds ratio probability or likelihood?

- Probability
- Likelihood

Question 20.

Suppose the OR = 3.2 and the 95% confidence interval = [1.6, 5.8], can we say that the OR is statistically significant? Explain why.

Question 21.

Appendices

Shelley is confused with the meaning of “statistically significant”, how would you help her understand what it tells you? Explain using three key phrases: “odds ratio (OR)”, “95% CI”, and “OR = 1”.

Question 22.

Which of the criteria for causality correspond to the following statement?

"Can the association be explained by another factor? Is there a factor that explains the association?"

- Strength of association
- Dose–response relationship
- Lack of confounder
- Biological plausibility

Question 23.

Which of the following most corresponds to the purpose of “matching” in case control studies?

- Finding as many respondents as possible
- Finding as many cases in different ethnic and racial groups as possible
- Finding as many controls in different ethnic and racial groups as possible
- Finding similar numbers of cases and controls in each ethnic and racial group as possible

Question 24.

Which type of error or bias most corresponds to the following statement?

"Although the amount of compensation given to the participants in this study was comparable to similar studies, the \$30 gift card was an incentive to participate for some people. Several of the case subjects mentioned that they had learned about the study and \$30 gift card at the homeless shelter they frequented. This group was almost entirely smokers, and they tended to be in the older range, skewing the average age for cases."

- Random error
- Selection bias
- Information bias
- Confounding

Question 25.

Which type of error or bias can be avoided or mitigated by “matching” in case control studies?

- Random error

Appendices

- Selection bias
- Information bias
- Confounding

Appendix B. Post Survey Items

Question 1.

Your Name

Read the scenarios below and answer the following questions.

Question 2.

What is the study population?

- 25 to 54 year–old adults from ethnic and racial groups in general
- 25 to 54 year–old adults from ethnic and racial groups found in the Washington State
- 25 to 54 year–old adults from ethnic and racial groups found in the Puget Sound Area
- 25 to 54 year–old adults from ethnic and racial groups in the database

Question 3.

What are the cases?

- Adults who are regular smokers
- Adults who had been regular smokers, but quit at some point before Adults who have tried smoking but did NOT continue
- Adults who have never smoked before

Question 4.

What are the controls?

- Adults who are regular smokers
- Adults who had been regular smokers, but quit at some point before Adults who have tried smoking but did NOT continue
- Adults who have never smoked before

Question 5.

Which smoking stage have all subjects (both cases and controls) experienced in the study?

- Initiation
- Maintenance
- Cessation
- Relapse

Question 6.

Which smoking stage have ONLY the cases experienced in the study?

- Initiation

Appendices

- Maintenance
- Cessation
- Relapse

Question 7.

Is the subject recruitment random or voluntary?

- Random
- Voluntary

Question 8.

Richard came up with his hypothesis that high stress can be a risk factor for regular smoking and tested his hypothesis with the same database you used for your project. He identified a question “As a teen, how much stress did you experience in your life?” and defined “A lot” as exposed, and “Some” and “None or very little” as non-exposed. The odds ratio was 1.98 and the confidence interval was [1.24, 3.17]. He applied the criteria for causality and concluded that high stress can be a risk factor for regular smoking. However, his team-mate Ellen disagreed because people make their own choice about whether they smoke or not regardless of the degree of stress. Whose position do you support? Explain why and refer to each position in your explanation.

Question 9.

Sarah came up with a hypothesis that feeling a buzz when one tried smoking can be a risk factor for regular smoking, and tested her hypothesis with the same database. She identified a question “During your experimental smoking phase, did you experience any of the following regularly? A buzz”, and defined “Yes” as exposed and “No” as not-exposed. The odds ratio was 6.33 and the confidence interval was [3.33, 12.02]. She applied the criteria for causality and concluded that feeling a buzz can be a risk factor for regular smoking. However, her teammate Tom disagreed and claimed that everyone does NOT necessarily smoke even if he/she felt a buzz. Whose position do you support? Explain why and refer to each position in your explanation.

Question 10.

Andrew came up with his hypothesis that having close friends who smoke can be a risk factor for regular smoking, and tested his hypothesis with the same database. He identified a question “During your experimental smoking phase, did any of your closest friends smoke cigarettes?”, and defined “Most or all” as exposed, and “Some” and “None” as not-exposed. The odds ratio was 3.61 and the confidence interval was [2.07, 6.3]. He applied the criteria for causality and concluded that having close friends who smoke can be a risk factor for regular smoking. However, his teammate Kari disagreed and claimed that the cause is one’s will, since people can make their own choice for smoking. Whose position do you support? Explain why and refer to each position in your explanation.

Appendices

Question 11.

Three students (Nancy, Fraser, and Jose) tested their hypothesis that having parents who smoke can be a risk factor for regular smoking. They identified a question “While you were growing up, did any of your parents/guardians smoke inside your home?” and defined “Yes” as exposed, and “No” as not-exposed. The odds ratio was 5.2 and the confidence interval was [3.12, 8.67]. They applied the criteria for causality and concluded that their hypothesis is valid, but they wondered to whom they could apply their results. Knowing that the people who participated in the study were all from the Puget Sound area, Nancy claimed that their results can apply to the people who live in this area, while Fraser claimed that their results should only apply to the people who participated in the survey. Jose claimed that their results can apply to other populations as well. Whose position do you support? Explain why and refer to each position in your explanation.

Question 12.

Knowing that the evidence from observational case control studies is associative rather than causal, Aaron wondered why epidemiologists often use observational studies rather than randomly manipulated experiments that can produce causal evidence. How would you help him understand the value of observational studies? In your answer, include the advantages and limitations of observational studies and under what conditions it is better to use them.

Question 13.

Tyler has never used odds ratio calculation to analyze data from case control studies. He wonders why scientists use odds ratios to analyze data. How would you help him understand the value of using odds ratio to analyze data? Explain its purpose and what the number indicates.

Question 14.

Sandra has never used 95% confidence intervals (CIs) to analyze data from case control studies. She wonders why scientists need it in addition to odds ratio to analyze data. How would you help her understand the value of using 95% CIs? Explain its purpose and what the interval indicates.

Question 15.

Steve has never used the criteria for causality to analyze data from observational case control studies. He wonders why scientists need it in addition to odds ratios and 95% confidence intervals. How would you help him understand the value of using the criteria for causality? Explain its purpose and the limitations of odds ratio and 95% CI.

Question 16.

Chloe wonders why tobacco/smoking research matters to society. How would you help her understand the value of this research? Explain problems caused by tobacco/smoking and the contributions of tobacco/smoking research.

Appendices

Read the story below and answer the following questions.

Mr. Limon notices that each day a number of students fall asleep in his World History class. He thinks that students are falling asleep because they don't get enough sleep at night. He decides to do a case control study to test his idea. He will ask students who fall asleep in class and those who don't whether or not they slept for 7 hours or more the night before and then compare these two figures.

The next day Mr. Limon passes out anonymous questionnaires to his four World History classes and collects the following data. Not everyone, but 120 students responded to the survey, and 28 reported falling asleep in class yesterday. Of these students, 20 had slept less than 7 hours the night before. Among students who had stayed awake, 39 had slept less than 7 hours.

Data Summary

	Fell asleep	Stayed awake
Slept LESS than 7 hours	20	39
Slept MORE than 7 hours	8	53

Question 17.

What is the study population?

- Students who attended Mr. Limon's History classes
- Students who fell asleep in class
- Students who stayed awake in class
- Students who slept more than 7 hours
- Students who slept less than 7 hours

Question 18.

What is the outcome?

- Attending Mr. Limon's History classes
- Falling asleep in class
- Staying awake in class
- Sleeping more than 7 hours
- Sleeping less than 7 hours

Question 19.

Who are the cases?

- Students who attended Mr. Limon's History classes
- Students who fell asleep in class
- Students who stayed awake in class
- Students who slept more than 7 hours
- Students who slept less than 7 hours

Question 20.

Who are the controls?

- Students who attended Mr. Limon's History classes
- Students who fell asleep in class

Appendices

- Students who stayed awake in class
- Students who slept more than 7 hours
- Students who slept less than 7 hours

Question 21.

What is the exposure?

- Attending Mr. Limon's History classes
- Falling asleep in class
- Staying awake in class
- Sleeping more than 7 hours
- Sleeping less than 7 hours

Question 22.

Is this study prospective or retrospective?

- Prospective
- Retrospective

Question 23.

What is the sample size?

- 20
- 28
- 92
- 120

Question 24.

Calculate the odds ratio (OR) between cases and controls.

Question 25.

Explain what the OR tells us.

Question 26.

Is odds ratio probability or likelihood?

- Probability
- Likelihood

Question 27.

Suppose the OR = 3.2 and the 95% confidence interval = [1.6, 5.8], can we say that the OR is statistically significant? Explain why.

Question 28.

Appendices

Shelley is confused with the meaning of “statistically significant”, how would you help her understand what it tells you? Explain using three key phrases: “odds ratio (OR)”, “95% CI”, and “OR = 1”.

Question 29.

Which of the criteria for causality correspond to the following statement?

"Can the association be explained by another factor? Is there a factor that explains the association?"

- Strength of association
- Dose–response relationship
- Lack of confounder
- Biological plausibility

Question 30.

Which of the following most corresponds to the purpose of “matching” in case control studies?

- Finding as many respondents as possible
- Finding as many cases in different ethnic and racial groups as possible
- Finding as many controls in different ethnic and racial groups as possible
- Finding similar numbers of cases and controls in each ethnic and racial group as possible

Question 31.

Which type of error or bias most corresponds to the following statement?

"Although the amount of compensation given to the participants in this study was comparable to similar studies, the \$30 gift card was an incentive to participate for some people. Several of the case subjects mentioned that they had learned about the study and \$30 gift card at the homeless shelter they frequented. This group was almost entirely smokers, and they tended to be in the older range, skewing the average age for cases."

- Random error
- Selection bias
- Information bias
- Confounding

Question 32.

Which type of error or bias can be avoided or mitigated by “matching” in case control studies?

- Random error
- Selection bias

Appendices

- Information bias
- Confounding

Appendix C. Coding Schemes

Item identifiers below are not same with those in pre and post surveys.

General Direction:

- Please read the questions and criteria for scoring first, before reading student responses
- If you think the response sufficiently meets any one of the criteria, please assign “+1”; if it partially meets the criteria, assign “0.5”; if it corresponds to the criteria but totally incorrect or just insufficient, assign “0”.
- If you find a response that doesn’t correspond to any criterion, please provide some category or description you would suggest in “other”.

Item No. = P7

Question and Criteria
Richard came up with his hypothesis that high stress can be a risk factor for regular smoking and tested his hypothesis with the same database you used for your project. He identified a question “As a teen, how much stress did you experience in your life?” and defined “A lot” as exposed, and “Some” and “None or very little” as non-exposed. The odds ratio was 1.98 and the confidence interval was [1.24, 3.17]. He applied the criteria for causality and concluded that high stress can be a risk factor for regular smoking. However, his teammate Ellen disagreed because people make their own choice about whether they smoke or not regardless of the degree of stress. Whose position do you support? Explain why and refer to each position in your explanation. a) Identifying statistical significance (+1): Articulating the condition for statistical significance (e.g., one is outside the interval) b) Strength of association (+1): Taking into account the <u>size</u> (e.g., large or small) of association c) Biological plausibility (+1): Articulating some biological explanation(s) for the risk factor d) (Lack of) confounding (+1): Articulating some confounder(s) or why it is not the case e) Statistical/probabilistic reasoning (+1): Articulating some statistical/probabilistic statement(s) (e.g., more likely, higher probability) rather than deterministic judgment (e.g., all or nothing) f) Indirect cause (+1): Taking into account some indirect cause/effect(s) g) Distinguishing association from causation, and also seeking causal explanation / (+1): Articulating that the OR indicates association, not directly causation, but also seeking some causal explanation (e.g., risk) Examples: “I agree with Richard. This is because his $OR > 1$ and his 95% CI doesn't contain 1 in it. This means his question is significant and correct. Ellen is right that there may be a confounder, but

there isn't enough evidence to prove a confounder is a problem.”

Scoring: $a(1) + d(1) = 2$ (note: however, the expression “correct” is misleading)

“In this case, I would support Ellen. Although stress makes people about 2x more likely to smoke, it doesn't mean that everyone will choose to smoke because they have stress.”

Scoring: $b(0.5) + e(0) = 1.5$ (note: somewhat deterministic due to “not everyone”)

“Having also tested stress as a factor, I'm more inclined to agree with Richard based off of my own results. Without getting too lengthy, smoking often causes a feeling a relaxation, which is experienced around six times as often in the cases, which definitely supports Richard's position (although his odds ratio and confidence interval are only "okay"). As for Ellie's closeminded viewpoint, she really should take into consideration that stress could be a subconscious factor, meaning they aren't necessarily aware of it 24/7, but may still act accordingly.” Scoring: $c(1) + b(0.5) + f(1) = 2.5$

Item No. = Q1

Sarah came up with a hypothesis that feeling a buzz when one tried smoking can be a risk factor for regular smoking, and tested her hypothesis with the same database. She identified a question “During your experimental smoking phase, did you experience any of the following regularly? A buzz”, and defined “Yes” as exposed and “No” as not-exposed. The odds ratio was 6.33 and the confidence interval was [3.33, 12.02]. She applied the criteria for causality and concluded that feeling a buzz can be a risk factor for regular smoking. However, her teammate Tom disagreed and claimed that everyone does NOT necessarily smoke even if he/she felt a buzz. Whose position do you support? Explain why and refer to each position in your explanation.

- a) Identifying statistical significance (+1): Articulating the condition for statistical significance (e.g., one is outside the interval)
- b) Strength of association (+1): Taking into account the size (e.g., large or small) of association
- c) Biological plausibility (+1): Articulating some biological explanation(s) for the risk factor
- d) (Lack of) confounding (+1): Articulating some confounder(s) or why it is not the case
- e) Statistical/probabilistic reasoning (+1): Articulating some statistical/probabilistic statement(s) (e.g., more likely, higher probability) rather than deterministic judgment (e.g., all or nothing)
- f) Indirect cause (+1): Taking into account some indirect cause/effect(s)
- g) Distinguishing association from causation, and also seeking causal explanation / (+1): Articulating that the OR indicates association, not directly causation, but also seeking some causal explanation (e.g., risk)

Examples:

“I support Sarah's position because of how high her OR and CI was and how far from 1 they are. There is a significant correlation between feeling a buzz and maintenance smoking. Feeling a buzz is a risk factor for continuing smoking. True, not everyone who smokes feels a buzz but it's simple a higher possibility to continue if they do.”

Scoring: $a(1) + e(1) + g(0.5) = 2.5$ (note: not so sure if he/she seeks causal explanation)

“i support Sarah because her CI did not include one and showed that there was a 6.33 times more chance of the people becoming a maintenance smoker. While Tom may be right with his idea that everyone doesn't feel a buzz, but her studies show that there is a more likely chance of continuing to smoke when feeling some sort of buzz or comfort feeling. Because her numbers were high, this shows there is a correlation between feeling a buzz and maintenance smoking.”

Scoring: $a(1) + b(1) + e(1) + g(0.5) = 3.5$ (note: not sure if he/she seeks causal explanation)

Item No. = Q2

Andrew came up with his hypothesis that having close friends who smoke can be a risk factor for regular smoking, and tested his hypothesis with the same database. He identified a question “During your experimental smoking phase, did any of your closest friends smoke cigarettes?”, and defined “Most or all” as exposed, and “Some” and “None” as not-exposed. The odds ratio was 3.61 and the confidence interval was [2.07, 6.3]. He applied the criteria for causality and concluded that having close friends who smoke can be a risk factor for regular smoking. However, his teammate Kari disagreed and claimed that the cause is one's will, since people can make their own choice for smoking. Whose position do you support? Explain why and refer to each position in your explanation.

- a) Identifying statistical significance (+1): Articulating the condition for statistical significance (e.g., one is outside the interval)
- b) Strength of association (+1): Taking into account the size (e.g., large or small) of association
- c) Biological plausibility (+1): Articulating some biological explanation(s) for the risk factor
- d) (Lack of) confounding (+1): Articulating some confounder(s) or why it is not the case
- e) Statistical/probabilistic reasoning (+1): Articulating some statistical/probabilistic statement(s) (e.g., more likely, higher probability) rather than deterministic judgment (e.g., all or nothing)
- f) Indirect cause (+1): Taking into account some indirect cause/effect(s)
- g) Distinguishing association from causation, and also seeking causal explanation / (+1): Articulating that the OR indicates association, not directly causation, but also seeking some causal explanation (e.g., risk)

Examples:

“I would agree with Andrew because based on their database their odd ratio was higher than a 1 but wouldn't say that it's actually a high risk factor but is definitely a risk factor meaning that it's significant.”

Scoring: $b(1) + g(0.5) = 1.5$ (note: though he/she saying “risk”, but not sure if he/she understands the concept)

“I agree with Andrew. Kari argues that people can choose whether or not to smoke. This is true but risk factors, like friends who smoke, can increase the likelihood that you are going to smoke. That is what Andrew is saying; he is not saying that it will cause them to smoke, he's saying that it increases their chances of becoming a smoker.”

Scoring: $e(1) + g(0.5) = 1.5$ (note: he/she clearly says, the finding indicates only association)

Item No. = Q3

Three students (Nancy, Fraser, and Jose) tested their hypothesis that having parents who smoke can be a risk factor for regular smoking. They identified a question “While you were growing up, did any of your parents/guardians smoke inside your home?” and defined “Yes” as exposed, and “No” as not-exposed. The odds ratio was 5.2 and the confidence interval was [3.12, 8.67]. They applied the criteria for causality and concluded that their hypothesis is valid, but they wondered to whom they could apply their results. Knowing that the people who participated in the study were all from the Puget Sound area, Nancy claimed that their results can apply to the people who live in this area, while Fraser claimed that their results should only apply to the people who participated in the survey. Jose claimed that their results can apply to other populations as well. Whose position do you support? Explain why and refer to each position in your explanation.

- a) Study population (+1): Articulating the study population (= the Puget Sound area)
- b) Sampling (+1): Articulating the representativeness based on random or voluntary sampling
- c) Statistical representativeness (+1): Articulating the representativeness based on the study population
- d) Biological representativeness (+1): Articulating the representativeness based on the similarity or dissimilarity with other populations
- e) Statistical significance (+1): Articulating the condition for statistical significance (e.g., one is outside the interval)

Examples:

“I support Nancy. Fraser says that it only applies to people who took the survey. This is true but how would you find OR and CI for the entire population of a town/city/county. You would have to individually go to every person and survey them. Jose says that it can apply to other populations but that's only partly true. Some populations there are less smokers than non-

smokers which would completely change the data. Nancy says it can apply to the people who live in that area. This would be the best option. There would be no other way to measure the whole population so you should just calculate the information for a chunk of it and apply it to the whole population.”

Scoring: $a(0.5) + c(1) + d(1) = 2.5$

“I support jose. I also believe that it can be applied to other populations because it is known that the Pudget Sound area is very diverse. “ Scoring: $d(0.5) = 0.5$ (note: there is no explicit explanation why the “diverse” nature allows for applying to other populations)

“Because they used the 95% CI to make sure there hypothesis is correct and significant, they can apply the 95% CI to the population. In other populations, the OR is 95% likely to fit into the range of 3.12 and 8.67. At the same time, we can apply the OR to the survey and the 95% CI range to the puget Sound population, but Jose is most correct when he says that the results apply to other populations.” Scoring: $a(1) + c(0) + e(0.5) = 1.5$ (incorrect understanding in the statistical representativeness from confidence intervals; indirectly indicates the understanding of 95% CIs)

Item No. = Q4

Knowing that the evidence from observational case control studies is associative rather than causal, Aaron wondered why epidemiologists often use observational studies rather than randomly manipulated experiments that can produce causal evidence. How would you help him understand the value of observational studies? In your answer, include the advantages and limitations of observational studies and under what conditions it is better to use them.

- a) Advantages: Articulating situations in which case control design is preferred over experimental design
 - 1. Ethics (+1): Articulating some ethical concern(s)
 - 2. Economical efficiency (+1): Articulating some economical efficiency
 - 3. The nature of observational study (+1): Understanding in the real world
 - 4. Other pragmatic reasons (+1): Articulating some other pragmatic reason(s)
- b) Limitations: Articulating limitations of case-control design or observational studies in general
 - 1. Confounding (+1)
 - 2. Association, not necessarily causation (+1)
 - 3. Controllability of variables (+1): Articulating that it is not possible to manipulate and control variables in research
 - 4. Others (+1): Articulating some other limitation(s)

Appendices

Examples:

“With an observational study there is no desired outcome. With an experiment variables will be changed to reach a desired outcome. Observational studies are also cheaper making them easier to fund”

Scoring: $a-2(1) = 1$ (“desired” or not is irrelevant)

“Epidemiologists are a different type of scientist, they study environmental things and they can not be controlled by manipulative variables or any sort of thing that can be done in a lab, they do real-world situations.”

Scoring: $a-3(1) = 1$

Item No. = Q5

Tyler has never used odds ratio calculation to analyze data from case control studies. He wonders why scientists use odds ratios to analyze data. How would you help him understand the value of using odds ratio to analyze data? Explain its purpose and what the number indicates.

- a) Quantifying (the strength of) association (+1): Articulating the purpose of using odds ratio in analysis
- b) Criteria for the strength of association (+1): Articulating the meaning of odds ratio when $OR > 1$, $OR = 0$, and/or $OR < 1$.
- c) Association rather than causation (+1): Articulating that OR indicates association, not necessarily causation; if only mentioned about association or correlation (+0.5)

Examples:

“An odds ratio can tell scientists whether data is significant or not. People like to see data in numbers, not necessarily just written out in a multi-page description. The odds ratio is a quick, easy way to see how a possible risk factor is associated with an outcome, if at all. The number indicates how many times more likely somebody is to have a certain outcome after being exposed. If the OR is below one, then it can be a protective factor. If the OR is equal to one, then the data is not significant. And if the data is greater than one then it is significant.”

Scoring: $a(1) + b(1) + c(0.5) = 2$ (note: the expression “significant” can be misleading, but I think it is fine for this question)

“Odds ratio shows if the numbers a person is studying is valid or not. If the odds ratio is clearly above 1 then it is a risk factor and it is positive. If it's below 1 then it is a protective factor and it could be negative. If the odds ratio is really close to 1 then we would say that the case is invalid.”

Scoring: $a(0) + b(0.5) = 0.5$ (note: it is not the point whether it is “valid” or not)

Item No. = Q6

Sandra has never used 95% confidence intervals (CIs) to analyze data from case control studies. She wonders why scientists need it in addition to odds ratio to analyze data. How would you help her understand the value of using 95% CIs? Explain its purpose and what the interval indicates.

- a) Significance/applicability of odds ratio to the study population (+1): Articulating the purpose of using 95%CI in analysis
- b) Meaning of 95%CI (+1): Articulating that the true OR lies inside the interval with the 95% confidence – and/or it could be false-positive by the 5% chance
- c) Criteria for statistical significance (+1): Articulating the condition of significant significance (e.g., one is outside the interval)

Example:

“Scientists use 95% confidence intervals (CIs) along with their odds ratio because if an odds ratio is 1.0 then there is no association, so if the odds ratio is 2.5 but the CI is [0.7,4.3] then there is also no association. This is because the data has at least a few test subjects whose odds ratio is 1.0 or at least very close to it. This would mean that their odds ratio has no association, and so if there are subjects whose odds ratios have no association, then the exposure has no association, and so the CI can tell you whether or not the odds ratio is applicable to the rest of a population, rather than just being a fluke in one case study.”

Scoring: $a(1) + b(0.5) + c(0.5) = 2$ (note: the last point is right, but the explanation is difficult to understand; he/she indirectly points to c), but it is not explicit enough)

“95% confidence interval is the probability in which most of the data relies on, so it is not showing outliers. It indicates where most numbers of the population rely on.”

Scoring: $a(0.5) = 0.5$ (note: point to the concept of population, but its explanation is insufficient enough)

Item No. = Q7

Steve has never used the criteria for causality to analyze data from observational case control studies. He wonders why scientists need it in addition to odds ratios and 95% confidence intervals. How would you help him understand the value of using the criteria for causality? Explain its purpose and the limitations of odds ratio and 95% CI.

- a) Distinguishing causation from association (+1): Articulating the purpose of using the criteria for causality
- b) Two or more specific criteria for causality (+1): Articulating at least two criteria for causality
- c) Limitations of OR and 95%CI (+1): Articulating OR and 95%CI indicate the strength and significance of association, not necessarily causation

Appendices

Examples:

“The reason for the use of the criteria for causality is to help figure out if the study that was made was actually correct in the sense that it was able to be observed and with accurate results.”

Scoring: $a(0) = 0$ (note: whether it is “accurate” or not is insufficient)

“One important criteria for causality is bias. If someone were to conduct a study on heronine addiction or obesity using only white caucasian males, you could come to the conclusion that both issues are predominant in white and male populations.

Another important criteria for casality is temporal sequence. Using the smoking study as an example, if you don't clearly identify the age of the participants, you may miss a variety of other "age-related" factors, such as work and marriage (not usually present in teens".

Scoring: $b(0.5) = 0.5$ (note: the explanation for bias is somewhat unclear)

“The criteria for causality is used to make sure that the CI and odds ratio calculated from a study are not being influenced by untested/unknown/unrecognized variables. If the data is being manipulated by other variables that were not brought up in the study, then it is invalid and cannot be applied to a larger population, which is the goal for ecology studies.”

Scoring: $a(0.5) + b(0.5) = 1.5$ (note: it indicates distinguishing causation from association through confounder, it is not so clear)

Item No. = Q8

Chloe wonders why tobacco/smoking research matters to society. How would you help her understand the value of this research? Explain problems caused by tobacco/smoking and the contributions of tobacco/smoking research.

- a) Problems in smoking (+1): Articulating one or more issues caused by smoking
- b) Contributions of smoking research (+1): Articulating the contributions of tobacco research (e.g., prevention)

Examples:

“It is important to society because tobacco has been shown to cause cancer, and is very addictive, and cancer is a leading cause of death all over the world. If something, one thing, is killing so many people than it makes sense that it would have a lot of attention drawn to it.”

Scoring: $a(1) = 1$

“Tobacco/smoking causes cancer and many other diseases. With this research, we can find what causes people to smoke, and then create laws to end the factor and get people to stop smoking.”

Scoring: $a(1) + b(1) = 2$

Appendices

Item No. = Q20

Shelley is confused with the meaning of “statistically significant”, how would you help her understand what it tells you? Explain using three key phrases: “odds ratio (OR)”, “95% CI”, and “OR = 1”.

- a) Criteria for statistical significance (+1): Articulating the condition (e.g., one is outside the interval)
- b) Statistical representativeness (+1): Articulating that it evaluates the applicability of OR to the study population

Examples:

“If there is a one in the CI, then the data isn't statistically significant.

If the OR is > 1 , there is a positive association.

If the OR=1, there is no association.

If the OR < 1 , there is a negative association.”

Scoring: a(1) = 1 (note: the strength of association is not directly relevant to this question)

“Statistically significant means that the data can be applied to other populations. If the either the OR or 95% CI are close to, or equal to 1, then the data is not statistically significant. This is because OR = 1 shows that the exposure and the outcome have no association, and it is the same if the CI = 1.”

Scoring: a(1) + b(0.5) = 1 (the criteria is not so clear, but indicating a right answer; the applicability is to the study population, not “other” populations)