In this assignment, you will develop your Methodology section. This section follows your Introduction and Literature Review sections(I have included those papers, please stick to a more qualitative approach) At best, it should build on the information included in these sections. Having introduced a brief outline of your study in the introduction, you will flesh that out and provide details in the Methodology section. Having identified gaps and missing elements in the existing body of current research in your literature review, you will attempt to fill those gaps and address unanswered questions related to your hypothesis with this design.

The Methodology section should be 6–8 pages in length. Using the Methodology Scoring Guide as a reference and the APA Proposal Template, write a paper in which you discuss seven specific points:

1. Purpose of the study.
2. Research method and study design.
3. Research hypotheses and questions.
4. Target population and sampling techniques.
5. Instruments (or measures in quantitative studies).
6. Data-collection procedures.
7. Data-analysis procedures.

To ensure you are meeting the objectives of this assignment, please review the Research Methods Scoring Guide. In your paper, include in-text citations and a reference section at the end. Make sure your paper is formatted according to current APA guidelines. Refer to APA Style Central for guidance in using proper APA style. See the APA Style and Format section of the Writing Center for instructions on accessing and using APA Style Central.

Discussion Question: For this discussion, refer to the quantitative pre-experimental, true-experimental, and quasi-experimental research studies that you found in the Capella library (I have included those articles). Summarize each study briefly using the Article Analysis Worksheet and discuss the following aspects:

- What are the differences in sampling among the studies?
- How did the design (pre-experimental versus experimental versus quasi-experimental) differ among the three studies?
- What were the findings of each of the studies?

- What are the defining features of each that make it pre-experimental, true experimental, or quasi-experimental?

Research Paper Introduction: Investigation of Stress Sources, Coping, and Response Strategies Among Nursing Students

Student's Name

Institutional Affiliation

Research Paper Introduction

Statement of the Problem

Healthcare professionals face some of the highest levels of stress of all careers. This is particularly true to nursing students, who are not only confronted with academic stress, but also stressful conditions at work and during their practical training period. How they cope with these stressors may influence their performance and productivity (Pulido et al. 2012). These stressors normally adversely affect their mental and physical wellbeing. Some of the most common stressors that nursing students undergo include time pressures, work overload, decision-making challenges, constant changes and economic mistakes within the work environment. While nurses continue to face these challenges, scholars are yet to develop long-standing remedies to work- and training related stress for nursing students. Healthcare professionals acknowledge the need to develop prevention of physical risks to a global approach. In this case, the emerging psychosocial risks that stem from work-related stress are of specific interests to stakeholders interested in improving the physical and emotional wellbeing of nurses (Pulido et al. 2012). The contributions of nurses to patients' quality of life cannot be overlooked. Therefore, nurses must be in their right state of mind if they are to deliver the desired health outcomes for patients.

Stress can have adverse effects on nursing students. These problems range from individual level, physical, psychosocial, and behavioral disorders. In turn, stress is strongly related to organizations-level problems, such as increased levels of absenteeism, reduction in quality of life and a sharp decline in performance and productivity (Pulido et al. 2012). While stress negatively affects almost all disciplines and careers, the degree of stress that nursing students are exposed to are not only higher, but they stem from multiple sources. The negative effects of stress go beyond their work environments to include the poor academic performance and imbalance in work and life-related issues (Pulido et al. 2012). Stress in nursing is regarded as one of the most significant problems in the contemporary world. In particular, psychosocial stress stem from poor relations between the nursing student and their immediate environments. This problem emerges from environmental conditions that are that nursing students regard as taxing or exceeding their resources and endangering their wellbeing (Labrague, 2014). The existence of imbalance between their environmental demands and their perceived resources that the individual nurses have available to cater for these demands can also generate adverse effects. In the event that the demands exceed the resources, stress can take place at the individual levels. Secondary appraisal takes place when a person defines their capability to manage and control the environmental demands (Labrague, 2014). Healthcare professionals and policymakers have the task of addressing the mental, physical, and psychosocial stressors that affect nursing students for them to provide high quality of care to patients.

Background of the Problem

Stress has been acknowledged in the nursing literature as one of the most challenging problems that nursing students confront. Some of the most renowned studies on stress and coping strategies among nursing students can be traced back to the works of Gorostidi et al. (2007) and Singh and Kohli (2015). In their attempt to investigate the evolution and changes in nursing students' perception of stressors associated to clinical practices and environments, Gorostidi et al. (2007) found that the most powerful stressors that students complained about include lack of competence, uncertainties and impotence, fear of being subjected to harm as a result of the relationships and interactions that they have with patients, and extreme emotional engagements that they are subjected to within their clinical contexts. Therefore, scholars agree that stressors differ in the way that they impact on nursing students' ability to work effectively.

Top on the list of sources of stress include feeling of having less control over their relationships with patients, contacts with suffering, and relationships with their senators and tutors. Further, nursing students complain of work overload, which interfere with their goal of attaining work-life balance in their career and social lives. In the same vein, Singh and Kohli (2015) argue that professional learning can prove to be psychologically overburdening and mentally distressing. This can result in negative effects on the physical and mental wellbeing of students. Of much concern by scholars has been the need to examine the psychological wellbeing of nursing students in order to determine the extent to which they are affected by their professional environments. Singh and Kohli (2015) identified the levels of stressors that nursing students are exposed to, their stress response mechanisms, and coping strategies that are applied by nursing students in response to the distressing environments. Seeking social support have been identified as some of the main coping mechanisms that nursing students employ. As such, there is need to develop effective interventions to reduce stress among nursing students.

Stress can be defined as coercive interaction between an individual and the environment. This interaction is comprised of work-related boundaries that may be threatening to overcome the individual's resources and skills. Stress is not synonymous to nervousness or anxiety. However, it offers the originality, capability and energies, although it can result in fatigue and illness at mental and physical levels (Inyat & Paveen, 2017). Stress is a psychological variable that affects the educational performance of nursing students. While all areas of professionalisms are affected by stress, healthcare professionals run a higher risk of stress, particularly nursing profession. Stress and anxieties are not just a workplace challenge, but different levels of stressors might exert major impacts on the students' ability to operate within their practical environments (Inyat & Paveen, 2017). Although the main causes of stress among nursing students have been identified by Singh and Kohli (2015), other causes of stress include negative insights towards nursing profession, disrespectful attitude towards nurses, and embarrassing behavior. Nursing students also struggle to overcome the feeling of lack of self-confidence, negative attitude, and low temper. In addition, nurses may experience tiredness and sleep disorders, accompanied by low satisfaction with their performance environment (Inyat & Paveen, 2017). However, timely detection of causes of stress has the prospect of bringing positive change in the student's performance and the instructors' perception towards the students, which is often more aggressive and ignoring in nature. The stressors can be defined in terms of the conditions or actions that have the potentiality to affect human health. Timely identification and mitigation of these stressors can improve nursing student's level of performance and productivity (Inyat & Paveen, 2017). Therefore, there is need o identify mechanisms through which nursing students can cope with stress in order to deliver proper services and quality of life to patients and clients later in their careers.

Purpose and Significance of the Proposal

The primary objective of the study is to identify the leading causes of stress among nursing students and identify ways of mitigating and responding to them. This study is informed by the fact that reducing distressful environmental conditions in the nursing students' learning and clinical environments can help to improve their academic and professional performance. Stress is an inevitable part of lives of all individuals across the world. It arises and affects people's internal sense of wellbeing, how they relate with their family members, friends, coworkers, and patients. It also adversely affects their ability to function within their educational and career settings (Zamshidi et al. 2016). When excessive, it might result in mental and physical symptoms. When it is prolonged, stress might result in disability and illnesses. Stress has the ability to modify a person's perceptions, feelings, attitudes, and behaviors and affects nursing students' organizations and activities that they carry out. Therefore, identifying the leading causes of stress helps to determine ways of developing effective interventions in order to generate positive outcomes for nursing students. For instance, it is important to develop strategies of improving nursing students' work-life balance in order to reduce stress. Improved work-life balance results in multiple additional benefits for nursing students. For example, they can be better able to cater for the sick and their family members and be available for important family and friends' commitments. Nursing students also need to reduce the overload that they go through in order to be able to attend to their personal needs.

Reducing stress is also associated with improvements in nursing students' overall quality of life, which can translate to excellent academic and professional performance. Stress may be a stumbling block to students' realization of quality of life. However, gaining insights into mitigation strategies can contribute to improvements in quality of life for nurses owing to many positive effects such as feeling good about the self and easing oneself of burdens. Nurses should be given some degree of flexibility for them to perform their tasks optimally. Job stress can be reduced by ensuring that students are able to separate their work from non-career issues. Stress might stem from having to grapple with work intruding on home life and vice versa. In most studies that have been undertaken on stress that affect students, it is commonly agreed across the scholarly divide that nurses hold that their work arrangements and conditions are the leading promoters of stress. Therefore, it is important to improve the operating environments and climate of nursing students for them to perform well academically and within their nursing environments.

Summary of Research Methods and Design

Random stratified sampling procedure will be used to select the participants of the survey. A total of 100 nursing students from first to third year will be selected. Half of the participants will be comprised of males and the rest will be female participants. Thereafter, the students will be subject to a mixed-method design in which questionnaire and interviews will be used to gather their view about the leading stressors and the steps that they normally take to cope with them. All the leading stressors will be identified and ranked using a 5-point Likert scale from least stressful to most stressful. The study will also attempt to investigate the best coping mechanisms and implications for the field of nursing and healthcare in general.

Guiding Research Questions and Hypothesis

Research Questions

1. What are the leading contributors of stress among nursing students?
2. What are nursing student's coping mechanisms to these stresses?
3. What are the effects of these stressors on nursing students' academic and clinical environments?
4. How can policymakers and healthcare professionals mitigate or address these challenges?

Hypothesis:

1. Stressful environments reduce the academic performance of nursing students.

Assumptions and Limitations of the Study

Assumptions

The study is founded on the assumption that all the nursing students who have participated in the research also participate in clinical activities. Another major assumption of the study is that it applies to all nursing students at college and university levels. In addition, the study presumes that both male and female nurses are exposed to the same clinical and academic settings within which they deliver their roles and responsibilities. Thus, the strength of this study rests in its inclusion of junior and senior student nurses in the university as respondents in the investigation. This reduces likelihood of selection biases because respondents are not selected purposively.

Limitations

One major limitation of this study is that it is going to be conducted among student nurses from solely one institution of higher learning. Therefore, the "n" value is going to be small. The fact that students from other universities have been excluded has the potential to limit the generalizability of this investigative research (Salazar, Crosby & DiClemente, 2015). In addition, the research design may be affected by the responses of the respondents because they are requested to remember their past clinical experiences, including important stressful experiences that they underwent.

Operational Definition of Terms

Stress- Coercive interaction between an individual and the environment, affecting social, mental, and physical wellbeing of nurses. This interaction is comprised of work-related boundaries that may be threatening to overcome the individual's resources and skills.

Stressors- Factors that contribute to stress among nursing students.

Expected Findings

There are various issues that this study seeks to address. One expected outcome of this study is an exploration of the leading factors that generate stress among nursing students in colleges and universities. Another expected outcome is the identification of the degree of severity of each of the stressors that have been investigated and identified. This will be attained by ranking them in order of most stressful to least stressful accordingly. Further, the expected study outcome would be an exploration of all the coping mechanisms that students have employed to respond or reduce the adverse impacts of these stressors. Most importantly, the study attempts to determine methods that can be used by healthcare professionals to address the stressful challenges that nursing students undergo.

References

Gorostidi, X. Z., Egilegor, X. H., Erice, M. J. A., Iturriotz, M. J. U., Garate, I. E., Lasa, M. B., & Cascante, X. S. (2007). Stress sources in nursing practice. Evolution during nursing training. *Nurse Education Today*, *27*(7), 777-787.

Inyat, S., & Paveen, A. (2017). Evaluation of factors of stress among Nursing Students. *Advanced Practices in Nursing,* 23, 136.

Labrague, L. J. (2014). Stress, stressors, and stress responses of student nurses in a government nursing school. *Health Science Journal.*

Pulido□Martos, M., Augusto□Landa, J. M., & Lopez□Zafra, E. (2012). Sources of stress in nursing students: a systematic review of quantitative studies. *International Nursing Review*, *59*(1), 15-25.

Salazar, L. F., Crosby, R. A., & DiClemente, R. J. (2015). *Research methods in health promotion*. New Jersey: John Wiley & Sons.

Singh, N., & Kohli, C. (2015). Stress reaction and coping strategies among nursing students in Delhi. *Asian Journal of Nursing Education and Research*, *5*(2), 274.

Jamshidi, N., Molazem, Z., Sharif, F., Torabizadeh, C., & Najafi Kalyani, M. (2016). The challenges of nursing students in the clinical learning environment: a qualitative study. *The Scientific World Journal.*

Literature Review: Major Sources of Stress among Nursing Students and How these Stressors Evolve Over the Period of Nursing Training

Stress is an inevitable part of any academic experience. However, student nurses have additional stressors within clinical and learning experiences that make the nursing populace distinct from other academic and health-related lines of professionalism. In the last ten years, there has been a significant rise in body of literature that attempt to investigate stress, its impacts on academic and career performance, and how the stressors evolve over their period of nursing training. For instance, Sharma and Kaur (2011) sought to investigate the main factors that contribute to stress among nursing students. Using a quantitative design, the researchers selected a total of 37 student participants for the study. In the first stages of the survey, stress was evaluated with the use of a standardized scale. The second part of the research entailed the examination of stressors using subjects experiencing moderate and serious stress using a validated instrument. Sharma and Kaur (2011) found that 97 percent of the participants had moderate levels of stress while 3 percent experienced severe stress.

According to Sharma and Kaur (2011), the leading causes of stress included environmental factors (40 percent) closely followed by interpersonal factors at 30 percent. Although the researchers acknowledge that nursing students are often exposed to different forms of stressors, including pressure of academics with an obligation to succeed, their study revealed that academic factors only accounted for 19 percent of stress while intrapersonal issues had minimal contributions. Sharma and Kaur (2011) further define stress as an emotional and physical strain brought by people's reaction to pressures from the outside world. Examples of environmen5tal stress in their order of influence included: change in living environments, insufficient telephone and communication facilities, insufficient provision of safety and security measures, and insufficient recreational facilities and canteen/ mess. Examples of interpersonal factors identified included change in social activities, conflicts with colleagues and seniors, lack of cooperation from colleagues, and conflicts with roommates.

While Sharma and Kaur (2011) group the leading sources of stress in order of severity as environmental, interpersonal, academic, and intrapersonal factors, Pulido□Martos, Augusto□Landa and Lopez□Zafra (2012) state that academic factors are the main sources of stress followed by clinical factors. The objectives of their study were to examine the main causes of stress among nursing students and their subsequent evolution and progress during training periods. This survey was driven by the assumption that the degree of stress is high among health professionals than other experts. Pulido□Martos, Augusto□Landa and Lopez□Zafra (2012) argue that mos5t of these stressors have adverse health implications, and are found particularly among nursing professionals. According to Pulido□Martos, Augusto□Landa and Lopez□Zafra (2012), stress is a mental factor that influences the academic performance and wellbeing of nursing students.

In order to comprehensively examine these factors, Pulido□Martos, Augusto□Landa and Lopez□Zafra (2012) undertook a systematic review of the scientific literature on stressors among nursing students. Their searches included all the articles that were previously published at the end of 2010. The study findings suggest that the common causes of stress are mainly associated with academic matters. These academic factors include reviews, work overload, as well as problems related to studying. The second most serious cause of stress include clinical factors such as fear of unknown circumstances, errors with handling patients, and lack of familiarity with technical equipments. In general sense, Pulido□Martos, Augusto□Landa and Lopez□Zafra (2012) observed that there were no significant evolutions or changes that took place at different periods of students' education. Consequently, the researchers suggest that stakeholders in the healthcare and education sectors should consider the time spent by nursing students on the proposed tasks in order to avert work overload. Further, alternative arrangements of evaluations beyond the acquisition of skills for testing, such as learning-oriented evaluation can be helpful in minimizing academic stress among students and the reduction of negative mental symptoms.

In a different survey, Aedh, Elfaki and Mohamed (2015) identified busy schedules, critical thinking examinations, and clinical experiences in hospital environments as the leading causes of stress among nursing students. According to Aedh, Elfaki and Mohamed (2015), students are often overburdened by the many requirements of nursing curricula. All these elements combined with outside responsibilities such as family, children, and work issues, have the potential to bring about intense stress in students' lives. In order to affirm this argument, Aedh, Elfaki and Mohamed (2015) evaluate the levels of stress among nursing students apart from investigating the stress sources and their subsequent impacts. Using a cross-sectional and descriptive research design, the researchers evaluated the levels of stress that affect nursing students in Najran University. After employing a convenient sampling strategy, 50 participants were selected in their third year of study. After administering the participants with a self-reported questionnaire to gather data, the results indicated that nursing students in their third year of study exhibited the highest levels of stress in comparison with their counterparts in their first and second years respectively. The most common form of stressor identified by the students mainly stemmed from insufficient professional knowledge and skills, followed by stress emerging from burdensome assignments and work overloads. On the other hand, the least perceived sources of stress identified by the researchers included stress from peers and colleagues and dail6y life. The findings further indicated that the nursing students' academic year or study was directly correlated with the student's level of stress. The study undertaken by Aedh, Elfaki and Mohamed (2015) also delved into the coping strategies employed by students who are exposed to different types of stressors. In this respect, Aedh, Elfaki and Mohamed (2015) revealed that the main coping strategy employed by students was transferring the stress to other leisure activities. The researchers infer that nursing students are exposed to different types of stressors. These stressors have the potential to inhibit positive learning experiences and outcomes. In order to reverse these negative effects, the scholars propose the introduction of stress management workshops and programs to enlighten students on different ways in which they can deal with and adapt to the unprecedented stressors that pose obstruction to their learning experiences and career goals.

Taking yet another different position, Alzayyat and Al□Gamal (2014) argue that the stress that nursing students go through mainly stems from their clinical environment and education. In response to the widespread concerns by scholars about stress among nursing students, Alzayyat and Al□Gamal (2014) argue that clinical education and activities generate the most stressful experiences for nursing students. In light of the above, the researchers critically evaluate previous researchers associated with levels of stress and the types of stressors that can be found among undergraduate nursing students during their clinical activities. The data gathering methodology employed by these scholars included use of databases such as Medline, CINAHL, and PubMed to gather medical and psychological data respectively. The study outcomes revealed four major themes associated with stress. They include early cl8inical exposure, comparison between different academic years, cross-cultural comparison, and eustress elements of clinical experiences. The mos5t stressful part of their clinical education is the initial clinical experiences. According to Alzayyat and Al□Gamal (2014), the early periods of clinical exposure is highly depressing for nursing students. During these stages, the most contributing factors include insufficient knowledge and skills, taking time to attend to patients, and assignment overloads. Therefore, the researchers propose that stakeholders in the nursing education sector can exploit these findings to direct their students during clinical practice in order to avert stressful experiences. This can be instrumental in enhancing the clinical education initiatives in order to foster the psychosocial wellbeing of the students, and as such improve the patients' quality of life. The researche5rs also suggest that nursing educators should offer supportive evidence-based interventions for students from their initial time of clinical exposure. Most significantly, Alzayyat and Al□Gamal (2014) urge educators to persuade students to openly discuss their feelings and stressors in order to offer appropriate interventions.

Further, Behere, Yadav and Behere (2011) compare the levels of stress that nursing, medicine, and engineering students are faced with. In one of the few available comparative studies on the subject, Behere, Yadav and Behere (2011) conducted a cross-sectionl study wih the use of questionnaires that were administered to 100 randomly selected participants from the nursing, medical, and engineering colleges. In order to boost reliability, the researchers maintained the anonymity of the participants. Thereafter, they utilized stress measurement scale that had closed ended questions as metrics for evaluating the degree of stress that affected each student groups. The results of the comparative study suggested that stress in totality, is common across different students in all these lines of oprofesional8ism, regardless of their gender, ages, and socioeconomic statuses. However, nursing students experienced the highest levels of stress compared to their engineering and medical counterparts. The study also suggested that students in all these three major lines of professionalism expressed denial to the existence of problems, with utmost denial students from the nursing profession. In view of these findings, Behere, Yadav and Behere (2011) observe that there is an emerging pattern among students of overlooking the current stress, which pose serious challenges and might worsen their mental and psychosocial wellbeing.

Behere, Yadav and Behere (2011) assert that there is a significant increase in pressure and stress among nursing students. This could be associated with studying, examination, peer pressures, educators, and partners' pressures. Stress is sometimes hard to define. However, Behere, Yadav and Behere (2011) develop a definition that borrows highly from Selye's description of stress as the rate of wear and tear in the human body. Based on this understanding, the researchers regard stress as the external stimulus that brings about wear and tear or to the resultant internal damages. Psychological models of stress often subscribe to the transactional approach, which stresses on intervening psychological processes. In such models, stress is regarded only to take place when the individual perceives an external demand as extending their capacity to deal with it. Therefore, the personal assessment of the kind of the demand of the available resources and personal skills, and the assumed outcomes, determine whether stress is felt. This approach both recognizes and touches on the individual differences.

Reading from the same script, Schwarze and Gerler Jr (2015) urge stakeholders in the healthcare sector to seek urgent ways of reducing key stressors among nursing students in order to improve patient outcomes. Using an exploratory design, Schwarze and Gerler Jr (2015) examine the efficacy of modified mindfulness-based cognitive therapy intervention with the use of individual counseling sessions to minimize stress and improve levels of mindfulness among nursing students. In order to realize this goal, the scholars utilize an AB experimental research that was replicated thrice. Their study showed reduced stress in two out of the three pa5rticuipants and widespread mindfulness levels among all the three participants. The research implies that university and college counselors who work with clients in high-stress occupations should consider the use of this model to reduce stress. This is particularly true for nursing students, who are constantly exposed to distressing clinical environments during most times of their studies. The mindful-based cognitive therapy (MBCBT) can be regarded as an important part of third generation of cognitive therapies. When examined together with dialectical behavioral therapy, the MBCBT integrates the concepts of mindfulness with standardized cognitive behavioral models. This framework can be traced back to the works of Kabat-Zinn's mindfulness-based stress minimization initiatives. It is an eight-week group-based intervention initiative that is comprised of mindfulness medication practices to support students to cope with stress and physical fatigue.

Finally, Turner and McCarthy (2017) argue that the stress that is experienced by nursing undergraduate students have the potential to inhibit their learning and increase attrition rates. To investigate this concern, the researchers selected a total of twenty-six intervention studies and examined them. Out of these studies, three major themes were identified: stressors, coping strategies, and appraisal. Most of these interventions sought to minimize cases of severity of stressors through the development of initiatives such as curriculum development and improve students' coping capabilities. While qualitative results indicated the existence of potential benefits of these strategies to reduce stress, there were no statistically significant studies that utilized reappraisal, either alone or when merged with other approaches. Turner and McCarthy (2017) also took note of progress from 2008 in terms of increased cases of studies and improved ut8ilization of validat5ed outcome evaluations. However, they also took note of the fact that the review needed further methodologically sound, sufficiently powered studies. This is particularly true when it comes to the need to use randomized controlled trials, which are required for the determination of the most appropriate interventions to tackle the problem of excessive stress and anxiety among undergraduate nursing students.

References

Aedh, A. I., Elfaki, N. K., & Mohamed, I. A. (2015). Factors associated with stress among nursing students (Najran University-Saudi Arabia). *IOSR Journal of Nursing and Health Science (IOSR-JNHS)*, *4*(6), 33-38.

Alzayyat, A., & Al□Gamal, E. (2014). A review of the literature regarding stress among nursing students during their clinical education. *International Nursing Review*, *61*(3), 406-415.

Behere, S. P., Yadav, R., & Behere, P. B. (2011). A comparative study of stress among students of medicine, engineering, and nursing. *Indian journal of psychological medicine*, *33*(2), 145.

Pulido□Martos, M., Augusto□Landa, J. M., & Lopez□Zafra, E. (2012). Sources of stress in nursing students: a systematic review of quantitative studies. *International Nursing Review*, *59*(1), 15-25.

Sharma, N., & Kaur, A. (2011). Factors associated with stress among nursing students. *Nursing and Midwifery Research Journal*, *7*(1).

Schwarze, M. J., & Gerler Jr, E. R. (2015). Using mindfulness-based cognitive therapy in individual counseling to reduce stress and increase mindfulness: An exploratory study with nursing students. *The Professional Counselor*, *5*(1), 39-52.

Turner, K., & McCarthy, V. L. (2017). Stress and anxiety among nursing students: A review of intervention strategies in literature between 2009 and 2015. *Nurse education in practice*, *22*, 21-29.

DISCUSSION QUESTION 1.

Review of Educational Research
September 2012, Vol. 82, No. 3, pp. 300–329
DOI: 10.3102/0034654312457206

Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis

Erin Marie Furtak
University of Colorado at Boulder

Tina Seidel
TUM School of Education

Heidi Iverson
University of Colorado Denver

Derek C. Briggs
University of Colorado at Boulder

Although previous meta-analyses have indicated a connection between inquiry-based teaching and improved student learning, the type of instruction characterized as inquiry based has varied greatly, and few have focused on the extent to which activities are led by the teacher or student. This meta-analysis introduces a framework for inquiry-based teaching that distinguishes between cognitive features of the activity and degree of guidance given to students. This framework is used to code 37 experimental and quasi-experimental studies published between 1996 and 2006, a decade during which inquiry was the main focus of science education reform. The overall mean effect size is .50. Studies that contrasted epistemic activities or the combination of procedural, epistemic, and social activities had the highest mean effect sizes. Furthermore, studies involving teacher-led activities had mean effect sizes about .40 larger than those with student-led conditions. The importance of establishing the validity of the treatment construct in meta-analyses is also discussed.

Keywords: science teaching, science learning, meta-analysis, inquiry-based teaching, student achievement.

In the past 50 years, major educational policy organizations in the United States and across the world have emphasized that students learn by engaging in the thinking processes and activities of scientists (e.g., American Association for the

Advancement of Science, 1993; Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009; National Research Council [NRC], 2001; Organisation for Economic Co-Operation and Development, 2009). This approach has often been described as inquiry-based teaching, or science as inquiry, and includes students drawing upon their scientific knowledge to ask scientifically oriented questions, collect and analyze evidence from scientific investigations, develop explanations of scientific phenomena, and communicate those explanations with their teacher and peers (NRC, 1996). Although inquiry has been emphasized as a popular reform in science education since the 1960s, it returned to prominence in the 1990s with the *National Science Education Standards* (NRC, 1996, 2001).

Despite this emphasis, the efficacy of inquiry-based teaching has been continually challenged (Kirschner, Sweller, & Clark, 2006; Mayer, 2004). Critics of inquiry-based teaching have argued that its minimally guided approach does not provide sufficient structure to help students learn the important concepts and procedures of science. These critics often characterize the inquiry-oriented teacher as staying in the background while students engage in self-guided, hands-on activities of dubious value (e.g., Kirschner et al., 2006). These critics have advocated for traditional, direct instruction in which teachers deliver content to students through carefully designed lectures and verification-style laboratory activities.

While the debate between inquiry-based and traditional instructional approaches has continued to simmer, researchers have investigated inquiry-based teaching reforms with particular interest in the specific features that appear to lead to increased student learning. These studies typically have an experimental or quasi-experimental design in which student outcomes are compared across two groups of students—a control group taught in a traditional lecture-based manner and a treatment group taught with some form of inquiry-based teaching. These studies assume that if inquiry-based teaching is indeed a more effective way to help students learn, students in the treatment conditions will outperform students in the control conditions on measures of conceptual understanding. In fact, a number of meta-analyses of reforms in science education have come to this conclusion (Bredderman, 1983; Schroeder, Scott, Tolson, Huang, & Lee, 2007; Shymansky, Hedges, & Woodworth, 1990; Weinstein, Boulanger, & Walberg, 1982).

Despite these findings, critiques of inquiry-based teaching have persisted, in part because of disagreements among researchers about what features define this instructional approach (e.g., Furtak, Shavelson, Shemwell, & Figueroa, 2012; Hmelo-Silver, Duncan, & Chinn, 2007). Inquiry has been defined in a number of ways, ranging from simple descriptions of students actively guiding their own learning with the teacher acting as facilitator to more elaborated lists of actions for the teacher, student, and curriculum (e.g., NRC, 1996, 2001). This situation is further complicated by the fact that the field of science education in the United States has moved away from using the term *inquiry* as the primary descriptor for science education reform and instead has focused on articulating how students can engage in science as a body of knowledge, a set of practices, and a process of participation in those practices (NRC, 2007). Indeed, the recently released *Framework for K–12 Science Education* focuses on identifying these practices, as well as crosscutting themes and core ideas for science education (Board on Science Education [BOSE], 2012).

The ways in which inquiry has been defined has consequences for the validity of meta-analyses that attempt to synthesize the causal inferences made by individual studies. For example, one study might define inquiry as students collecting data in response to a question their teacher had posed and contrast this with a condition in which students analyze data provided by their teacher. A second study might have the teacher guiding students to develop explanations about a scientific phenomenon and contrast this condition with students who also developed explanations, but with the support of an online tool. Although both studies belong within the domain of inquiry-based teaching, the activities and cognitive processes in which students engage are quite different. In addition, the role of the teacher varies between the conditions. These variations in the way that inquiry-based teaching is defined have consequences for the inferences made in research syntheses about the effectiveness of the approach (Briggs, 2008).

This purpose of this article is to reevaluate the effect of inquiry-based reforms on student learning of science in the 10 years (1996–2006) immediately following the publication of the *National Science Education Standards* (NRC, 1996). To do so, we will apply a framework for inquiry-based reforms based on two dimensions: (a) the cognitive domains of inquiry in which students are engaged and (b) the extent to which activities are guided by the teacher or student. These distinctions will be made for both the experimental and control groups.

Previous Inquiry Meta-Analyses

In past years, several meta-analyses have been conducted on studies of inquiry-based science teaching. These studies fall roughly into two categories: those that focused explicitly on the outcomes of reform-oriented versus traditional curricula and those that contrasted inquiry-based teaching with other approaches. These studies generally computed effect sizes by taking the difference between mean achievement outcomes for treatment and control groups and dividing by the standard deviation of the control group (e.g., Glass's Δ; Glass, 1976). The following section will briefly review these meta-analyses for the purpose of identifying how each one conceptualized inquiry-based teaching.

A set of meta-analyses conducted in the 1980s framed inquiry-based teaching as *innovative* (Weinstein, 1982), *activity based* (Bredderman, 1983), *process oriented* (Shymansky, Hedges, & Woodworth, 1990), and *discovery oriented* (Wise & Okey, 1983). Table 1 summarizes their mean effect sizes. Three studies contrasted reform-oriented with traditional curricula, presenting mean effect sizes for several types of outcomes (Bredderman, 1983; Shymansky et al., 1990; Weinstein et al., 1982). In addition, two meta-analyses coded subsets of studies that explored the extent to which the teacher or the student guided the activity and the level of inquiry (Lott, 1983; Wise & Okey, 1983). Mean effect sizes across these studies ranged from .06 to .35, with larger effects found for inquiry-discovery and guided exploration as compared to teacher direction and structured guidance (Lott, 1983; Wise & Okey, 1983). Medium-level inquiry had a larger effect on student outcomes than low-level inquiry (Lott, 1983).

The most recently published meta-analysis on science teaching, conducted by Schroeder et al. (2007), reviewed experimental and quasi-experimental studies.

TABLE 1
Effect sizes of previous inquiry meta-analyses

Study	Contrast	*N* studies	Effect size estimate
Weinstein, Boulanger, and Walberg (1982)[a]	Innovative versus traditional curricula	151	.31
Bredderman (1983)[b]	Activity-based versus non–activity-based elementary curricula	57	.35
Shymansky, Hedges, and Woodworth (1990)[b]	Process oriented versus traditional curricula	136	.26
Wise and Okey (1983)[b]	Total	309	.35
	Inquiry-discovery	38	.41
	Teacher direction	28	.18
Lott (1983)[b]	Composite	39	.06
	Structured	8	–.14
	Guided exploration	15	.43
	Low level inquiry	12	.29
	Medium level inquiry	5	.41
Schroeder, Scott, Tolson, Huang, and Lee (2007)[c]	Inquiry strategies	12	.65

a. Weighted.
b. Unweighted.
c. Mean of both weighted and unweighted effect sizes.

Their coding system included inquiry strategies that they defined as less recipe-style, more student-centered instruction in which students answer scientifically oriented research questions. The authors found an effect size of .65 for this subset of studies, a number that exceeds those found in the 1980s meta-analyses.

Although not a meta-analysis, the Inquiry Synthesis Project is the most recent effort to survey the results of studies of inquiry-based teaching (Minner, Levy, & Century, 2010). This project included 138 studies and did not exclude studies on the basis of design. As a result, unlike the preceding syntheses, their final analysis merged results of qualitative studies with experimental, quasi-experimental, and nonexperimental studies. Studies were coded according to design, methodological rigor, inquiry saturation, and impact. Within their larger sample, Minner et al. (2010) identified a subset of 42 comparative studies that contrasted inquiry-based teaching with other approaches. Of these studies, 55% found students in the condition with "higher inquiry saturation" to outperform students in the comparison group (Minner at al., 2010, p. 483). However, the authors did not calculate an effect size for the subset of these studies. Additionally, the authors also found that six of the subset of nine studies that directly studied student versus teacher direction found a "statistically significant increase in student conceptual learning when there was more student responsibility in the instruction (and higher inquiry saturation)" (Minner at al., 2010, p. 492). Overall, the authors argued that their synthesis indicates a "clear, positive trend" in favor of inquiry-based teaching" (p. 474).

Motivation for a New Meta-Analysis

The preceding summary illustrates the variability in the ways that inquiry-based teaching has been defined and coded, something that has a direct bearing on a meta-analysis's *construct validity*. Shadish, Cook, and Campbell (2002) defined construct validity as "the validity of inferences about the higher order constructs that represent sampling particulars across persons, treatments, outcomes and settings" (p. 17). In other words, the generalizability of the inferences one can make after combining effect sizes in a meta-analysis will be sensitive to the way that the sample of students has been selected, the way that the outcome variable has been measured, and the way that the treatment under investigation has been defined. It is important for meta-analysts to pay close attention to these implicit choices because when they have been carefully coded they can be used to help explain the variability in effect sizes from study to study. We argue that insufficient attention has been given to the operationalization of the inquiry construct in the case of prior meta-analyses of inquiry-based teaching and that this has masked important differences in the efficacy of distinct features of this instructional approach.

The majority of previous meta-analyses have relied upon expansive definitions of inquiry-based teaching (e.g., Bredderman, 1983; Shymansky et al., 1990; Weinstein et al., 1982). Although some have explicitly defined some aspects of inquiry teaching (e.g., Lott, 1983; Wise & Okey, 1983), others are much less clear about the parts that make up the whole (Schroeder et al., 2007). We argue that coding inquiry as a dichotomy, as opposed to existing on a spectrum, fails to capture the range of activities and thinking processes in which students might be engaged (Hmelo-Silver et al., 2007). The problem with using inquiry as a "black box" category is that it does not allow for distinctions between activities that are guided more by the teacher and those guided more by the student (NRC, 2001). Only a few of the prior reviews explore the amount of guidance provided by the teacher as a dimension of the reform being enacted in the study (Lott, 1983; Minner et al., 2010; Wise & Okey, 1983).

The shortcomings in the conceptualization of inquiry-based teaching in previous meta-analyses call into question the generalizability of claims being made about the effectiveness of inquiry as an instructional approach. In order to move research forward with regard to reforming traditional teaching, we need to know more about the specific effects of inquiry-based teaching on student learning by better articulating a model for this instructional approach. The following section will present a two-dimensional framework that we will use in this meta-analysis to compare and contrast the effects of inquiry-based teaching on student learning that have been found in the research literature.

Dimensions of Inquiry-Based Teaching

Part of the disagreement surrounding the effectiveness of inquiry-based teaching may be that the term *inquiry* itself has taken on multiple meanings in the science education reform literature. Inquiry has been alternatively used to describe (a) scientific ways of knowing (i.e., the work that scientists do), (b) a way for students to learn science, (c) an instructional approach, and (d) curriculum materials (Furtak et al., 2012). Some in the science education community have even argued that there is an inquiry threshold below which educational interventions can no longer be identified as inquiry based (Minner et al., 2010). Given this, it is important to be precise and transparent about how the construct of inquiry is to be defined when conducting a meta-analysis. In this study, we define inquiry in terms of two dimensions: the cognitive and social activities of the student and the guidance provided to students by their teacher, their peers, or curriculum.

Cognitive Dimension of Inquiry

For the cognitive dimension of inquiry, we draw upon Duschl's (2003, 2008) conceptualization of inquiry-based teaching and learning, which was developed on the basis of research and scholarly findings from cognitive and social psychology, the history and philosophy of science, and educational research. Duschl identified three categories of inquiry that included *conceptual* structures and cognitive processes that are used during scientific reasoning, *epistemic* frameworks used when scientific knowledge is developed and evaluated, and *social* interactions that shape how knowledge is communicated, represented, argued and debated.

The conceptual domain consists of the facts, theories, and principles of science, that is, science as a body of knowledge (NRC, 2007). All activities in science take place within an orienting framework of conceptual knowledge, connected not only to students' prior knowledge but also to the more sophisticated understandings they are expected to develop as a result of instruction. Atkin (1966) argued that the process of science is embedded within the content it seeks to enrich and define (Rutherford, 1960).

The epistemic domain is based upon students knowing how scientific knowledge is generated. Evidence collected through students' own scientific investigations is essential in this regard. Students should examine and evaluate the quality of evidence and then interpret evidence to develop explanations for phenomena (NRC, 1996, 2001, 2007). Then, a bridge should be provided to students so that they learn that their own process of collecting, evaluating, and interpreting evidence is similar to the practice of real scientists (e.g., Bell, Lederman, & Abd-El-Khalick, 1998). Furthermore, students should learn that scientific knowledge is subject to change in the face of new evidence or new interpretations of old evidence.

The social domain consists of the collaborative and communicative processes by which scientific knowledge is constructed, or the process by which students participate in scientific practices (NRC, 2007). Duschl (2003) described the social processes involved with communicating scientific ideas and understandings and emphasizes the importance of students making public their ideas through argument, modeling, and other modes of representation to help them learn to examine and evaluate their developing understanding of science. Since science is a collaborative enterprise, students should be encouraged to work in groups to reason collectively and reach decisions together.

To these three domains we add a fourth category, which we will call *procedural*, subdivided from Duschl's epistemic domain. The procedural domain describes the *methods* or "heuristics of discovery" (Bruner, 1961, p. 30): that is, asking scientifically oriented questions, designing experiments, executing procedures, and creating data representations (NRC, 2001; Schwab, 1962). This distinction is important to make because inquiry-based teaching reforms are often described as

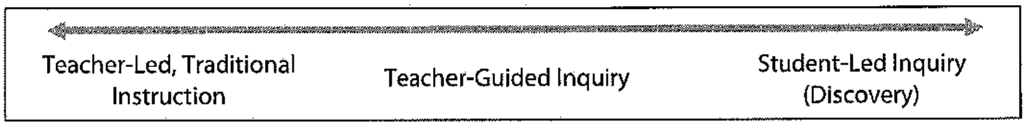

FIGURE 1. *Continuum of guidance in inquiry-based science teaching reforms.*

hands-on activities where students manipulate materials and collect their own data, independent of engaging in the process of evaluating the data they collected.

Guidance Dimension of Inquiry

In addition to the cognitive demands placed on the student as defined earlier, another feature of reform-based science teaching is a constant series of transitions of responsibility for learning from the teacher to the student, and then back to the teacher. Rather than being passive recipients of scientific knowledge, students in inquiry-based lessons are actively engaged in constructing their own understandings. Guidance in science education is often characterized as varying with respect to what is left open to students to define or provided by the teacher or curriculum (Schwab, 1962; Shulman & Tamir, 1973). The NRC's *Inquiry and the National Science Education Standards* defined a continuum of directedness across the "essential features" of inquiry, varying across the extent of learner self-direction and direction from the teacher or materials (NRC, 2001, p. 29).

Thus, the guidance dimension of inquiry distinguishes between reform conditions in which the teacher or student leads the activity. In this way, inquiry-based teaching can be thought of as part of a continuum of guidance. This continuum is bordered on one end by traditional *teacher-led* instruction in which the teacher tells students the answers they are expected to learn. At the other end of the continuum, students lead the activity, an approximation of discovery learning. We recognize a middle ground of *teacher-guided inquiry* in which the teacher actively guides students' activities in a reform-oriented science lesson (Figure 1).

Method

The purpose of this review is to compare and contrast the effects that have been found for inquiry-based teaching as a function of the different dimensions we have defined earlier. The sections that follow will describe in detail our literature search methods and selection criteria, coding process, and meta-analytic procedure.

Literature Search Methods and Selection Process

This article was developed as an extension of a larger meta-analysis (Seidel & Shavelson, 2007) that explored the effects of different teaching interventions on student learning across disciplines. The sample of studies selected for the present meta-analysis began with the subset of studies from that previous report, which were conducted in science education, and extended that sample with additional search terms and years of publication.

We chose to focus upon the effects of inquiry-based teaching during the 10 years following publication of the *National Science Education Standards* (NRC, 1996), a time period during which the reform spotlight was intensely focused upon this approach to science teaching reform (1996–2006). In this way, this meta-analysis presents a historical analysis of the effects of a particular approach during the time it was a popular "buzz word" in science education.

When adding to Seidel and Shavelson's (2007) list of search terms, we considered that science educators and science education researchers use a rich vocabulary beyond inquiry to describe the reforms they are developing and studying. To name just a few, reforms in science teaching in the past 15 years have been called inquiry-based teaching and learning (e.g., NRC, 2001), authentic inquiry (Chinn & Malhotra, 2002), model-based inquiry (Windschitl, Thompson, & Braaten, 2008), modeling and argumentation (McNeill, Lizotte, Krajcik, & Marx, 2006), project-based science (Singer, Marx, Krajcik, & Chambers, 2000), hands-on science (Pine et al., 2006), and constructivist science (Hardy, Jonen, Möller, & Stern, 2006). Klahr and Li (2005) argued that this profusion of terminology has made difficult the advancement of research in science education because of the lack of clear operational terms for what does and does not constitute a reform. In fact, Hmelo-Silver et al. (2007) argued that critiques of these approaches to science teaching (e.g., Kirschner et al., 2006) indicate a misunderstanding as to what constitutes each kind of reform.

To capture the diversity of terms used to describe inquiry-based reforms, we appended the keywords from Seidel and Shavelson's (2007) meta-analysis and conducted a new search in *Web of Science* and ERIC with the following teaching keywords: *effective instruction, instructional effectiveness, direct instruction, teacher effectiveness, mastery learning, constructivist teaching, science instruction, classrooms, science teaching*, and *inquiry*. We crossed each of these keywords with the output keywords *achievement, competencies, interest, motivation, engagement*, and *attainment*. We did not use the term *learning* alone because it yielded too many different kinds of studies.

In order to focus upon the effects of science teaching reforms on student learning in mainstream K–12 classroom settings, we excluded studies that did not have experimental or quasi-experimental designs and that focused specifically on students with disabilities, special education, or interventions that took place outside regular K–12 science classrooms. Our search terms yielded an initial sample of 1,625 studies. We iteratively reduced this initial sample to smaller and smaller pools of studies by reading abstracts and, when necessary, full papers, eliminating studies that did not satisfy our inclusion criteria. This process yielded a reduced sample of 59 papers. We then coded each paper according to features of its experimental design and outcome measure, retaining only papers published in English; those that had a pre-post, two-group design; and those that used a cognitive outcome measure with data provided to calculate an effect size. This left a total of 15 papers. Five of these studies did not include data necessary for calculating effect sizes. After contacting these authors, we were able to obtain necessary data for 4 papers. The first search thus yielded 14 papers for inclusion.

However, this initial search did not yield important experimental studies known to the first author, such as Chang and Mao's (1999) comparison of inquiry-group versus traditional instruction and Hardy et al.'s (2006) study of constructivist versus low instructional support. We therefore cross-referenced these articles with the original search terms and conducted an additional search with new terms. We

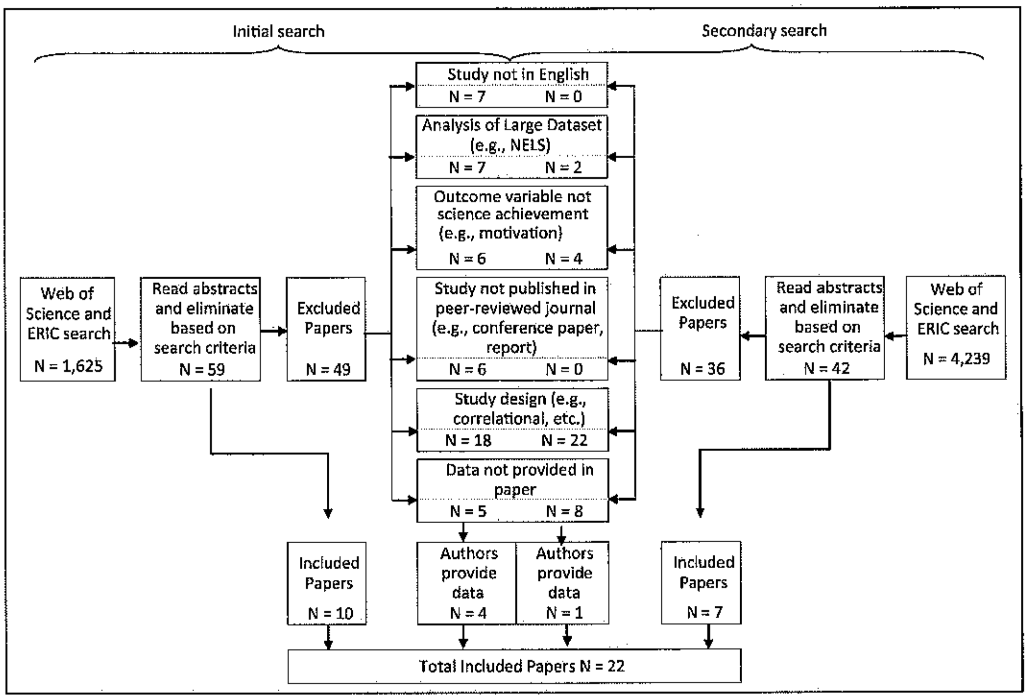

FIGURE 2. *Study selection flowchart.*

crossed *support, traditional,* and *instruction* with the reduced outcome search terms *understanding, achievement, attainment,* and *effect.* After reading abstracts and eliminating studies based on our inclusion criteria, this secondary search yielded 15 additional papers, 7 of which had complete data. After emailing authors we received data from 1 additional paper. Thus, in the end, our two searches yielded a total of 22 papers for inclusion in our meta-analysis (see Figure 2 for a schematic representation of the selection process).

As we describe earlier, we attempted to retrieve as many published studies as possible that met our search criteria. Despite the fact that our review of the literature was exhaustive, we are aware that we still may have missed some relevant publications. However, we believe that the search was thorough enough to present a collection of publications that is representative of the field between the years 1996 and 2006.

Multiple Contrasts

Multiple contrasts were reported in several of the papers in our final sample. To preserve these individual studies within papers, we included contrasts between different permutations of the treatment (e.g., teacher led vs. student led), contrasts that separated out the effects of the treatment in different subject areas, and contrasts that showed the effect of the treatment based upon multiple outcome measures. Three papers (Hardy et al., 2006; Huffman, 1997; Pedaste & Sarapuu, 2006) disaggregated outcome measures by specific item or content-specific groups of items. In these situations we reported only an aggregate for the outcome measure as a whole. These criteria yielded 37 studies from our sample of 22 papers.

TABLE 2
Codes and subcategories

Domain of inquiry	Description
Procedural	Asking scientifically oriented questions
	Experimental design
	Executing scientific procedures
	Recording data
	Representing data
	Hands-on
Epistemic	Nature of science
	Drawing conclusions based on evidence
	Generating and revising theories
Conceptual	Drawing on/connecting to prior knowledge
	Eliciting students' ideas/mental models
	Providing conceptually oriented feedback
Social	Participating in class discussions
	Arguing/debating scientific ideas
	Presentations
	Working collaboratively

Classification Process

A central question of this meta-analysis was whether the use of the cognitive and guidance dimensions of inquiry would help us better identify the instructional approaches that were contrasted in each study. As we described previously, each study contrasted different elements of these dimensions; thus, we coded the experimental and control condition for each study according to the cognitive and guidance dimensions to identify which dimensions were being explicitly contrasted.

Design and Structure Codes

Two coders used a process of consensus coding to code studies according to the research design, the way the studies self-identified their treatments (e.g., inquiry, discovery), the grade level of the study, and the country in which the study took place.

Inquiry Codes

The four inquiry domains were operationalized into subcategories, which were then used to code the description of the teaching in each condition. In this way, we were able to identify the contrast between the control and treatment groups (e.g., conceptual only control vs. full model of inquiry treatment). The inquiry codes are shown in Table 2. Two raters independently applied the inquiry codes to the experimental groups in 11 of the 22 papers (representing 19 of 37 studies, roughly half of the total sample). Raters used each code dichotomously to indicate the presence or absence of each of the four inquiry domains for a total of 44 codes given by each rater. The two raters applied the same code 33 out of 44 times, establishing 75% direct interrater agreement. Disagreements were resolved through adjudication, and one rater then proceeded to code the remainder of the studies.

Guidance Codes

In coding the studies, we identified four different kinds of guidance contrasts: traditional instruction versus student-led reform, traditional instruction versus teacher-led reform, student- versus teacher-led, and guidance dimension undefined. Traditional comparison groups were always treated as controls. As a rule of thumb, we labeled the student-led condition as the control in studies that contrasted student-led conditions with teacher-led conditions. We made this distinction following literature emphasizing the importance of teacher guidance in inquiry-based science teaching reforms (e.g., Hmelo-Silver et al., 2007). Studies falling into the undefined group contrasted some kind of visual representation (e.g., Kim, 2006) with another kind of representation or one computer-based intervention with another (e.g., White & Frederiksen, 1998), and thus guidance was not explicitly a feature of the study's design. In some studies, we coded the contrasts differently than the authors had identified them to ensure consistency across studies within our own sample. For example, Gläser-Zikuda, Fuβ, Laukenmann, Metz, and Randler (2005) identified their treatment as student centered, but the intervention was a phase of student-centered instruction followed by teacher-centered time focusing on correction of mistakes and assessment. Thus, we coded this study as teacher-guided reform. Two raters applied the guidance codes through a process of consensus coding.

Computing Effect Sizes

The effect size statistic allows for the comparison of treatment effects across studies where the dependent variable may not necessarily be operationalized in the same way (Lipsey & Wilson, 2001, p. 48). A commonly used effect size formula is Glass's Δ (Glass, 1976), represented as

$$ES = \frac{\left(\overline{X}_{Treatment} - \overline{X}_{Control}\right)}{SD_{Control}}, \tag{1}$$

where $\bar{X}$ represents the average scores for the treatment and control groups and $SD_{Control}$ is the standard deviation of the control group.

This formula can be altered slightly when different experimental designs are employed (see Lipsey & Wilson, 2001), such as in situations where the study involves pre- and posttest measures for both a treatment and control group (Becker, 1988; Morris, 2008). In these situations the effect size may be different from that calculated in the standard approach (i.e., Glass's Δ) if the mean pretest score and standard deviations across the treatment and control groups are taken into account. This is particularly important in our meta-analysis because over half of the studies (23 of 37) had higher pretest scores for the treatment group than the control group. In order to mitigate the threat of selection bias in these data, pretest results will be included in the effect size calculation. Morris (2008) conducted an analysis of several different methods of calculating effect size for studies with pretest-posttest-control designs and found the following equation to yield more precise effect size estimates as a result of a smaller sampling variance when compared with other approaches.

$$ES_{\text{Pre-/Post-Test Two Groups}} = \frac{\left(\overline{X}_{\text{T\_Post}} - \overline{X}_{\text{T\_Pre}}\right) - \left(\overline{X}_{\text{C\_Post}} - \overline{X}_{\text{C\_Pre}}\right)}{SD_{\text{Pooled\_Pre}}}, \quad (2)$$

In this equation, $\bar{X}$ represents a test score mean for treatment and control conditions (subscripts "T" and "C") administered at the beginning and end of a study period (subscripts "PRE" and "POST"), and *SD* is computed as the weighted average of the standard deviations across treatment and control groups.

However, Morris' (2008) equation is not directly applicable to our data as his analysis was conducted in the domain of organizational interventions. In the case of achievement data, one would expect variance to increase from pretest to posttest, and thus using the pooled pretest *SD* as the denominator would lead to inflated effect sizes. Indeed, Morris' approach yielded a larger effect size with greater variance as compared to using $SD_{\text{Control\_Post}}$ in the denominator.[1] Therefore, we used a modified version of Morris' equation as follows:

$$ES_{\text{Pre-/Post-Test Two Groups}} = \frac{\left(\overline{X}_{\text{T\_Post}} - \overline{X}_{\text{T\_Pre}}\right) - \left(\overline{X}_{\text{C\_Post}} - \overline{X}_{\text{C\_Pre}}\right)}{SD_{\text{C\_Post}}}, \quad (3)$$

Since all the studies in our meta-analysis—even those with random assignment—involve a pre-post, two-group design, and since the majority of posttest *SD*s were greater than pretest SDs, we calculated mean effect sizes using Equation 3. Use of $\text{SD}_{\text{Control\_Post}}$ is also consistent with the denominator used in the most recent meta-analysis of student achievement in science (Schroeder et al. 2007).

Although many meta-analysts often weight study effect sizes based on the number of students in the study, we have chosen not to do so for two reasons based on critiques of weighting by sample size. First, the concept of weighting by sample size stems from the assumption that each study can be considered a replication of the other and that each study is like a random draw from some implicit population of studies, with each sample of students a random draw from some implicit population of students; however, this assumption falls short since research studies take place at different times and at schools with vastly differing student populations and curricula. Second, weighting based on sample size is most relevant when the focus of a meta-analysis is to produce a single number that represents the main effect of a given intervention. The principal focus in this study is to examine the variability in effect sizes by the way that inquiry-based teaching has been defined.

Results

Descriptive Information

The studies represented a range of subject matters, were performed in a variety of countries, and examined teaching in elementary, middle, and secondary school. Descriptive codes are listed in Table 3 for all 37 studies from the 22 papers. Of the 22 papers included in our analysis, 5 presented studies conducted in only the

TABLE 3
List of included studies, codes applied, and effect sizes

Author(s)	Study number	Country	Subject matter	Level	Treatment	Control	Inquiry contrast	Guid-ance	Total *n*	Effect size
Alexander, Fives, Buehl, and Mulhern (2002)	1	United States	Astronomy	Middle	PECS	C	PES	1	110	1.737
Alexander et al. (2002)	2	United States	Astronomy	Middle	PECS	C	PES	0	126	0.956
Ardac and Akaygun (2004)	1	Turkey	Chemistry	Middle	C	C	None	3	49	0.622
Balci, Cakiroglu, and Tekkaya (2006)	1	Turkey	Biology	Elementary	PECS	C	PES	1	67	0.779
Balci et al. (2006)	2	Turkey	Biology	Elementary	EC	C	E	1	68	0.545
Cepni (2006)	1	Turkey	Biology	High	C	C	None	3	52	1.214
Chang (2002)	1	Taiwan	Earth science	High	PEC	PEC	None	2	244	0.043
Chang and Mao (1999)	1	Taiwan	Earth science	Middle	PECS	C	PES	0	612	0.148
Chang and Tsai (2005)	1	Taiwan	Earth science	High	PEC	PEC	None	2	347	0.007
Chang, Hsiao, and Barufaldi (2006)	1	Taiwan	Earth science	High	PECS	C	PES	1	155	0.049
Diakidoy and Kendeou (2001)	1	Cyprus	Astronomy	Elementary	ECS	C	ES	1	54	1.369
Ertepinar and Geban (1996)	1	Turkey	Physical science/intro chemistry	Secondary	PECS	C	PES	1	43	0.652
Gläser-Zikuda (2005)	1	Germany	Physics	Grades 8–9	CS	C	S	1	378	0.185

(continued)

TABLE 3 (continued)

Author(s)	Study number	Country	Subject matter	Level	Treatment	Control	Inquiry contrast	Guid-ance	Total n	Effect size
Gläser-Zikuda (2005)	2	Germany	Biology	Grades 8–9	CS	C	S	1	281	–0.005
Hardy et al. (2006)	1	Germany	Physical science	Elementary	PECS	0	PECS	0	95	0.236
Hardy et al. (2006)	2	Germany	Physical science	Elementary	PECS	0	PECS	1	102	0.251
Hardy et al. (2006)	3	Germany	Physical science	Elementary	PECS	PECS	None	2	125	0.013
Huffman (1997)	1	United States	Physics	High	CS	C	S	3	145	1.045
Huffman (1997)	2	United States	Physics	High	CS	C	S	3	145	0.206
Huffman (1997)	3	United States	Physics	High	CS	C	S	3	145	–0.240
Huffman (1997)	4	United States	Physics	High	CS	C	S	3	145	–0.168
Keselman (2003)	1	United States	Multivariable causality	Middle	PECS	PS	EC	3	46	0.625
Keselman (2003)	2	United States	Multivariable causality	Middle	PES	PS	E	3	46	0.786
Kiboss (2005)	1	Kenya	Physical science	Secondary	C	C	None	3	79	1.547
Kiboss (2005)	2	Kenya	Physical science	Secondary	C	C	None	3	79	0.929
Kim (2006)	1	United States	Earth science	Elementary	PC	PC	None	3	41	0.472
Lazarowitz, Baird, Bowlden, and Hertz-Lazarowitz (1996)	1	Israel/ United States	Biology	High	PCS	PC	S	0	113	0.187

(continued)

TABLE 3 (continued)

Author(s)	Study number	Country	Subject matter	Level	Treatment	Control	Inquiry contrast	Guid-ance	Total n	Effect size
Lazarowitz et al. (1996)	2	Israel/ United States	Biology	High	PCS	PC	S	0	113	–0.300
Pedaste and Sarapuu (2006)	1	Estonia	Environmental science and ecology	Grades 6–12	PEC	PEC	None	3	125	0.393
Reid et al. (2003)	1	United King-dom	Physics	Middle	PEC	P	EC	2	38	–0.040
Reid et al. (2003)	2	United King-dom	Physics	Middle	PEC	P	EC	2	38	–0.012
Reid et al. (2003)	3	United King-dom	Physics	Middle	PC	P	C	2	38	0.021
Ugwu and Soyibo (2004)	1	Jamaica	Biology	Middle	CS	CS	None	3	266	1.190
Ugwu and Soyibo (2004)	2	Jamaica	Biology	Middle	C	C	None	3	351	0.705
Ugwu and Soyibo (2004)	3	Jamaica	Biology	Middle	C	C	None	3	315	1.617
Uzuntiryaki and Geban (2005)	1	Turkey	Chemistry	Middle	EC	C	E	1	64	0.920
White and Frederik-sen (1998)	1	United States	Physics	High	PECS	PECS	None	3	302	–0.064

Note. P = procedural; E = epistemic; C = conceptual; S = social; 0 = traditional versus student led; 1 = traditional versus teacher led; 2 = student centered versus teacher centered; 3 = guidance not contrasted.

United States, 5 in Turkey, 4 in Taiwan, and 2 in Germany, with additional studies from Cyprus, Estonia, Israel/United States, Jamaica, Kenya, and the United Kingdom. This international sample of studies spanned a range of subject areas, with 4 taking place in biology, 8 in Earth and space science, 6 in physics/physical science, 3 in chemistry, and 1 that studied causality without a specific content focus. In terms of school type, 4 studies took place in elementary school, 7 in middle school, 7 in high school, and 4 in what was termed secondary school spanning Grades 6 to 12.

The number of papers published each year between 1996 and 2006 varied. Two studies were published in 1996; 4 papers were published each year in 1997, 1998, 1999, and 2001; 6 papers were published between 2002 and 2004; and finally, 10 papers were published between 2005 and 2006. This suggests a possible upward trend in experimental studies in response to growing governmental initiatives in the United States in the early 2000s (e.g., Shavelson & Towne, 2002).

Only three papers (Cepni, Tas, & Köse, 2006; Kiboss & Ogunniyi, 2005; Reid, Zhang, & Chen, 2003) involved the random assignment of students to conditions, or true experimental studies. A second group of papers ($n = 13$) involved random assignment of whole classes to conditions. Four papers involved nonrandom assignment of whole classes to conditions, or quasi-experimental studies, and the remaining two studies provided no information about assignment.

Only four of the papers included in this analysis self-identified as studying what they called inquiry-based teaching. The other papers used a variety of terms to describe the treatments, including cooperative learning, problem solving, discovery learning, instructional support, investigative approach, and constructivist learning. A subset of papers focused on computer-based interventions described as simulations and computer-assisted or multimedia approaches. The control groups in each study were described with a different set of terms: textbook approach, traditional instruction, direct instruction, individual mastery learning, regular instruction, baseline, and practice alone.

Overall Mean Effect Size

The distribution of individual effect sizes for all 37 studies included is shown in Figure 3. The overall mean effect size was .50 ($SD = 0.56$). This mean is greater than all of the mean effect sizes reported by the previous meta-analyses reviewed at the beginning of this article (see Table 2) with the exception of Schroeder et al. (2007), who found an effect of .65.

As Figure 3 indicates, although 7 studies had negative effects, 23 studies had effect sizes between 0 and 1, and 7 studies had an effect size between 1 and 2. When we explored the relationship between effect size and sample size, we found that the studies with larger sample sizes had effect sizes within the same range of the other studies with smaller sample sizes (Figure 4).

Effect Size Contrasts on the Cognitive Dimension

The first step in our analysis was to compare effect sizes using the cognitive dimension of inquiry-based teaching. The studies included in our meta-analysis used multiple reform terms to describe their treatment groups and multiple terms to describe their control groups. In this section, we present results of the meta-analysis according to the domain of inquiry being contrasted. As shorthand, we

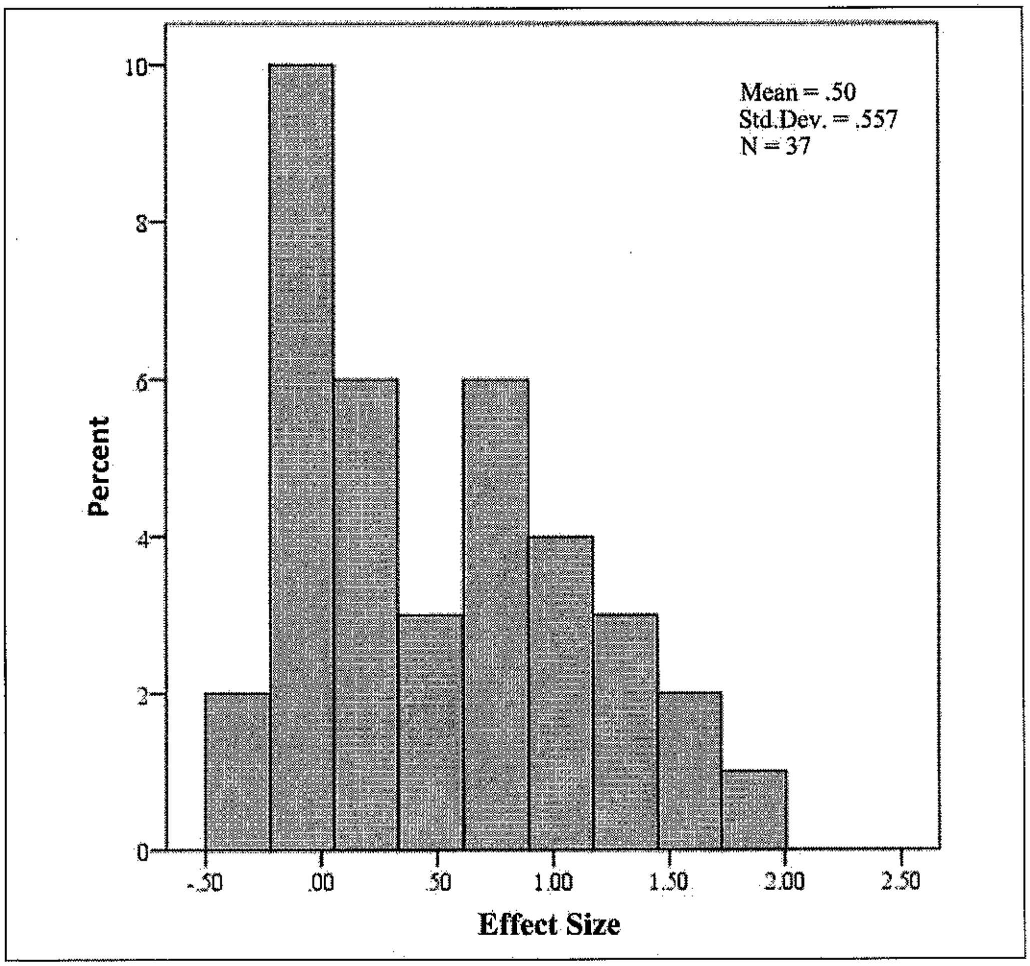

FIGURE 3. *Distribution of effect sizes (*N *= 37 studies).*
Note. Bin width equals one standard deviation.

will refer to each combination of domains by abbreviations. For example, a study coded for all four domains of the cognitive dimension—procedural, epistemic, conceptual, and social domains—will be referred to as PECS, and a study with only the procedural and conceptual domains will be referred to as PC. We will also provide examples from the studies that follow.

The most common code given to the treatment group was PECS; 10 of the 37 studies included this kind of treatment in its analysis. The next most commonly coded treatment conditions were CS (7 studies), C (6 studies), and PEC (5 studies). In total, there were 10 different combinations of the four domains represented in the treatment groups in the sample of 37 studies. With respect to the control group, 21 studies were only coded for the conceptual domain. In total, there were eight different combinations of the domains in the control groups. These results indicate that there were a variety of treatments and control groups represented in the sample of studies.

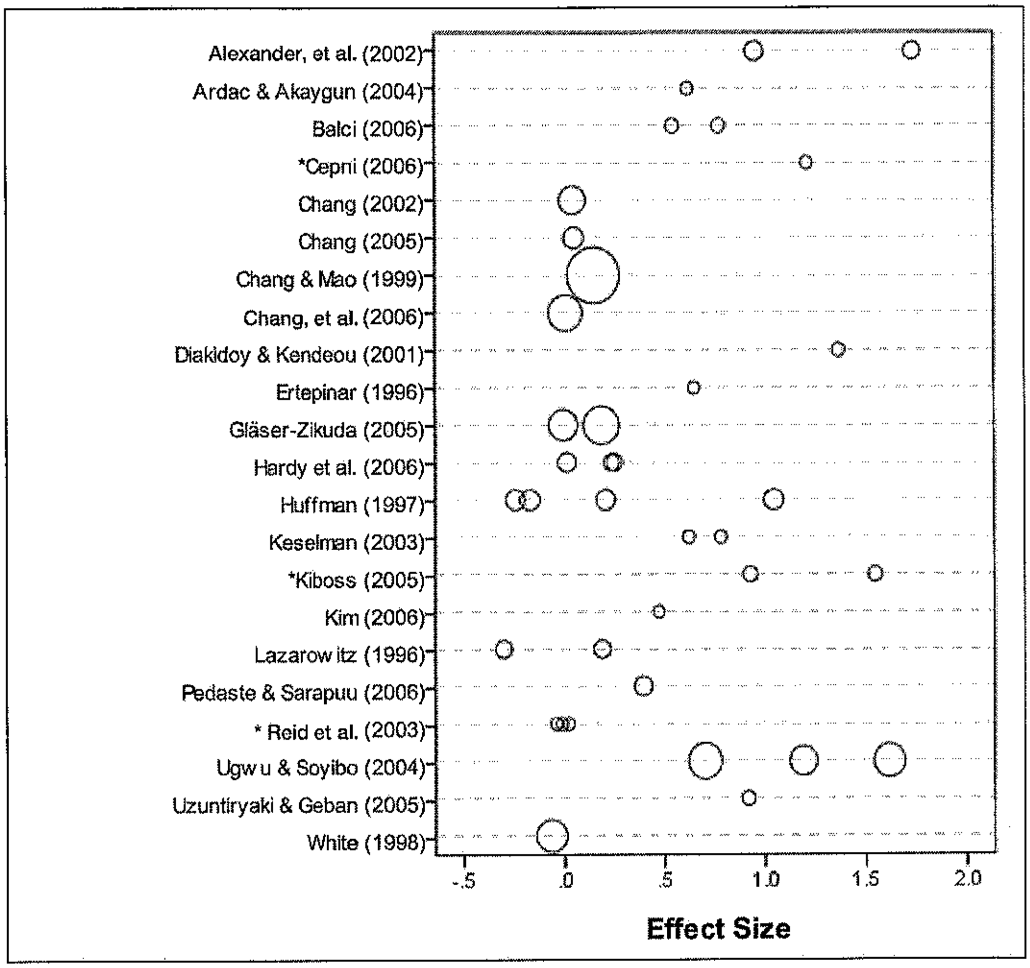

FIGURE 4. *Effect size by sample size.*
Note. Distribution of study effect size along with the relative sample size indicated by the size of the bubble. Bubble size indicates sample size; studies with random assignment are indicated with an asterisk.

We do not argue that the four cognitive inquiry domains were independent of each other; thus, we did not interpret effect sizes by inquiry domains included in the treatment. Rather, we explored the difference between the domain codes given to the treatment and control groups in each contrast. This difference tells us what elements of the intervention were being varied in the study, and thus to which domain of inquiry the effect of the intervention might be attributed. For example, the conditions contrasted in Chang and Mao (1999) were coded as PECS versus C, or a contrast of PES. In the experimental condition, students worked in teams to generate data (procedural), present and discuss data (social), and construct explanations on the basis of that data (epistemic), all embedded within the content of the earth-moon-sun system (conceptual). In the control condition, students experienced lectures based on the same concepts with occasional demonstrations for key ideas (conceptual). In the case of Diakidoy and Kendeou (2001), students in the experimental condition began by discussing their prior ideas about the shape of the earth and gravity and moved on to focusing on the concepts of speed of rotation

TABLE 4
Mean effect size by model of inquiry contrasted

Contrast	*N* studies	*N* papers	Min	Max	*SD*	Mean	Median
EC	3	2	–.04	0.63	.38	.19	–.01
S	8	3	–.30	1.05	.43	.11	.09
PECS	2	1	.24	0.25	.01	.24	.24
PES	6	5	.05	1.74	.61	.72	.72
E	3	3	.55	0.92	.19	.75	.79

Note: Overall mean effect size = .50 across the 37 studies. Table does not provide mean effect size for studies that did not explicitly study guidance or for which there was only one study in a category. P = procedural; E = epistemic; C = conceptual; S = social.

and perception of movement (conceptual), discussing the results of a demonstration (social) and formulating hypotheses and explanations about how day changes into night (epistemic). In the control condition, students read booklets that focused on the same concepts, where the teacher described and defined each concept, wrote those concepts on the board, and had students answer questions from the booklet orally and in writing (conceptual).

Surprisingly, for 13 of the studies, there was no difference in the inquiry domain codes in the treatment and control groups, and thus there was no contrast in the cognitive dimension of inquiry to report. For several of the studies (Ardac & Akaygun, 2004; Cepni et al., 2006; Chang, 2002; Chang & Tsai, 2005; Kiboss & Ogunniyi, 2005; Pedaste & Sarapuu, 2006), the contrast did not show up in the cognitive dimension codes because the contrast was between different types of computer-assisted environments or the same curriculum when it was given through regular teaching or led by a computer. However, two other studies (Hardy et al., 2006; White & Frederiksen, 1998) were coded as having PECS present in both the control and experimental groups and thus were contrasting features other than the cognitive domains of inquiry as defined for this analysis. In Hardy et al. (2006), the contrast between the conditions was based on the extent to which the student activities were guided by the teacher or students, and this contrast is captured with the guidance codes. In the case of White and Frederiksen (1998), the contrast between the conditions was in the extent to which metacognition was integrated into students' activities, a variation that is subsumed within our definition of the conceptual domain. However, our coding system did not capture this variation because the control condition in this study was also embedded in conceptual content.

For the remaining 22 studies, we were able to examine contrasts in treatment and control conditions as a function of our PECS codes. Table 4 shows that the 3 studies that explicitly contrasted the epistemic domain of inquiry had the largest mean effect size on student learning, followed by the 6 studies that contrasted the procedural, epistemic, and social domains. The most common contrast was social, which had the lowest effect size from the five contrasts presented previously.

Effect Size Contrasts on the Guidance Dimension

The next step in our analysis was to compare effect sizes using the guidance dimension of inquiry-based teaching. As described in the methods section, we

TABLE 5
Effect sizes by guidance contrasted in study

Guidance	*N* studies	Min	Max	*SD*	Mean	Median
Student led versus teacher led	6	−.04	0.04	.03	.01	.01
Traditional versus student-led reform	5	−.30	0.96	.45	.25	.19
Traditional versus teacher-led reform	10	−.01	1.74	.57	.65	.60

Note. Overall mean effect size = .50 across the 37 studies. Table does not provide mean effect size for studies that did not explicitly study guidance or for which there was only one study in a category.

TABLE 6
Median effect size by guidance and inquiry domain contrast

	Inquiry domain contrast					
Guidance	None	EC	S	PECS	PES	E
Traditional versus student-led reform			$n = 2$ −.06		$n = 2$ .55	
Traditional versus teacher-led reform			$n = 2$ .09		$n = 4$ .80	$n = 2$ .73
Student led versus teacher led	$n = 3$ .02	$n = 2$ −.03				
Guidance undefined	$n = 10$ .86		$n = 4$ .20			

identified four categories of guidance within the set of studies included in the meta-analysis: teacher-led reform, student-led reform, student versus teacher led, and other. A total of 16 studies that did not explicitly contrast guidance between the treatment and control groups were coded as guidance undefined.

The first three categories were applied as contrasts in the 21 studies shown in Table 5. Table 5 indicates that the 10 studies that explicitly contrasted teacher-guided reform versus traditional conditions had a higher mean effect size than the 5 that contrasted student-led reform versus traditional. The 6 studies that contrasted student- versus teacher-led reform had a mean effect size of .01. Finally, we compared the studies by guidance and inquiry domain contrast (Table 6). For the 10 studies that we coded as guidance undefined and that had no inquiry domain contrast, the mean effect size was large, suggesting that the computer-based interventions and representations explored in these studies successfully support students in learning science. We found the largest effect sizes for studies that contrasted traditional versus teacher-led reform and that varied PES between groups or E between groups (Figure 5).

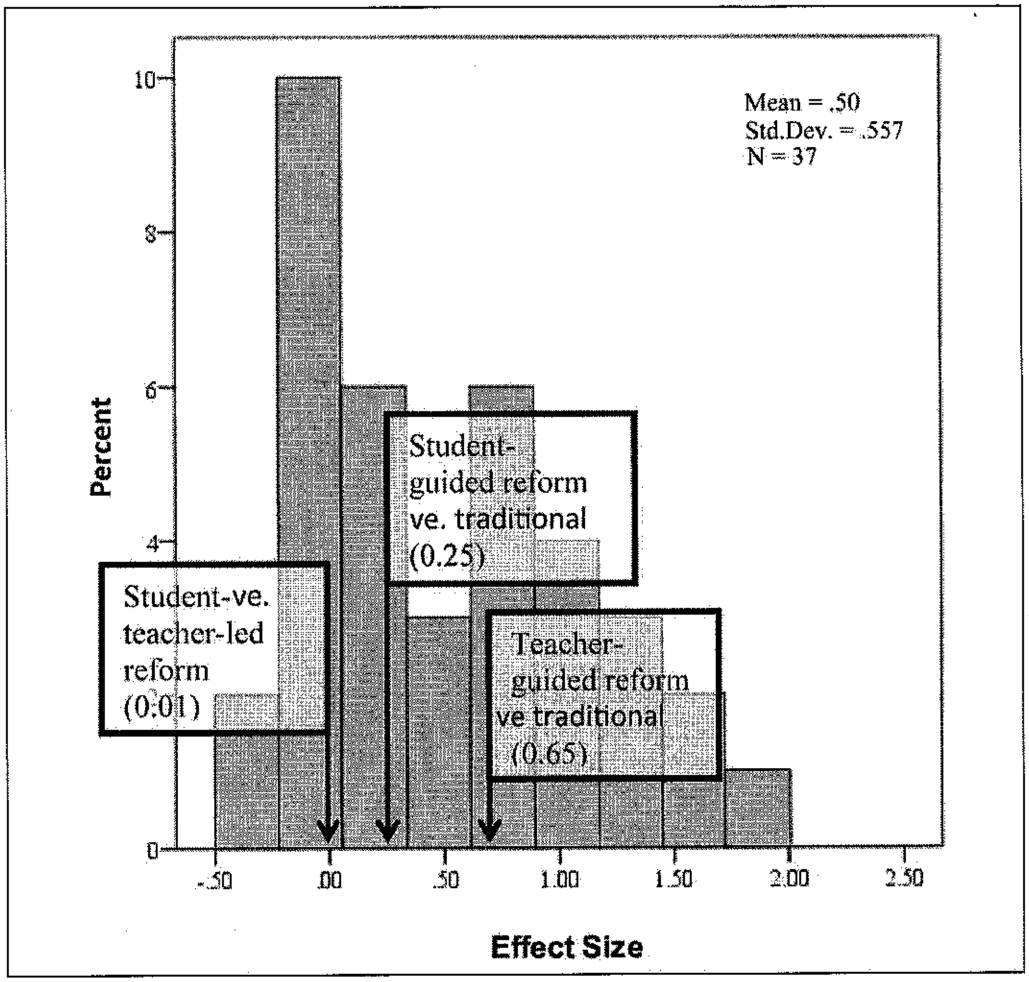

FIGURE 5. *Location of effect sizes for guidance contrasts.*

Discussion

The purpose of this review was to compare and contrast the effects that have been found for studies of inquiry-based teaching published in the 10 years following release of the *National Science Education Standards* (NRC, 1996). We separated inquiry-based science teaching into cognitive and guidance dimensions and used this framework to arrive at a more nuanced understanding of the construct of inquiry-based science teaching for the sample of studies identified for inclusion.

The 37 studies that met our inclusion criteria had a mean effect size of .50. This mean was higher than any of the means reported in meta-analyses conducted in the 1980s (which ranged from .06 to .43) and approaches the .65 mean effect size found by Schroeder et al. (2007), even though there is no overlap in the papers included in the two studies. We attribute this lack of overlap to differences in search terms and inclusion criteria. Schroeder et al. used a different set of search terms that focused more on study design and content area. In addition, Schroeder et al.'s inclusion criteria differed in two significant ways. First, Schroeder et al. had a wider range of search years (1980–2004) and included only studies performed in the United States. Since only one of the U.S. studies in our sample fell outside Schroeder et al.'s search years (Kim, 2006), the remaining lack of overlap in studies must be attributed to differences in search techniques and terms.

In the following sections, we begin by discussing limitations of our study sample, the need for data standards for publication, and construct validity and limitations of the meta-analytic approach. We will then summarize and discuss our findings and their implications for science teaching and learning. Finally, we will close with implications for reform science teaching terminology.

Limitations of Study Sample

Our sample was drawn exclusively from papers published in peer-reviewed, research-oriented journals. We made this decision to focus on papers that had met the levels of rigor necessitated by the peer review process. Since these journals often do not publish studies with negative results, one might argue that we missed out on a larger range of findings available in unpublished reports and theses. Despite excluding these reports, we still found several studies within our sample that did present negative results. Furthermore, our mean effect size was within the range of those found in previous meta-analyses with different selection criteria, suggesting that our sample of studies, while limited, was still roughly representative of the field.

Unfortunately, we found very few experimental and quasi-experimental studies performed in K–12 science classrooms during the years sampled. In addition, although the field of science education scarcely needs a reminder, our study reminds us yet again that reform-oriented science teaching practices are difficult to describe, difficult to enact, and even more difficult to characterize.

Implications for Meta-Analytic Approach

Need for Data Standards for Publication

An unexpected and disheartening finding of this analysis was the number of studies that met our design criteria for selection but did not present the necessary data for calculating effect sizes. Many of these studies presented group means with standard errors or skipped presentation of means altogether in favor of F, p, σ, and χ^2 values. We contacted the authors of 13 studies, and of the nine authors who responded, three could not send their data. Of those three authors, one could no longer access the original data, and one had lost the data. Although we acknowledge that these data may not always contribute to the substantive story being told through the answering of the research questions posed by original studies, and that strict word limits may force authors to omit data that they view as peripheral to their argument, omission of these data nevertheless makes it impossible to advance our understanding through the use of meta-analytic techniques. We argue that in order for the field to move forward, studies with experimental designs should always include basic relevant statistics such as means and standard deviations so that they may be included in future meta-analyses.

Construct Validity and Limitations of the Meta-Analytic Approach

A meta-analysis represents an attempt to synthesize causal inferences across studies, and each study will vary with respect to its context. Although a meta-analysis can facilitate causal generalization, these generalizations remain constrained by the quality and scope of the underlying studies that have been gathered (Briggs, 2008; Seidel & Shavelson, 2007; Shavelson & Towne, 2002). In this sense, we harbor no assumptions that this or any other meta-analysis will definitively answer the questions raised about a particular instructional approach. A concern about meta-analysis that we have tacitly raised in this study is that there may often be unexplored issues with regard to the construct validity of the treatment or intervention (in this case inquiry-based teaching). Nonetheless, this same concern could be raised with our attempt to establish a more refined coding approach since, obviously, we were not able to retrospectively attend these classrooms to determine for ourselves the extent to which any of the dimensions of inquiry were or were not enacted and instead had to rely on descriptions of the actual instruction for our process of coding, an inherent limitation in this necessarily retrospective research synthesis.

Furthermore, we experienced a tension between better defining the construct and the size of the sample of studies. On the one hand we wanted a better definition of the construct to put the studies into smaller bins that would reflect more accurately what students experienced in the studies. However, at the same time, by sorting into smaller categories we reduced the number of studies in each bin, making the computations of mean effect sizes across studies more sensitive to any single study with an unusually low or high effect size. This tension brought into sharp relief a tradeoff that must be navigated when conducting any meta-analysis.

Implications for Science Teaching and Learning

The results of this meta-analysis have several implications for the field of reform-oriented science teaching and learning. Overall, the results indicate a positive effect of inquiry-based teaching reforms on student learning of science. Second, results suggest the importance of the role of the teacher in actively guiding student activities in the context of inquiry learning. The current move in science education is away from using *inquiry* as a catch-all term for reform and toward more detailed descriptions of the practices in which the teacher and student engage (e.g., BOSE, 2012; NRC, 2007). This meta-analysis, by taking the approach of disaggregating different domains of a two-dimensional model of inquiry, has helped to identify particular features of those practices that appear to be more strongly linked to improved student learning of science. We discuss these features in the sections that follow.

Model of Inquiry

We found considerable variability among effect sizes when they were considered as a function of the cognitive and guidance dimensions of inquiry. For the 13 studies where the distinguishing cognitive feature of the inquiry treatment was EC, S, or PECS, the mean effect sizes were a modest .19, .11, and .24, respectively. Yet for the 9 studies where the distinguishing cognitive feature was E or PES, the mean effect sizes were about three times larger (.75 and .72).

With respect to our comparison of studies on the basis of model of inquiry, we found that studies contrasting the epistemic domain and the procedural, epistemic, and social domains had the largest effect size. This finding suggests that engaging students in generating, developing, and justifying explanations as part of other science activities is an important element to helping students learn science. This finding also supports multiple lines of research in the past 20 years that have argued for student engagement in evidence and explanation (NRC, 1996, 2001, 2007), as well as new policy directives in the conceptual framework for the new science standards (BOSE, 2012).

The next largest effect size was found for students that contrasted procedural, epistemic, and social. These studies usually had a traditional treatment group that involved students engaging in content only and therefore represent a classic comparison between a content-focused control and a treatment that integrates the other three domains of inquiry. The size of the effect of this subset of studies on student learning supports the contention of the science education community that inquiry-based teaching has a large, positive effect on student learning.

Surprisingly, a large number of studies in our sample ($n = 13$) did not contrast any of the four domains of inquiry we defined between the experimental and control groups. Given the stature of one of these studies within the science education community (White & Frederiksen, 1998), this finding suggests that there may be nuances in enactment of inquiry that were not captured by our coding system. However, many other studies that fell into this category studied representations that support student engagement in inquiry, but not inquiry itself, such as 2-D versus 3-D representations of the same content in support of inquiry learning (Kim, 2006) or diagrams to support students' conceptual understanding (Ugwu & Soyibo, 2004). The presence of so many studies in our sample indicate that many of the experimental studies performed in the 1996–2006 period did not actually study inquiry-based teaching and learning per se, but rather contrasted different forms of instructional scaffolds that did not substantively change the ways in which students engaged in the domains of inquiry, at least as those domains were conceptualized in this study. Future meta-analyses could explore the idiosyncrasies of these instructional scaffolds to better understand the ways in which these technologies can support inquiry-based teaching.

Model of Teacher Guidance

Critics of reform-oriented science teaching have argued that minimally guided instructional approaches are not successful in helping students to learn (Kirschner et al., 2006; Mayer, 2004). However, these critics have equated inquiry-based teaching, which is commonly conceived of as being supported by the teacher (Hmelo-Silver et al., 2007), with discovery learning, which has largely been discredited in mainstream science education circles as a means for getting students to learn specific scientific concepts (Shulman & Keislar, 1966). In this study, we considered discovery learning as occupying a position on the same continuum as teacher-guided inquiry, that is, a variation of the same reform, but with much less guidance provided by the teacher.

The subset of 10 studies that contrasted teacher-led reform conditions with traditional instruction had a mean effect size that was more than twice as large (.65) as the 5 studies that contrasted student-led reform with traditional instruction (.25). One might argue that in the conditions structured or guided by the teacher, students experienced the domains of inquiry more directly than in the student-led conditions; however, in the absence of detailed implementation data, we cannot draw this conclusion. Nonetheless, the evidence from these studies suggests that teacher-led inquiry lessons have a larger effect on student learning than those that are student led.

This finding illustrates an important nuance in the ongoing debate over the impact of reform-based teaching on student learning. Several papers published in recent years have argued for the failure of reform-based teaching featuring student guidance, notably Kirschner et al. (2006), Klahr and Nigam (2004), and Mayer (2004). However, Hmelo-Silver et al. (2007) argued that Kirschner et al. (2006) inappropriately grouped discovery learning with structured inquiry approaches that actually include guidance from the teacher. The results of this meta-analysis affirm this argument that engaging students in guided inquiry contexts does lead to learning gains when contrasted with comparison groups featuring traditional lessons or unstructured student-led activities.

Reform Science Teaching Terminology

Another finding of this analysis is the variety of terms used by the authors of these studies to describe the treatments and controls being contrasted, which made difficult the search and synthesis process. For example, some of the rigorous studies that featured instructional strategies related to inquiry-based teaching and learning did not label themselves as such (e.g., Hardy et al., 2006) while other studies that did label themselves as inquiry provided few details as to how inquiry was actually operationalized in the instruction (Pine et al., 2006). Furthermore, the different terms used to describe inquiry sometimes refer to the amount of guidance provided by the teacher (e.g., discovery learning) and other times refer to the cognitive activities of the student (e.g., model-based inquiry).

We suggest that greater consistency in keywords would make studies more accessible through search engines, and editors and reviewers could help authors more accurately select keywords for their work. The greater problem of inconsistency in reform science teaching jargon might be addressed through sustained discussion of what all of these terms mean, focusing on teaching and learning practices rather than labels alone. Furthermore, experimental and quasi-experimental studies might benefit by greater consistency and transparency in the type of instruction students experience in the control groups, not only to facilitate future research syntheses, but also to promote consensus within the field as to which approaches to science instruction are most effective.

Conclusion

In closing, we will summarize what we believe are three key findings of this meta-analysis. First, our synthesis of a historical sample of studies from the years in which inquiry-based teaching was returning to prominence indicates a positive effect of this teaching approach on student learning, with a particularly large effect of students engaging in the epistemic domain of inquiry and the procedural, epistemic, and social domains combined. Second, the meta-analysis also indicates higher effect sizes for studies that involved teacher-led activities. Third, and extending beyond the domain of inquiry-based teaching, this meta-analysis has illustrated how a refined model for an instructional approach can yield more nuanced interpretations of the effects of that approach on student learning.

Note

This research was initiated while Erin Marie Furtak was a visiting scholar at the Leibniz-Institute for Science and Mathematics Education (IPN) in Kiel, Germany, with the support of a German Chancellor Fellowship from the Alexander von Humboldt Foundation. The authors thank Manfred Prenzel for his support of this research project while Director of the IPN, Rich Shavelson and Katherine Eggers for reading and providing feedback on multiple drafts of this manuscript, as well as the editors and reviewers who provided helpful guidance in improving its quality and readability.

1. When comparing the three approaches to calculating effect size, we found that Equation 3 yielded the lowest effect size and variance as follows: Glass's Δ: .70 (SD = .86), $ES_{Pooled\_SD\_Pre}$ = .65 (SD = .73), $ES_{Control\_SD\_Post}$ = .50 (SD = .56).

References

Alexander, P. A., Fives, H., Buehl, M. M., & Mulhern, J. (2002). Teaching as persuasion. *Teaching and Teacher Education, 18*, 795–813. doi:10.1016/S0742-051X(02)00044-6

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York, NY: Oxford University Press.

Ardac, D., & Akaygun, S. (2004). Effectiveness of multimedia-based instruction that emphasizes molecular representations on students' understanding of chemical change. *Journal of Research in Science Teaching, 41*, 317–337. doi:10.1002/tea.20005

Atkin, J. M. (1966). Science education: "Process" and "content" in grade schools. *Science, 151*(3714), 1033. doi:10.1126/science.151.3714.1033

Balci, S., Cakiroglu, J., & Tekkaya, C. (2006). Engagement, exploration, explanation, extension, and evaluation (5E) learning cycle and conceptual change text as learning tools. *Biochemistry and Molecular Biology Education, 34*, 199–203. doi:10.1002/bmb.2006.49403403199

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257–278. doi:10.1111/j.2044-8317.1988.tb00901.x

Bell, R. L., Lederman, N. G., & Abd-El-Khalick, F. (1998). Implicit versus explicit nature of science instruction: An explicit response to Palmquist and Finley. *Journal of Research in Science Teaching, 35*, 1057–1061. doi:10.1002/(SICI)1098-2736(199811)35:9<1057::AID-TEA6>3.0.CO;2-C

Board on Science Education. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

Bredderman, T. (1983). Effects of activity-based elementary science on student outcomes: A quantitative synthesis. *Review of Educational Research, 53*, 499–518. doi:10.3102/00346543053004499

Briggs, D. C. (2008). Synthesizing causal inferences. *Educational Researcher, 37*, 15–22. doi:10.3102/0013189X08314286

Bruner, J. (1961). The act of discovery. *Harvard Educational Review, 31*, 21–32.

Cepni, S., Tas, E., & Köse, S. (2006). The effects of computer-assisted material on students' cognitive levels, misconceptions and attitudes towards science. *Computers and Education, 46*, 192–205. doi:10.1016/j.compedu.2004.07.008

Chang, C.-Y. (2002). The impact of different forms of multimedia CAI on students' science achievement. *Innovations in Education and Teaching International, 39*, 280–288. doi:10.1080/13558000210161052

Chang, C.-Y., Hsiao, C.-H., & Barufaldi, J. P. (2006). Preferred-actual learning environment "spaces" and earth science outcomes in Taiwan. *Science Education, 90*, 420–433. doi:10.1002/sce.20125

Chang, C.-Y., & Mao, S.-L. (1999). Comparison of Taiwan science students' outcomes with inquiry-group versus traditional instruction. *Journal of Educational Research, 92*, 340–346. doi:10.1080/00220679909597617

Chang, C.-Y., & Tsai, C.-C. (2005). The interplay between different forms of CAI and students' preferences of learning environments in the secondary science class. *Science Education, 89*, 707–724. doi:10.1002/sce.20072

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*, 175–218. doi:10.1002/sce.10001

Diakidoy, I.-A. N., & Kendeou, P. (2001). Facilitating conceptual change in astronomy: comparison of the effectiveness of two instructional approaches. *Learning and Instruction, 11*, 1–20. doi:http://dx.doi.org/10.1016/S0959-4752(00)00011-6

Duschl, R. A. (2003). Assessment of inquiry. In J. M. Atkin & J. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41–59). Arlington, VA: NSTA Press.

Duschl, R. A. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education, 32*, 268–291. doi:10.3102/0091732X07309371

Ertepinar, H., & Geban, Ö. (1996). Effect of instruction supplied with the investigative-oriented laboratory approach on achievement in a science course. *Educational Research, 38*, 333–341. doi:10.1080/0013188960380306

Furtak, E. M., Shavelson, R. J., Shemwell, J. T., & Figueroa, M. (2012). To teach or not to teach through inquiry: Is that the question? In S. M. Carver & J. Shrager (Eds.), *The journey from child to scientist: Integrating cognitive development and the education sciences* (pp. 227–244). Washington, DC: American Psychological Association.

Gläser-Zikuda, M., Fuß, S., Laukenmann, M., Metz, K., & Randler, C. (2005). Promoting students' emotions and achievement—Instructional design and evaluation of the ECOLE-approach. *Learning and Instruction, 15*, 481–495. doi:10.1016/j.learninstruc.2005.07.013

Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher, 5*, 3–8. doi:10.3102/0013189X005010003

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking." *Journal of Educational Psychology, 98*, 307–326. doi: 10.1037/0022-0663.98.2.307

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*, 99–107. doi:10.1080/00461520701263368

Huffman, D. (1997). Effect of explicit problem solving instruction on high school students' problem solving performance and conceptual understanding of physics. *Journal of Research in Science Teaching, 34*, 551–570. doi:10.1002/(SICI)1098-2736(199708)34:6<551::AID-TEA2>3.0.CO;2-M

Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching, 40*, 898–921. doi:10.1002/tea.10115

Kiboss, J. K., & Ogunniyi, M. B. (2005). Outcomes of first year secondary students in a computer-augmented physics program on measurement. *Learning, Media, and Technology, 30*, 313–326. doi:10.1080/17439880500251442

Kim, P. (2006). Effects of 3D virtual reality of plate tectonics on fifth grade students' achievement and attitude toward science. *Interactive Learning Environments, 14*, 25–34. doi:10.1080/10494820600697687

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86. doi:10.1207/s15326985ep4102_1

Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology, 14*, 217–238. doi:10.1007/s10956-005-4423-5

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*, 661–667. doi:10.1111/j.0956-7976.2004.00737.x

Lazarowitz, R., Baird, J. H., Bowlden, V., & Hertz-Lazarowitz, R. (1996). Teaching biology in a group mastery learning mode: High school students' academic achievement and affective outcomes. *International Journal of Science Education, 18*, 447–462. doi:10.1080/0950069960180404

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Lott, G. W. (1983). The effect of inquiry teaching and advance organizers upon student outcomes in science Education. *Journal of Research in Science Teaching, 20*, 437–451. doi: 10.1002/tea.3660200507

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*, 14–19. doi:10.1037/0003-066X.59.1.14

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences, 15*, 153–191. doi:10.1207/s15327809jls1502_1

Minner, D., Levy, A., & Century, J. (2010). Inquiry-based science instruction—What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*, 474–496. doi: 10.1002/tea.20347

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*, 364–386. doi:10.1177/109442810 6291059

Mullis, I. V. A., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Boston, MA: TIMMS & PIRLS International Study Center.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academies Press.

National Research Council. (2001). *Inquiry and the national science education standards*. Washington, DC: National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.

Organisation for Economic Co-Operation and Development. (2009). *PISA 2009 assessment framework: Key competencies in mathematics, reading and science.* Paris, France: Author.

Pedaste, M., & Sarapuu, T. (2006). Developing an effective support system for inquiry learning in a Web-based environment. *Journal of Computer Assisted Learning, 22*, 47–62. doi:10.1111/j.1365-2729.2006.00159.x

Pine, J. P., Aschbacher, P. A., Roth, E., Jones, M., McPhee, C., Martin, C., . . . Roth, E. (2006). Fifth graders' science inquiry abilities: A comparative study of students in textbook and inquiry curricula. *Journal of Research in Science Teaching, 45*, 467–484. doi:10.1002/tea.20140

Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning, 19*, 9–20. doi:10.1046/j.0266-4909.2003.00002.x

Rutherford, F. J. (1960). The role of inquiry in science teaching. *Journal of Research in Science Teaching, 2*, 80–84.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44*, 1436–1460. doi:10.1002/tea.20212

Schwab, J. J. (1962). *The teaching of science as enquiry.* Cambridge, MA: Harvard University Press.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*, 454–499. doi:10.3102/0034654307310317

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Shavelson, R. J., & Towne, L. (2002). *Scientific research in education.* Washington, DC: National Academies Press.

Shulman, L. S., & Keislar, E. R. (Eds.). (1966). *Learning by discovery: A critical appraisal.* Chicago, IL: Rand McNally.

Shulman, L. S., & Tamir, P. (1973). Research on teaching in the natural sciences. In R. M. W. Travers (Ed.), *Second handbook of research on teaching* (pp. 1098–1148). Chicago, IL: Rand McNally.

Shymansky, J. A., Hedges, L. V., & Woodworth, G. (1990). A reassessment of the effects of inquiry-based science curricula of the 60's on student performance. *Journal of Research in Science Teaching, 27*, 127–144. doi:10.1002/tea.3660270205

Singer, J., Marx, R. W., Krajcik, J., & Chambers, J. C. (2000). Constructing extended inquiry projects: Curriculum materials for science education reform. *Educational Psychologist, 35*, 165–178. doi:10.1207/S15326985EP3503_3

Ugwu, O., & Soyibo, K. (2004). The effects of concept and vee mappings under three learning modes on Jamaican eighth graders' knowledge of nutrition and plant reproduction. *Research in Science and Technological Education, 22*, 41–58.

Uzuntiryaki, E., & Geban, Ö. (2005). Effect of conceptual change approach accompanied with concept mapping on understanding of solution concepts. *Instructional Science, 33*, 311–339. doi:10.1016/j.compedu.2008.11.017

Weinstein, T., Boulanger, F. D., & Walberg, H. J. (1982). Science curriculum effects in high school: A quantitative synthesis. *Journal of Research in Science Teaching, 19*, 511–522. doi:10.1002/tea.3660190610

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*, 3–118. doi:10.1207/s1532690xci1601_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*, 941–967. doi:10.1002/sce.20259

Wise, K. C., & Okey, J. R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. *Journal of Research in Science Teaching, 20*, 419–435. doi:10.1002/tea.3660200506

Authors

ERIN MARIE FURTAK is assistant professor of education with a specialization in science education in the Department of Curriculum and Instruction at the University of Colorado at Boulder, 249 UCB, School of Education, Boulder, CO 80309; e-mail: *erin.furtak@colorado.edu*. A former high school science teacher, she received her PhD in curriculum and teacher education from Stanford University. Her research explores issues pertaining to the training and retention of secondary science teachers, and the improvement of student learning through reform-based science teaching.

TINA SEIDEL is the Friedl Schöller Chair of Teaching and Learning Research at TUM School of Education, Technische Universität München, Germany. She finished her PhD with excellence at the Leibniz-Institute for Science and Mathematics Education in Kiel, Germany. She is the author of numerous books and articles on the topics of classroom research, teacher and teaching effectiveness, teacher learning, and teacher professional development. Her special interest is in video-based analysis of classrooms, teaching adaptivity and aptitude-treatment-interaction and development of multimedia-based tools in teacher research.

HEIDI IVERSON is a postdoctoral researcher at the School of Education and Human Development at the University of Colorado, Denver. She received her PhD in curriculum and instruction with an emphasis on science education from the University of Colorado at Boulder. Her research interests center around the development and evaluation of innovative approaches to teaching in the college environment, specifically in physics education, and more broadly on innovative approaches to instruction and assessment in science education.

DEREK C. BRIGGS is associate professor and chair of the Research and Evaluation Methodology Program at the University of Colorado at Boulder, where he also serves as an associate professor of quantitative methods and policy analysis. He received his PhD from the University of California at Berkeley. In general, his research agenda focuses upon building sound methodological approaches for the valid measurement and evaluation of growth in student achievement. His daily agenda is to challenge conventional wisdom and methodological chicanery as they manifest themselves in educational research, policy, and practice.