Summer August 21st, 2015

A Meta-Analysis of the Effects of Problem- and Project-based Learning on Academic Achievement in Grades 6-12 Populations

Kimberly J. Jensen
Seattle Pacific University

A Meta-Analysis of the Effects of Problem- and Project-based Learning on Academic Achievement in Grades 6-12 Populations

by

Kimberly J. Jensen

Dissertation presented to the

Faculty of the Graduate School of Education at

Seattle Pacific University

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy in Education

May 2015

A Meta-Analysis of the Effects of Problem- and Project-based Learning on Academic Achievement in Grades 6-12 Populations

by

Kimberly J. Jensen

A dissertation submitted in partial fulfillment of the degree of

Doctor of Philosophy in Education

Seattle Pacific University

2015

Approved by ____

Arthur K. Ellis, EdD, Chair, Dissertation Committee

____

William Nagy, PhD, Dissertation Committee

____

John B. Bond, EdD, Dissertation Committee

Program Authorized to Offer Degree

____

Date ________________________

____

Rick Eigenbrood, PhD, Dean, School of Education

Copyright Page

Signature ________________________________

Date ________________________________

Table of Contents

List of Figures

List of Tables

List of Appendices

.

Acknowledgements

This dissertation is the product of encouragement and support of many individuals. I cannot even begin to thank each one enough.

To my husband, Shane, thank you for supporting my educational journey, keeping me grounded, and protecting my time. Your unfailing love continues to amaze me. I am looking forward to post-dissertation "freedom" and new adventures with you. Thank you to my children, Zachary and Jessica: your love, patience, and encouragement keep me fueled. Jessica, the eleventh hour support was much appreciated! Thank you to my parents, Keith and Jan Possinger, for believing in me, listening to my ideas as I sorted out my thoughts, and praying continually for wisdom and stamina.

I am indebted to Dr. Arthur K. Ellis, my dissertation chair. Your insight, encouragement, and careful and direct criticism assisted me in being a better writer, researcher, thinker, and practitioner. Thank you for turning my ideas into a vision and my vision into a reality. I am still the "live wire" you met three years ago! I am also grateful to several professors at Seattle Pacific University for pushing me to be my best. In particular, thank you Dr.'s William Nagy and John B. Bond for serving on my dissertation committee, providing me with additional sources and insight, reading drafts, and calming my Type A personality when it needed tempering. Many thanks to Mrs. Cynthia Strong, SPU SOE Librarian and student advocate. Your research support, encouragement, boldness, and prayers were and are much appreciated.

Dr. Alison Brynelson, I cannot begin to quantify what your friendship and support means to me. We passed comprehensive exams together, read each other's papers and dissertation drafts, agonized over stats together (all for "fun!"), and exhausted a couple of Starbucks locations. Michael Gallagher, your encouraging words, prayers, and flexibility were sanity savers in more ways than you know. Alison and Michael, thank you for setting me on my new professional journey! Alex Johns, thank you for the encouragement, reading drafts, and allowing me to freak out in your office when needed. Thank you to my dear friend, Ivonna Seaton, for believing in me and always picking up from where we left off, even after long expanses of time.

It is without question that this dissertation would not be possible without the expertise and diligence of Dr.'s James T. Rosenbaum and Thomas S. Hwang of Casey Eye Institute, Portland, Oregon. Thank you for the gift of sight and finding solutions to otherwise unsurmountable eye issues. Dr. Hwang, I promise to rest my eyes after this!

The process and product of this meta-analysis is the result of an outpouring of support, encouragement, and contributions from the PBL and educational research community. I would like to acknowledge the following individuals for contributing unpublished or in press articles and dissertations, and/or further information, data, or leads on already published studies: Dr. Charles Igel, Dr. Betina Malhotra, Dr. Deanna Kuhn, Frank Rösch, Dr. Kobus Maree, Dr. Mahnaz Moallem, Dr. Casey Creghan, and Dr. Terry Goodin. Your assistance and thoughtfulness demonstrates the collegiality of educational researchers.

Finally, but foremost, thank you to my Heavenly Father for giving me the capacity to accomplish this endeavor. Your promises never fail. (Proverb 19:21; Isaiah 46:4).

This dissertation is dedicated to my "study buddy," Sophie: in my heart forever.

Abstract

A Meta-Analysis of the Effects of Problem- and Project-based Learning on Academic Achievement in Grades 6-12 Populations

by

Kimberly J. Jensen

Seattle Pacific University Dissertation Chair: Dr. Arthur K. Ellis

Researchers and proponents of problem- and project-based learning (PBL) indicate that PBL as a curriculum and instruction approach (Savery, 2006; Schmidt, Loyens, Van Gog, & Paas, 2007) provides an effective way for teachers to respond to students' needs, provides opportunities for students to actively engage in and take responsibility for learning by engaging in meaningful and relevant work, and provides students opportunities to directly apply their knowledge and skills (Hmelo-Silver & DeSimone, 2013; McCombs, 2010; Parker et al., 2011). Although primary research within secondary (6-12) contexts indicated that problem-and project based learning (PBL) is often superior to traditional, lecture-based instruction (Mergendoller, Maxwell, & Bellisimo, 2006; Wirkala & Kuhn, 2011) and meta-analyses at the post-secondary level indicated that PBL is at par with or superior to traditional, lecture-based instruction (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Vernon & Blake, 1993; Walker & Leary, 2009), a synthesized and quantified exploration of the strength of relationship between PBL and academic achievement within middle high school student populations (Grades 6-12) was needed. The results in this meta-analysis indicate that overall, PBL students outperformed traditionally instructed students, g = 0.54, on content and skills exams across academic subject types and grade levels. Analysis of the funnel plot suggests publication bias;

however, an adjustment of the mean effect using Duval and Tweedie's (2000) *Trim and Fill* rendered a similar summary effect of $g = 0.50$. Although the mean summary effect is relatively robust, effect sizes varied depending on subject area and specific types of outcome measures. The test of homogeneity indicated that 90.49% of the variance between studies was unexplained. An insufficient number of studies rendered meta-regression unfeasible, hindering exploration of possible explanations for this variance.

Keywords: meta-analysis, problem-based learning, project-based learning, metacognition, reflective thinking, reflective assessment, academic discussion, constructive discourse, collaboration, adolescence, middle school, junior high school, high school

Chapter 1: Introduction

Background

Increased accountability measures inherent in No Child Left Behind (2003) and Race to the Top (2009) legislation, and current Common Core State Standards (CCSS) (2010) and CCSS initiatives provide impetus for district and school leaders to implement curricular and pedagogical practices that promote raised academic achievement and the preparation of students for democratic participation (CCSS Initiative, 2012; National Governors Association Center for Best Practices [NGACCSSO], 2010; U.S. Department of Education, 2009). Such practices include cooperation and collaboration, problem-solving, problem-posing, considering one's environment, and investigating alternatives (Hmelo-Silver, 2004; Hmelo-Silver & DeSimone, 2013; Kuhn, 2015; NGACCSSO, 2010). Democratic participation includes doing, thinking, and discussing (Katz & Chard, 2000) and through these processes individuals build knowledge and understanding, both personally and socially, about the world around them (Brown, 1977, 1992; Bruner, 1996; Dewey, 1899, 1922, 1933; Flavell, 1976, 1979). Democratic participation also depends upon individuals who are willing to enter into the perspective of others and are able to see themselves from the perspective of others (Banks, 1995, 2008; Bruffee, 1999; Palmer, 1993).

There is a long held belief that knowledge is power and knowledge is necessary to improve society (Bagley, 1939; Banks, 2008; Hirsch, 1996; Ravitch, 2000; Ross & Marker, 2005; Stanley, 2005). Dewey (1937/1991), however, argued that knowledge alone does not lead to understanding or the ability to apply that knowledge. Dewey (1937/1991) advocated teaching methods that connect "knowledge, understanding, and skills" to the "ways things are done socially and how they may be done" (p. 184). He further explained "[f]or only in this connection of knowledge and social action can education generate the understanding of present forces, movements, problems, and needs that is necessary for the continued existence of democracy" (p. 185). Thus, the acquisition of merely knowledge and skills is not enough; teachers must provide opportunities for students to apply knowledge and skills to real world, or authentic, contexts (Barron & Darling-Hammond, 2008) and foster lifelong learning (Dewey, 1933, 1938; Ertmer & Simons, 2006; Kilpatrick, 1921).

Researchers and proponents of problem- and project-based learning (PBL) indicate that PBL as a curriculum and instruction approach (Savery, 2006; Schmidt, Loyens, Van Gog, & Paas, 2007) provides an effective way for teachers to respond to students' needs, provides opportunities for students to actively engage in and take responsibility for learning by engaging in meaningful and relevant work, and provides students opportunities to directly apply their knowledge and skills (Hmelo-Silver & DeSimone, 2013; McCombs, 2010; Parker et al., 2011). PBL encompasses doing, thinking, and discussing (Barron et al., 1998) through active learning, reflective assessment, and academic discussion (Hmelo-Silver, 2004; Hmelo-Silver & DeSimone, 2013; Schmidt et al., 2007). Further, PBL encompasses the desired knowledge sets, skills, and outcomes endorsed by advocates of CCSS (Barron & Darling-Hammond, 2008) and articulated by Partnership for 21st Century Skills (2011): critical thinking and problem-solving, collaboration, communication, and creativity and innovation. PBL also prepares students for lifelong learning through development of self-regulation, inquiry, and metacognition (Ertmer & Simons, 2006; Hmelo-Silver & DeSimone, 2013; Schmidt et al., 2007). However, few studies investigate the effect of PBL on academic achievement in Grades 6-12 populations (Hmelo-Silver & DeSimone, 2013; Ravitz, 2009; Wirkala & Kuhn, 2011) compared to traditional, lecture discussion instruction. Even fewer studies explore the moderating effects of reflective assessment and academic discussion on that achievement (Hmelo-Silver, 2004; Kuhn, 2015; Ravitz, 2009; Wirkala & Kuhn, 2011; Zohar & Ben David, 2008).

Problem- and Project-Based Learning Defined and Differentiated

The Project Method

Problem- and project-based learning (PBL) are descendants of the Project Method. William Kilpatrick (1918, 1921) envisioned an instructional method that might unify knowledge, understanding, skills, and preparation for civic life. Kilpatrick termed this unifying method "the project method" and defined it as "wholehearted purposeful activity proceeding in a social environment" that leads to a "worthy life" (p. 4) and further learning (p. 13). To this end, Kilpatrick conceived of the project method to encompass more than just vocational skills.

Kilpatrick (1921) proposed that three types of projects serve active, educative ends: creating some sort of product (p. 283); solving a problem that requires thinking and clarifying ideas (p. 285); and experiences that drive the acquisition of knowledge and skills (p. 286). In each of these types, the project itself drives inquiry and acquisition of knowledge and skills, subsequent learning, and problem-solving, which in turn are necessary for achieving the success of the project (see Appendix A). Kilpatrick (1918) wrote that these types of projects require "purposing, planning, executing, and judging" (p. 17), skills requisite for decision making and critical thinking in daily life. This method is fundamentally opposite to traditional methods in which a project (or product) follows acquisition of knowledge as a means to demonstrate and assess one's "learning" (Parker et al., 2011; Thomas, 2000) (see Appendix A).

Criticisms of the Project Method

The project method was not without its critics. William Chandler Bagley (1921, 1939) argued that the project method is haphazard, lends itself to instrumental knowledge (only that knowledge needed to complete the project), and questioned whether fixation on project completion might stymie higher order thinking skills and transferability of knowledge. Bagley's (1939) concerns led to his instrumental role in orchestrating the essentialist counter movement in reaction to what he perceived as the ills of progressivism: a de-emphasis in intellectual and academic rigor, and thoughtless planning.

Neither Dewey (1933, 1938) nor Kilpatrick (1918, 1921) advocated haphazard, non-intellectual experiences. Both understood that not all student interests are equally meaningful or educative. Thus, Dewey and Kilpatrick argued it is essential that effective teachers are both content and pedagogical experts in order to guide students to experiences that proffer interest, meaning, and further inquiry and growth. Proponents of PBL attempt to ameliorate the project method by eliminating misconceptions and misapplications of its purpose, implementation, and practice.

Problem- and Project-based Learning Defined

PBL can be described as students, working in small, collaborative groups, confronting "real-world" (Barron & Darling-Hammond, 2008), authentic (Parker et al., 2011), or "ill-structured" (Hmelo-Silver & DeSimone, 2013; Stepien & Gallagher, 1993) problems that are complex and initiate learning and the acquisition of higher-order thinking skills (Barron & Darling-Hammond, 2008; Hmelo-Silver, 2004; Hmelo-Silver & DeSimone, 2013; Wirkala & Kuhn, 2011). Though similar, problem- and project-based learning emerged from different disciplines and methodologies (Graaff & Kolmos, 2007; Hanney & Savin-Baden, 2013). Graaff and Kolmos (2007) explained that problem-based learning developed out of the medical field while project-based learning developed simultaneously out of the engineering field. However, both involve students completing complex tasks.

Some critics of PBL construe the emphasis of student autonomy as “minimally guided” (cf. Kirschner, Sweller, & Clark, 2006) in that teachers simply put students into groups and expect that they will learn by doing a self-directed project or solving-problems. Kirschner, Sweller, and Clark (2006) contended that such an approach is antithetical to the brain’s natural need for structure; especially in novice learners, and delimits students’ ability to retain knowledge and concepts in long term memory. Although some teachers misapply PBL by equating group work with learning, Schmidt, Loyens, Van Gog, and Pass (2007) and Wirkala and Kuhn (2011) insist that good project-based instruction includes structured activities, clear goals and objectives, scaffolding, facilitation and monitoring, discussion, and reflection, each of which is important for supporting content acquisition, conceptual understanding, completion of projects, and skill application (Barron et al., 1998; Hmelo-Silver, 2004; Hmelo-Silver & DeSimone, 2013).

Problem- and project-based learning differentiated. In problem-based learning, the process of solving a problem is the heart of the learning experience. This learning process often requires collaboration and interdisciplinary skills and knowledge, and creativity is prized (Graaff & Kolmos, 2007; Hanney & Savin-Baden, 2013). Therefore the process and "solution" to the problem is the product.

In project-based learning the end result is "a realistic product, event, or presentation to an audience" (Barron & Darling-Hammond, 2008). Larmer (2013) explained project-based learning as a "broad category" that includes designing or creating a product or performance; solving a real world problem, or investing a topic in order to answer an open-ended question. Thus, project-based learning is "defined in terms of the assignment or task" students are expected to complete (Hanney & Savin-Baden, 2013).

Project-led, problem-based learning: A combined approach. In reality the terms problem- and project-based learning are used interchangeably by teachers (Larmer, 2013) and are considered a form of collaborative learning (O'Donnell & Hmelo-Silver, 2013; Hmelo-Silver & DeSimone, 2013). Many projects begin with a problem. Hanney and Savin-Baden (2013) argued for a hybrid model: project-led problem-based learning. In this model, process is valued over product and emphasis is placed on the problem-solving, creativity, and collaboration aspects of creating a product. It is these skills, Hanney and Savin-Baden explained that reinforce the learning objectives and stimulate learning itself.

For the purpose of this study, problem- and project-based learning (PBL) are combined and discussed as one because they are closely related and promote the same ends: collaborative learning that promotes self-directed learning and deep understanding. Several advocates of PBL recommend the use of problem-based learning as a scaffold for project-based learning, because problem-based learning provides students the opportunity to practice defining a problem, exploring alternatives, sharing ideas, proposing and presenting solutions, and continual reflection before taking on "open-ended projects" (Barron et al., 1998: Hanney & Savin-Baden, 2013).

PBL, Reflective Assessment, Academic Discussion, and "Traditional" Instruction

Researchers in PBL indicate that academic achievement is improved when students construct their knowledge through contextual learning, particularly when PBL is augmented with reflective assessment and academic discussion (Barron et al., 1998; Hmelo-Silver, 2004; Hmelo-Silver & DeSimone, 2013; Kuhn, 2015; Parker et al., 2011; Wirkala & Kuhn, 2011; Zohar & Ben David, 2008). PBL, a student-centered approach, is often compared to traditional, lecture-discussion instructions, a teacher-centered approach. Reflective assessment, academic discussion, and traditional instruction are each defined to provide context for further discussion in subsequent chapters.

Reflective Assessment Defined

Barron et al. (1998) contend that active reflection is essential in PBL, otherwise students get caught up in the doing, but fail to make meaning of it (the learning) (cf. Boud, Keogh, & Walker, 1985; Hmelo-Silver & DeSimone, 2013). Dewey (1933) defined *reflective thinking* as the "active, persistent, and careful consideration of any belief or supposed form of knowledge" (p. 9; cf. Johnson & Johnson, 1992; Kohlberg, 1976; von Glasersfeld, 1989, 2001/2006). The term reflective thinking, however, is subject to ambiguity. It is also referred to as metacognition (Brown, 1977; Flavell, 1976), consciousness (Vygotsky, 1962), self-regulated learning (Wolters, 2010; Zimmerman & Schunk, 2001), reflective assessment (Ellis, 2001; White & Frederiksen, 1998), and simply, reflection (Boud et al., 1985; Ertmer & Simons, 2006; Hmelo-Silver &

DeSimone, 2013). The term *reflective assessment* includes a variety of formative assessment techniques that provide students the opportunity to think actively about, articulate, and evaluate what they are learning (Ellis, 2001; Ellis, Bond, & Denton, 2012; White & Frederiksen, 1998). Thus, reflective assessment (RA) fosters reflective thinking.

Academic Discussion Defined

Academic discussion (AD) (Elizabeth, Ross Anderson, Snow, & Selman, 2012) is another essential feature of PBL. *Academic discussion* refers to the structured exchange of ideas, knowledge, and/or feedback for the purpose of better understanding academic material, concepts, or ideas (Barron et al., 1998; Hmelo-Silver, 2004; Wirkala & Kuhn, 2011). The word academic is intentional in that it indicates a trained or disciplined form of discussion. In this way, discussion leads to a purposeful end (Elizabeth et al., 2012). Academic discussion is often described as a form of social cognition, or reflection (Boud et al., 1985; Kuhn, 2015; Hmelo-Silver & DeSimone, 2013; Schmidt et al., 2007). This verbal exchange of intellectual activity is also referred to by other names: collaboration (Bruffee, 1999; Bruner, 1985; Hmelo-Silver & DeSimone, 2013; Worsham, 1992; Yager, Johnson, & Johnson, 1985); oral discussion (Johnson & Johnson, 1992; Yager et al., 1985); oral interaction (Forman & Cazden, 1985; Johnson, Johnson, Roy, & Zaidman, 1985; Webb, 1982; Webb & Kenderiski, 1984); constructive conversation (Bruffee, 1999), and classroom discourse (Anderson, Zuiker, Taasoobshirazi, & Hickey, 2007).

Traditional, Lecture-based Instruction Defined

The effects of PBL are typically compared to traditional, lecture-based instruction. Ravitz (2009) suggested that there is perhaps a “false dichotomy” created between PBL and traditional instruction, because “it is doubtful that either traditional instruction or PBL exists in ‘pure’ forms” (p. 6). Traditional instruction is referred to by different names: lecture-based instruction (Visser, 2003), lecture-discussion (Wirkala & Kuhn, 2011), expository method (Anyafulude, 2013; van Loggerenberg-Hattingh, 2003), and explicit teaching or direct instruction (Rosenshine, 1987), and direct-interactive teaching (Chang, 2001). Similar to the variation in PBL definition and practice, depending on the academic discipline in which it is applied or the form of PBL implemented (Walker & Leary, 2009), the definition of traditional instruction varies (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Ravitz, 2009).

Traditional instruction is commonly defined or described as whole group instruction, where lectures by the teacher and discussion between and among teacher and students characterize the instructional method (Dochy et al., 2003; van Loggerenberg-Hattingh, 2003). Discussion is typically teacher directed and student talk is relegated to responding to questions with pre-specified, answers with “brief phrases or in single disconnected sentences” (Dewey, 1933, p. 245), or it is simply used to “‘test’ what students already know” (Applebee, Langer, Nystrand, & Gamoran, 2003, p. 690). Learning is typically individualistic and the prominent emphasis is transmission of information from the teacher to the students. This approach is often criticized as a passive approach to learning (van Loggerenberg-Hattingh, 2003). Rosenshine (1987), however, described direct instruction as a “systematic method of teaching with emphasis on proceeding in small steps, checking for student understanding, and achieving active and successful participation by all students” (p. 34). Whereas some forms of PBL omit any forms of direct, whole group instruction, contemporary iterations encourage mini-lectures as needed to scaffold learning at "teachable moments" (Mergendoller, Maxwell, & Bellisimo, 2006). The difference between the use of lecture-discussion as the primary mode of teaching and learning versus the use of mini-lectures in PBL is that the latter is a supplemental approach based on student need. That is, in PBL student-directed and generated questions, inquiry, and problem-solving are primary and teacher-directed instruction is secondary and minimal, and used solely to support student directed needs.

Purpose of the Study

The majority of studies related to PBL are from the medical profession and higher education (Hmelo-Silver, 2004; Walker & Leary, 2009). Few empirical studies measure the effectiveness of PBL in K-12 education (Hmelo-Silver, 2004; Wirkala & Kuhn, 2011) and even fewer in middle (Wirkala & Kuhn, 2011) and high school contexts (Finkelstein., Hanson, Huang, Hirschman, & Huang, 2011; Mergendoller et al., 2006; Sungar, Tekkaya, & Geban, 2011). Of the PBL et al research in Grades 6-12, the majority of these studies are conducted in math and science contexts (Mergendoller et al., 2006; Wirkala & Kuhn, 2011). Few studies in PBL in Grades 6-12 are conducted in the social sciences (Finkelstein et al., 2010; Wirkala & Kuhn, 2011) and English/Language Arts. Further, there is a lack of research that specifically measures the effects of reflection and discussion on PBL outcome achievement; although many research designs include these elements (Hmelo-Silver, 2004; Mergendoller et al., 2006; Parker et al., 2006; White & Frederiksen, 1998; Wirkala & Kuhn, 2011; Zohar & Ben David, 2008).

Although primary research within secondary (6-12) contexts indicates that PBL is often superior to traditional, lecture-based instruction (Mergendoller et al, 2006; Wirkala & Kuhn, 2011) and meta-analyses at the post-secondary level indicate that PBL is at par with or superior to traditional, lecture-based instruction (Dochy et al., 2003; Vernon & Blake, 1993; Walker & Leary, 2009), a synthesized and quantified exploration of the strength of relationship between PBL and academic achievement within middle school, junior high, and high school student populations (Grades 6-12) is needed. Further, this exploration includes the "conditions and practices associated with differences in effectiveness" (U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, 2010, p. 2), which will provide context for implementation considerations. The U.S. Department of Education, Office of Planning, Evaluation, and Policy Development (2010) in meta-analysis exploring the effect of online learning describes "conditions and practices" in the following way. *Conditions* refer to the design features of a study, such as school, teacher, and student demographics, year of publication, and "state accountability systems" (p. 2). *Practices* refer to implementation of an intervention, e.g. role of the facilitator/teacher or how reflection and discussion are used—e.g. to complete a task or to promote understanding and application of concepts and skills.

Significance of Study

In an age of accountability, teachers and administrators are less inclined to implement instructional approaches that may negatively impact adequate yearly progress (AYP) mandated in NCLB (2003) legislation (Ertmer & Simon, 2006; Grant & Hill, 2006). As such, high stakes testing that is associated with AYP typically undermines pedagogical approaches that deter teachers from using "skill and drill" or lecture/discussion instructional methods as a means to prepare students for these high stakes tests (Grant & Hill, 2006; Ravitch, 2010). Compounding the issue are newly adopted teacher evaluation systems such as the Washington Teacher/Principal Evaluation Project (TPEP), curriculum adoptions that support CCSS, and continually changing high stakes testing mediums, such as the newly implemented Smarter Balanced Assessment. Within the TPEP (2013) model, teachers are held accountable to develop student (whole- and subgroup) growth goals (Criterions 3.1 and 6.1), demonstrating that students are academically achieving (Criterions 3.2 and 6.2). The compilation of these demands limits teachers' time and willingness to implement student-centered approaches, such as PBL, that may not prove effective. Teachers and administrators need cohesive information about the effects of PBL on academic achievement to warrant the time, energy, and resources required to change current practices, and subsequently implement and sustain ongoing professional development and curriculum for PBL

Research Questions and Null Hypotheses

Research Questions

As noted, meta-analyses on the effects of PBL in post-secondary education indicate that PBL is at par with or superior to traditional instruction, but there is wide variability in effect size among PBL studies, especially in the sciences and math (Vernon & Blake, 1993; Walker & Leary, 2009). In regard to assessment types, traditionally instructed students tend to outperform PBL students on knowledge acquisition (declarative knowledge) tests, but PBL students outperform traditionally instructed students on application of knowledge (skills/procedural knowledge) tests (Dochy et al., 2003; Vernon & Blake, 1993; Walker & Leary, 2009). Thus, the investigator explored and analyzed the effects of PBL on academic achievement in Grades 6-12 populations with several questions in mind:

1. What is the overall effect of PBL on Grades 6-12 populations across subject areas and grade levels compared to traditional instruction?
2. What effect does PBL have on individual academic achievement among Grades 6-12 populations compared to traditional instruction (lecture-discussion based) as measured by immediate content and/or skills posttests?
3. What effect does PBL have on individual academic achievement among Grades 6-12 populations compared to traditional instruction (lecture-discussion based) as measured by content and/or skills retention tests?
4. To what extent is this mean effect, if any, moderated by a.) academic subject b.) grade level ; c.) location (country); and/or d.) ability level of students (i.e. low, medium, or high ability students)?
5. To what extent is this mean effect, if any, moderated by a.) the use of reflective assessment and/or b.) the use of academic discussion?

After a careful review of the literature and coding of studies an additional question was added:

6. To what extent, if any, is the effect of PBL moderated by facilitator type (researcher or teacher) and prior PBL training (brief or extensive) or experience?

Null Hypotheses

There are two null hypotheses investigated in this study using meta-analytic procedures:

1. There are statistically non-significant differences in individual academic achievement among PBL and traditionally instructed conditions in Grades 6-12 populations as measured by immediate content and/or skills posttests.
2. There are statistically non-significant differences in individual academic achievement among PBL and traditionally instructed conditions in Grades 6-12 populations as measured by content and/or skills retention tests.

For both null hypotheses the independent variable is PBL (the intervention) and the dependent variable is academic achievement assessed through teacher developed, researcher developed, or standardized post- and/or retention tests.

Content of the Following Chapters

The subsequent sections of this dissertation are divided into four chapters, titled: Literature Review, Research Methods, Results, and Discussion of Results. The Literature Review includes an extensive overview of the theoretical framework and empirical research that supports the efficacy of PBL, reflective assessment, and academic discussion on academic achievement. The Research Methods chapter outlines the research design, inclusion and exclusion criteria for study selection, methodology, and data analysis used to conduct this meta-analysis. In the Results chapter, the investigator reports descriptive and meta-analytic results related to literature search and the stated hypotheses. The final chapter, Discussion of Results, contains the discussion of the results in light of the hypotheses, primary and moderator variables, former empirical studies, and theoretical assumptions. Suggestions for further research are included.

Chapter 2: Literature Review

Advocates of problem- and project based learning (PBL) tout it as a viable, learner-centered approach, among many, to promote higher order thinking skills, autonomy, collaborative learning, belonging, and self-regulation in students. PBL is not a conglomeration of teaching techniques, or rigid activities and strategies (Katz & Chard, 2000). Rather, as a curriculum and instruction approach (Savery, 2006; Schmidt et al., 2007), PBL provides a way for teachers to respond to students' needs so that students can actively participate in and take responsibility for learning by engaging in meaningful and relevant work (Katz & Chard, 2000; McCombs, 2010). Theory and research related to PBL provide a framework for analyzing the efficacy of PBL in developing cognitive and psychological skills that promote lifelong learning.

Theoretical Framework

There is no singular or unifying theoretical framework for problem- and project-based learning (PBL) (Finkelstein et al., 2011; Wirkala & Kuhn, 2011). However, doing, thinking, and discussing are fundamental components (Katz & Chard, 2000). Dewey (1933, 1938) argued that experiential learning, reflective thinking, and discussing are functions of active experience (or doing), and through these processes students are afforded opportunities to find meaning, which leads to further investigation, reflection, and discussion. Learning theories relevant to PBL, reflective thinking, and academic discussion are constructivism, meta-cognition and co-cognition, and social cognitive theory (SCT) (Brooks & Brooks, 1993; Hmelo-Silver, 2004; Wirkala & Kuhn, 2011). Each theory is discussed independently and connections are made between the theory and salient features of PBL. Although constructivism includes reflective thinking and academic discussion (von Glasersfeld, 1989, 2001/2006), which are essential and often embedded elements of PBL, theories related specifically to reflective thinking and discussion are discussed separately.

Educational Constructivism: Individual and Collaborative Active Learning

Constructivism is a theory of learning that addresses *how* knowledge is constructed by individuals and negotiated within social contexts (Brooks & Brooks, 1993; Phillips, 1995). Zimmerman (2001) stated that "learning is not something that happens *to* students; it is something that happens *by* students" (p. 33, original emphasis). However, there is no single theory of constructivism (Phillips, 1995). Phillips (1995) noted that there a several "sects" of constructivism, some of which have implications for education. The categorizations defined by Phillips are used here to discuss constructivism as it relates to PBL: *psychological*, *social*, and *radical.* Despite epistemological differences, the fact that theories of educational constructivism address how individuals learn directly impacts instructional practices and therefore influences teaching methods.

Psychological constructivism. Psychological constructivism is grounded in the work of Piaget (Applefield, Huber, & Moallem, 2001). Piaget (1967) held that each learner constructs his/her knowledge internally as a result of a problem to solve (disequilibrium) or a specific curiosity to discover by accommodating new experiences (information) to existing schemas of knowledge. In this way, when students are presented with a problem-statement or solution to solve in PBL, they are confronted with new information that creates disequilibrium, which in turn stimulates inquiry to solve the problem. As students are confronted with new information they must then process that new information into existing schema; thus, creating new understandings (O'Donnell & Hmelo-Silver, 2013).

O'Donnell and Hmelo-Silver (2013) further explained that Piaget's assertions regarding peer-to-peer interactions have important implications for collaborative learning methods. Namely, Piaget (1967, 1977) asserted that adult-to-child relationships create situations in which the child (student) is more likely to comply with the adult's (teacher's) thinking and ways of doing. However, in peer-to-peer interactions, students are "more likely to develop cognitively in contexts in which peers have equal power and all have opportunities to influence one another" (O'Donnell & Hmelo-Silver, 2013, p. 8).

Social constructivism. Dewey (1938), Vygotsky (1962, 1978), and Bruner (1960, 1985, 1996) each addressed the ideas that students learn more productively by working together than individually and that language is the instrument through which socially developed knowledge, norms, beliefs, and rules are transmitted. Thus, each argued that knowledge construction begins with and proceeds through social interaction with others. Vygotsky (1962) wrote, "[t]he relation between thought and word is not a thing but a process, a continual movement back and forth from thought to word and from word to thought.…Thought is not merely expressed in words; it comes into existence through them" (p. 126). Therefore, language [both verbal and non-verbal], and, thus, interaction, is a natural and mandatory requisite for thought and knowledge construction (Bruner, 1996; Dewey, 1933; Vygotsky, 1962). The key function of language, then, is to derive meaning, resolve problems, and reciprocally transfer knowledge and ideas to others.

Collaborative interaction within the classroom takes the form of teacher-to-student and peer-to-peer relationships. Vygotsky (1962, 1978) theorized that when a learner can complete a task successfully through scaffolded help by a teacher or more capable peer, that individual is working within his/her *zone of proximal development*. When applied effectively, scaffolding can be slowly reduced until the learner can succeed on his/her own at the new level of learning. Dewey (1938) and Bruner (1960) advocated facilitated, concrete experiences in which the learner is an active participant in the learning experience, and an active observer of how the teacher or mentor interacts with the learning experience. The social construction of learning, then, is a collaboration of all individuals in the classroom: the teacher as facilitator and guide (Bruffee, 1999; Bruner, 1996; Dewey, 1938), and the students as co-contributors (Dewey, 1938; Wolters, 2010; Zimmerman, 2001), inquirers (Bruner, 1966), problem-solvers (Resnick & Glaser, 1976), and problem-posers (Costa & O'Leary, 1992).

Radical constructivism. Von Glasersfeld (1989) argued that each individual constructs knowledge based on conceptions of perceived reality. Von Glasersfeld maintained that teachers cannot assume that all learners have the same conceptions, or that the words they use produce the same mental models for all students (p. 134). Further, not all students will construct knowledge the way the teacher intends, despite that teacher's best efforts. Therefore, von Glasersfeld (1989, 2001/2006) emphasized the necessity for reflective practices in the classroom coupled with opportunities for students to "discuss their view of a problem and their own tentative approaches [to solve that problem]" (p. 5). In this way students have opportunities to think about and discuss their learning to make deeper, more informed, and (hopefully) more accurate understandings.

Salient features of PBL and educational constructivism. PBL and constructivist practices include encouraging cooperation and collaboration, problem-solving, problem-posing, considering one's environment, and investigating alternatives (Barron et al., 1998; Larmer & Mergendoller, 2012a, 2012b; McCombs, 2010). Further, PBL methods encourage knowledge construction because students communicate their ideas with others and actively participate in the learning experience.

William Kilpatrick envisioned the role of the teacher as one who facilitates learning in such a way that the project develops self-regulated skills (Kirkpatrick, 1918) and is educative (Kirkpatrick,1921). The teacher is an active facilitator of the learning process by asking probing questions, providing feedback, and encouraging students to "dig deeper" or utilize each other and outside sources as resources (Zohar & Ben David, 2008). In this way, Kilpatrick (1918) argued that if a teacher is successful in facilitating the learning process, the teacher should "gradually eliminate" him or herself from the "success of the procedure" (p. 13). Contemporary advocates of PBL (cf. Ertmer & Simons, 2006; Hmelo-Silver & DeSimone, 2013; Schmidt et al., 2007; Wirkala & Kuhn, 2011) emphasize the necessity of scaffolding inquiry, collaboration, discussion, and reflection skills so that students acquire and practice these skills in order to be successful in a PBL learning environment. Wirkala and Kuhn (2011) explained that good PBL includes structured activities and interaction among students and teacher so that students reach their optimal learning level.

Metacognition and Co-Cognition: The Thinking and Discussing Paradigm

Reflective assessment reiterated. Dewey (1933) defined reflective thinking as the "active, persistent, and careful consideration of any belief or supposed form of knowledge" (p. 9; cf. Johnson & Johnson, 1992; Kohlberg, 1976; von Glasersfeld, 1989, 2001/2006). The term *reflective thinking*, however, is subject to ambiguity. It is also referred to as metacognition (Brown, 1977; Flavell, 1976), consciousness (Vygotsky, 1962), self-regulated learning (Wolters, 2010; Zimmerman & Schunk, 2001) and reflective assessment (Ellis, 2001; White & Frederiksen, 1998). The term *reflective assessment* includes a variety of formative assessment techniques that allow students to think actively about, articulate, and evaluate what they are learning (Ellis, 2001; Ellis et al., 2012; White & Frederiksen, 1998). Thus, reflective assessment fosters reflective thinking.

Metacognition defined. Flavell (1976) coined the term metacognition to describe an individual's ability to take what one has learned (i.e., a problem-solving skill), organize it, and integrate that knowledge into practice. As Flavell explained, metacognition "refers, among other things, to the active monitoring and consequent regulations and orchestration of these processes" (p. 232). Flavell (1987) later extended the scope of the term to "include anything psychological" and suggested that metacognition might also be attributed to processes of self-regulation that "are not conscious and perhaps not even accessible to consciousness" (p. 21).

Other researchers have also suggested that metacognition is not always conscious, especially if a practice has become self-regulatory (Bandura, 1986; Veenman, Van Hout-Walters, & Afflerbach, 2006). In this way, metacognitive strategies, such as "planning, monitoring, and evaluating" (Schraw, 1998, p. 114) must be explicitly taught and practiced in meaningful contexts in order for students to conceptualize the significance of these strategies on their learning (Brown, 1992; Ellis et al., 2012; White & Frederiksen, 1998; Zohar & Ben David, 2008). Once metacognitive strategies become innate skills, an individual can in turn utilize these strategies without having to consciously think about them (Barron et al., 1998; Schraw 1998; Veenman et al., 2006) and can subsequently utilize them in different contexts (Black & William, 2009).

Metacognition as a personal construct. Dinsmore, Alexander, and Loughlin (2008) wrote that "[a]t the broader level, the foundation of metacognition is in the mind of the individual" (p. 393). Thus, knowledge is personally constructed as the learner is confronted with new information, evaluates his/her thoughts and ideas, and attempts to make meaning of it. Brown (1977, 1987) suggested that when students are explicitly taught monitoring strategies, there is a positive effect on retention and transference, regardless of initial ability and one's previous beliefs about one's abilities.

The personal construct is elicited through reflective assessment (formative assessment) practices (Black & William, 2009). Arends and Kilcher (2010) emphasized the importance of "teaching students how to learn, [because] [p]roviding students with metacognitive skills helps them become aware of their own cognitive processes so they can monitor their progress and take responsibility for their own learning" (p. 54). Black and William (2009) explained the role of metacognition in augmenting a student's cognitive and psychological development:

> [M]etacognition is regarded as a higher level psychological process. By challenging learners to reflect on their own thinking, teachers and their peers help them to make unconscious processes overt and explicit and so making the more available for future use." (p. 19)

The role of reflective assessment, however, is not entirely individual. Black and William further asserted that the individual aspect of reflection is augmented by social interaction with the teacher and peers.

Academic discussion reiterated. *Academic discussion* refers to the structured exchange of ideas, knowledge, and/or feedback for the purpose of better understanding academic material or concepts (Elizabeth et al., 2012). The word *academic* is intentional in that it indicates a trained or disciplined form of discussion. In this way, discussion leads to a purposeful end (Elizabeth et al., 2012). This verbal exchange of intellectual activity is also referred to by other names: collaboration (Bruner, 1985; Worsham, 1992; Yager et al., 1985); oral interaction (Forman & Cazden, 1985; Johnson et al., 1985; Webb, 1982); constructive conversation (Bruffee, 1999), and classroom discourse (Anderson et al., 2007), and constructive collaboration (Kuhn, 2015).

Co-cognition defined. While metacognition is largely a personal construct, others suggest that metacognition can be socially stimulated in the form of co-cognition (Costa & O'Leary, 1992; Worsham, 1992), or social cognition (Brown, 1977). Brown (1977) explained that "social cognition, role taking, and communication" (p. 6) become relevant areas of research and consideration, as metacognition requires that an individual not only judge one's own capabilities, thinking, or ideas, but judge these against the perspective of others. Costa and O'Leary (1992) wrote that "co-cognition is…collaboratively developing concepts, visions, and operational definitions of intelligent behavior, which in turn are used to guide, reflect upon, and evaluate one's own performance while in groups (co-cognition) or when alone (metacognition)" (p. 53).

Co-cognition as a social construct. This reciprocal interaction of thinking and speaking, and speaking and thinking within a social setting develops meaningfulness and cognitive development (Vygotsky, 1962). Kuhn (2015) argued that there is distinct difference between transfer of knowledge and genuine collaboration. Mere transfer of knowledge occurs when a more competent member, which Kuhn stated exists in most groups, simply explains what he/she thinks or knows and the others in the group take that knowledge but contribute nothing intellectually in return. In contrast, genuine collaborations are those in which "participants directly engage one another's thinking. They listen and respond to what their peers say. In less successful collaborations, participants are more likely to work in parallel and ignore or dismiss the other person's contributions" (Kuhn, 2015, p. 47; cf. Black & William, 2009).

Whether to reflect on one's performance within the group or to engage intellectually about the task or problem at hand, co-cognition provides a social mechanism for extending one's thinking. Hmelo-Silver and DeSimone (2013) and Kuhn (2015) advocate explicit teaching and modeling of constructive collaboration (academic discussion) and sustained practice. As Kuhn explained, "Intellectual collaboration does not come naturally…it is not enough to put individuals in a context that allows for collaboration and expect them to engage in it effectively" (p. 51; cf. Johnson & Johnson, 1992).

Social Cognitive Theory: Self-Efficacy and Regulation of Learning

Bandura's *social cognitive theory* (SCT) (1986; formerly *social learning theory*, cf. Bandura, 1977) extends the theories of Dewey and Vygotsky in that Bandura more clearly articulates the free-agency (or "self-regulatory") aspect of humans. Human agency, or what Bandura (2001) called the agentic individual, describes the ability of an individual to take control of the quality of his/her experiences through contemplative processes. These processes are described as forethought, monitoring, regulation, and self-reflection (Bandura, 2001; Schunk, 2001; Wolters, 2010). This agentic, self-regulatory aspect of Bandura's SCT is significant because it supports the notion that *reflective assessment*, a self-regulatory, metacognitive activity, can influence academic achievement, because humans have the ability to change their current status through reflection and action. Though individuals may work more productively when working with others, and learn from one another through interaction and collaboration, one ultimately controls what or whether one learns (Bandura, 2001; Dewey, 1933, 1938; Flavell, 1979; Schunk, 1999, 2001).

Bandura (2001) stated that SCT differs from Vygotsky's sociocultural theory (social constructivism theory) in that a central tenant of SCT is self-efficacy. Regardless of social influences, an individual must believe him/herself capable. However, one must contend that the social environment (the classroom) often has tremendous influence upon a student's beliefs. As Bandura explained, "The likelihood that people will act on the outcomes they expect prospective performances to produce depends on their beliefs about whether or not they can produce those performances" (p. 10). One of the goals of PBL is to produce self-efficacious and self-directed learners (Hmelo-Silver & DeSimone, 2013). Bandura (1986, 2001) argued, as do advocates of PBL (see Ertmer & Simons, 2006; Hmelo-Silver & DeSimone, 2013; Schmidt et al., 2007) that the environment a teacher creates directly influences a student's ability to develop and foster self-efficacy and self-direction: important skills for success PBL outcomes and lifelong learning. A teacher creates an environment of potential success when he/she provides students with authentic problems to solve, and scaffolds and models effective inquiry, collaboration (to complete a task *and* discuss ideas academically), and reflection skills (Ertmer & Simon, 2006; Hmelo-Silver, 2004; Hmelo-Silver & DeSimone, 2013; Kuhn, 2015).

Internalization: Reciprocal Roles of Personal and Social Constructs

Taken together, theorists in educational constructivism, metacognition and co-cognition, and SCT suggest that learning is both a personal and social co-construction. These theories and PBL share three salient features: Engagement in authentic, real world experiences, personal meaning making, and social interaction. Appendix B includes a diagram of the reciprocal interaction of personal and social constructs, mediated by active learning (doing).

A**uthentic, challenging, and real world learning experiences.** Thinking and discussing are stimulated by active engagement with a challenging problem to solve or project to complete (Barron et al., 1998; Dewey, 1933, 1938; Hmelo-Silver, 2004; Kilpatrick, 1918; Kuhn, 2015; Wirkala & Kuhn, 2011). Barron and Darling-Hammond (2008) argued that authenticity in PBL includes realistic culminating projects, driving questions, or real-world issues that require solutions that have the potential to be implemented. However, some, but not all, advocates of PBL suggest going into the places that the problem/issue takes place, in order to conquer challenges, learn flexibility, and to think creatively within a natural context (Brooks & Brooks, 1999). The primary emphasis, whether in a classroom or natural setting is emphasis on concrete rather than theoretical experience (Dewey, 1938; McCombs, 2010).

Personal meaning making. Reflective assessment strategies elicit initial thinking and ultimately inform the learner in his/her discussions and further inquiry (Ellis, 2001; Clark, 2012). Knowledge is personally constructed as the learner is confronted with new information, evaluates his/her thoughts and ideas, and attempts to make meaning of it (Brooks & Brooks, 1993; Bruner, 1996; Piaget, 1967, 1977). Further, meaning making and active engagement in a problem or project is dependent on tapping into students' prior knowledge and suppositions (Brooks & Brooks, 1993; Ellis et al., 2012; Hmelo-Silver & DeSimone, 2013; Kuhn, 2015; Schmidt et al., 2007). Barron et al. (1998) contend that active reflection is necessary to bring meaning to active learning experiences; otherwise, students get caught up in the "doing," but do not make connections with the learning goals or develop deep understanding (p. 274). Thus, reflection affords students the opportunity to actively think about, articulate, and evaluate what they are learning (Ellis, 2001; Ellis et al., 2012; White & Frederiksen, 1998).

Social interaction. Students need opportunities to discuss their ideas as well as the ideas of others (Banks, 1995; Brooks & Brooks, 1993; Bruffee, 1999; Palmer, 1993). Discussion with others acts as a mediator and monitor of one's thoughts, ideas, and perceived understandings or knowledge (Kuhn, 2015). This interaction further stimulates reflective thinking (Applebee, et al., 2003; Bruffee, 1999; Bruner, 1996; Nystrand, Wu, Gamoran, Zeiser, & Long, 2003). The salient features of PBL related to learning and cognitive theory are highlighted in Table 1.

Table 1

Salient Features of Problem- and Project-based Theory Related to Theory

	Salient Features of Problem and Project-based Learning
Theory	Description
Educational Constructivism	• Promote active learning and cognitive growth through challenging and authentic problems to solve or projects to complete; • Scaffold (ZPD) inquiry, collaboration, and reflection skills; • Incorporate reflective thinking to promote meaning making; • Facilitate discussion to elaborate thinking;
Metacognition	• Model and foster reflective thinking as a means for students to connect doing to learning (connection to objectives); • Utilize individual reflective assessment strategies to solicit monitoring of task completion and understanding of what is being learned.
Co-cognition	• Model and foster academic discussion as a means for students to share ideas, judge those ideas, and monitor and adjust own ideas against others' ideas/beliefs; • Utilize social reflective assessment strategies (task and learning objectives).
SCT (Social Cognitive Theory)	• Develop self-regulation and self-efficacy skills so that the learner has the confidence to self-direct his/her learning; • Create a learning environment that promotes success by scaffolding essential skills: inquiry, collaboration, and reflection.

Review of Research: Effects of PBL on Academic Achievement

The majority of studies related to PBL are from the medical profession and higher (post-secondary) education (Hmelo-Silver, 2004; Walker & Leary, 2009). Few empirical studies measure the effectiveness of PBL in K-12 education (Hmelo-Silver, 2004; Wirkala & Kuhn, 2011) and even fewer in middle (Wirkala & Kuhn, 2011) and high school contexts (Finkelstein et al., 2010; Mergendoller et al., 2006; Sungar et al., 2011). Of the PBL research in Grades 6-12, the majority of these studies are conducted in math and science contexts (Mergendoller et al., 2006; Wirkala & Kuhn, 2011). Few studies in PBL in Grades 6-12 are conducted in the social sciences (Finkelstein et al., 2011; Wirkala & Kuhn, 2011).

There is a fundamental problem with reviewing literature from higher education, medical fields in particular, on PBL effectiveness in relation to secondary (Grades 6-12) students. One issue is the difference between the cognitive development and experiences of the average 6-12th grader and that of a college student. Further, college students typically demonstrate sufficient academic achievement and learning process capabilities to attend and succeed in college. For many middle/junior high and high school students, these traits are still developing. In regard to research in PBL within medical disciplines, medical students and the average middle/junior high and high school student are incomparable motivationally or academically (Mergendoller et al., 2006). Further medical school students are typically high achievers and undergo a competitive selection process. In secondary classrooms, on the other hand, there is wide variation in achievement and ability. Despite the incomparability, the breadth of literature from these fields provides a baseline of potential PBL effectiveness in Grades 6-12 settings and a basis for comparison; therefore, select meta-analyses of the effect of PBL on academic achievement in post-secondary contexts are discussed. The majority of the research presented and discussed in this review, however, was conducted with Grades 6-12 participants. These studies provide an overview of the effect of PBL on academic achievement in secondary contexts.

Effects of PBL on Academic Achievement within Post-Secondary Education

In the three meta-analyses discussed in this section (listed in order of review: Vernon & Blake, 1993; Dochy et al., 2003; and Walker & Leary, 2009), a positive (+) sign in front of the effect size (ES) indicates that the PBL students outperformed traditional, lecture/discussion instructed (LD) students and a negative (-) sign indicates that LD students outperformed PBL students. It must be noted that the meta-analysis by Walker and Leary (2009) contained studies that were conducted in K-12 settings; however, the preponderance of studies in their meta-analysis were conducted in higher education.

Meta-analyses within health and medical-related disciplines. Vernon and Blake (1993) conducted a meta-analysis of 22 research studies within health-related disciplines that compared PBL programs to traditional, LD programs. The authors examined the effect of PBL on four outcome indictors: *program evaluation* (i.e., student and teacher attitudes toward PBL v. LD programs; attendance, etc.); *academic process* (use of resources; approaches to learning—self- or instruct-facilitated, and information-seeking skills); *clinical functioning* (performance-based assessments: use of clinical (contextual) knowledge; clinical reasoning; and 'independent study of clinical problems"); and *academic achievement* (performance on standardized, e.g. National Board of Medical Examiners, Part I (NBME-I), and other knowledge tests) (pp. 554, 556, 560). Overall, Vernon and Blake stated that "the results of our meta-analyses support the superiority of the PBL approach over more traditional methods on several of the outcome domains examined" (p. 557). This conclusion was based on three outcome indicators: program evaluation (d_w = +0.55), academic process (ranging from d_w = +0.32 to 0.79, depending on the skill measured), and clinical reasoning (d_w = +0.28). A more extensive discussion of the effect of PBL on academic achievement in this study is needed.

In regard to academic achievement, PBL students did not outperform LD students (d_w = -0.18). However, Vernon and Blake (1993) noted that the “advantage” (p. 560) of LD approaches on academic achievement, based on results of the meta-analysis, might be contributed to significant variation among ES based on location of program. Two institutions, New Mexico and Michigan State, varied greatly in reported outcomes: New Mexico studies indicating “consistently negative and the latter [Michigan State] had values that were consistently positive” (p. 556). Thus, Vernon and Blake suggested that how a program (teacher) implements PBL may have an impact on what is emphasized. Most iterations of PBL emphasize skills (process) over knowledge (content) (Vernon & Blake, 1993; cf. Barron et al., 1998; Hmelo-Silver & DeSimone, 2013; Wirkala & Kuhn, 2011). However, this emphasis on skills does not suggest that content is unimportant; rather, content is considered secondary to practical application and long term use of content (Wirkala & Kuhn, 2011).

Dochy, Segers, Van den Bossche, and Gijbels (2003) conducted a meta-analysis of 43 quasi-experimental studies comparing PBL to LD approaches in medical programs. The primary analysis of their study focused on the main effect of PBL on knowledge (content acquisition) and skills (application of knowledge). The authors also investigated four moderating variables: methodological factors of the study (design and scope), expertise level of participants (year of program), type of assessment methods used, and effect of a retention outcome component on knowledge acquisition. Similar to the Vernon and Blake (1993) meta-analysis, results indicated that for knowledge acquisition

LD students outperformed PBL students (d_w = - 0.22) and for application of knowledge (skills) PBL students outperformed LD students (d_w = +0.46). Results of moderating variables on knowledge acquisition help explain the heterogeneity among effect sizes and the ensuing negative effect for PBL students. A few of these moderators are highlighted.

When accounting for level of expertise (year of program) on knowledge acquisition, there is a negative trend for PBL students in the first two years of the program (Year 1: d_w = - 0.15; Year 2: d_w = - 0.32). Dochy et al. (2003) explained that programmatic difference are noteworthy; namely in traditional programs students receive "a two-year basic science segment composed of formal courses drawn from various basic disciplines;" whereas PBL students are required to begin applying knowledge immediately (p. 542). After the first two years, the traditional and PBL programs look more similar as LD students begin to apply their knowledge. Interestingly, in Year 3 of medical programs, effect sizes of knowledge acquisition tests favor PBL students (d_w = +0.39), there is nearly negligible difference between PBL and LD students in Year 5 (d_w = - 0.04), and by graduation PBL students slightly outperform LD students (d_w = +0.17). Results of retention as a moderator on knowledge acquisition (content) achievement indicate that while PBL students may acquire "less" content as measured by immediate, posttest knowledge exams, PBL students retain more of the acquired content over time (d_w = +0.14).

Assessment type (instrumentation) also appears to have a significant effect on knowledge acquisition and skill application outcomes. The more accurately an instrument measures application of skills (i.e. procedural knowledge) the more favorable results are for PBL students (d_w ranges from +0.08 to +0.48 depending on the instrument)

(Dochy et al., 2003, p. 246-247). Instruments that measure knowledge acquisition demonstrate a similar trend, but in the opposite direction: the more the instrument measures "recognition tasks" (i.e., rote memorization or declarative knowledge), the more favorable LD students perform; whereas if the knowledge test includes "retrieval skills" (contextual, short-answer, or free recall), PBL students perform equally to or better than LD students (p. 548).

Meta-analysis of the effects of PBL across disciplines. Walker and Leary (2009) conducted a meta-analysis of 82 problem-based learning studies implemented between 2002 and 2007 to determine the extent to which academic discipline, problem type (degree of structure, difficulty, and solvability), method (e.g., lecture-based, case-based, problem-based learning, or closed loop problem-based learning), and assessment level (e.g., application, concept, principle) affect achievement outcomes. Inclusion criteria included reported effect sizes (or enough data to calculate it), a quantitative measure of student knowledge or skills, and a comparison between problem-based learning and traditional (lecture/discussion) conditions. Effect sizes were calculated using Cohen's *d*, dividing the difference in means by the pooled standard deviation (Walker & Leary, 2009). The PBL structure had to include the centrality of the ill-structured problem or scenario, "student-directed learning," and the teacher as facilitator not information giver (Walker & Leary, 2009 p. 19). For the purpose of this review, only academic discipline, method, and assessment type results are discussed and reported.

Results indicate that overall PBL had only a slightly favorable effect on achievement ($d_w = 0.13$, $p < .05$). However, effect size by discipline provides an interesting insight. PBL was practically significant in social science ($d_w = +0.30$) and teacher education (d_w = +0.64) compared to the nearly negligible effects in science (d_w = +0.06), medical education (d_w = +0.09), and engineering (d_w = +0.05) (Walker & Leary, 2009, p. 21). Only five of the studies reported the type of PBL method employed, all of which were closed loop designs (CLD). A problem-based CLD utilizes reflection, in which students "revisit the problem to determine any improvements they could make to their reasoning process" (Walker & Leary, 2009, p. 18). CLD demonstrated a moderately large effect on outcome achievement (d_w = +0.54). Assessment outcomes demonstrate that PBL students performed better on hypothesis-driven outcome measures, but less well on data-driven measures (Walker & Leary, 2009). At the concept (declarative knowledge) level, LD students barely outperformed PBL students (d_w = -0.043), however PBL students typically performed better on principle (d_w = +0.21) and application (d_w = +0.33) level outcome assessments.

Effect of PBL on Academic Achievement within Secondary Education

Batdi (2014) conducted a meta-analysis of 26 Turkish master's theses and doctoral dissertations investigating the effect of PBL on academic achievement compared to traditional instruction. The meta-analysis included studies conducted in primary, secondary, and post-secondary contexts, the majority of which were conducted at the secondary level (f = 17). Effect sizes were calculated across three subject areas: science, mathematics, and social science. Results indicated large to very large mean effects of d = 1.32 for science (f = 15), d = 0.79 for math (f = 6), d = 1.88 for social sciences (f = 5), all favoring PBL. All subjects combined yielded a summary mean effect of d = 1.30. It should be noted that most of these studies are unpublished and therefore not subjected to peer review.

Primary research on the effect of PBL on academic achievement within secondary contexts across disciplines suggests promising, but mixed results. Most of the primary studies discussed herein include a traditional, LD comparison. Reflection and discussion were key components of the intervention, but these components were not specifically measured. Wirkala and Kuhn (2011), however, included a LD comparison but also investigated the effect of discussion on PBL achievement. Three studies (Jewett & Kuhn, 2015; White & Frederiksen, 1998; Zohar & Ben David, 2008) in this review do not include a traditional, LD component, but are included as the intent of the authors was to isolate and investigate components of PBL (e.g., reflective assessment: White & Frederiksen, 1998; Zohar & Ben David, 2008; or discussion, Jewett & Kuhn, 2015) that effect achievement gains. While these three studies were not included in the meta-analysis due to their exclusion of a traditional, LD comparison (see Chapter 3), these studies are germane to understanding the effect of reflective, formative assessment practices and discussion on academic achievement with the use of PBL.

Experimental (PBL) conditions in each of the studies include students defining and solving problems or completing projects in small, collaborative groups. Learning activities include entry points/events, defining the problem, research, self-monitoring and evaluation assessments to activate prior knowledge and initiate further inquiry, and small and whole group discussion. The role of the teacher in PBL conditions includes monitoring group collaboration, fostering critical thinking, and directing students with open ended questions “when guidance was needed” (Sungar et al., 2011, p. 157) or looking for “teachable moments” (Mergendoller et al., 2006, p. 50). Teachers within traditional/LD conditions were directed to teach students as they had always done, using district and state required curriculum and standards. Thus, teachers typically delivered unit content through lecture, textbook readings, worksheet completion, and teacher-driven explanations and questioning.

Primary research comparing PBL and LD on academic achievement. In an often quoted study by Mergendoller, Maxwell, and Bellisimo (2006), advocates of PBL point to the overwhelming positive effect of PBL on academic achievement. Mergendoller et al. compared the effect of PBL (n = 139) and LD (n = 107) on Grade 12 economics achievement in a quasi-experimental, pre-posttest design across four schools and five teachers. Each teacher taught both an experimental and comparison group. The study lasted a single curriculum unit (5-10 days in length depending on the teacher) of an eight unit curriculum designed for a semester-long economics course developed by the Buck Institute for Education (BIE) for whom Mergendoller is the executive director and a principal researcher. Slavin (2008) suggested that one must be diligent in reviewing literature produced for predominantly commercial means, where possible conflict of interest issues arise, as these studies often tend to report overwhelmingly positive results.

Mergendoller et al. (2006) measured economics achievement using a 16-point multiple-choice exam developed by BIE using "items drawn from the Test of Economic Literacy and the test bank accompanying a widely used high-school economics textbook" (p. 56), which included application and analysis objectives as well as general knowledge items. Cronbach's alpha for this instrument was not reported. Results indicated that students in problem-based learning classes outperformed the comparison classes ($p < .05$) with an ES of d = 0.59 for the PBL classes, and d = 0.29 for LD classes. These analyses were determined by using a series of t-tests between pre- and posttest scores by class and condition, rather than by treatment (whole) and comparison (whole). Thus, these results are not only subject to Type I error, but also misleading. The use of pretest to posttest by individual class and condition as a means to determine effect size is problematic: if one is simply trying to demonstrate that PBL students "grew more" than LD students, this method is perhaps understandable, but methodologically and statistically questionable. Another issue is that calculations of the *t*-tests and subsequent effect sizes per teacher and condition do not add up to the reported *t*-values and effect sizes, leading one to question how analyses and subsequent effect sizes were conducted. It appears that the authors calculated the overall ES for each condition ("All Teachers," p. 60) by dividing the condition's mean change in score by the SD of the condition's mean score; thus, overestimating the effect of the PBL condition. Under the assumption that the mean score of the posttest for each condition was the sum of the overall condition pretest mean and posttest change in score, the ES for overall treatment effect using the Cohen's *d* (dividing the difference in means between the two conditions by a pooled SD) indicates that the PBL condition outperformed traditional instruction, but by and ES of $d = 0.15$, not 0.59. However, as already discussed, applying Cohen's *d* to the overall pretest and pretest-posttest change means is problematic, because one cannot assume the bases of the reported data.

It should also be addressed that a positive ES does not necessarily indicate acquisition of "academic achievement." An examination of the overall mean achievement (the sum mean pretest and pretest-posttest change) indicates a troubling realization. If one takes the PBL condition's mean change in score (1.48) and adds it to the PBL condition's mean pretest score (6.37), the mean posttest score for all PBL condition students is 7.85 (*SD* = 2.52) of 16 possible, or 49% correct. The mean posttest result for the LD condition was 7.45, or 47% correct. While this difference was statistically significant at $p < .05$, in actuality, it could be argued that students in both conditions failed the unit exam.

Finkelstein, Hanson, Huang, Hirschman, and Huang (2011) conducted a quasi-experimental study on the effectiveness of BIE Problem Based Economics (PBE) curriculum (five of eight modules) on economics knowledge (content) and skills (application of knowledge: problem-solving) achievement over a one semester term. Participants were 64 volunteer teachers with a range of experience across 72 schools randomly assigned to condition and 4,350 Grades 11 and 12 student participants nested within those teachers. Teachers in the PBL condition underwent 40 hours of professional development based on BIE recommendations. Throughout the study, teachers in the PBE condition received ongoing professional development from BIE. Teachers in the comparison, LD condition participated in their normal, school directed professional development. Thus, Finkelstein et al. cautioned that results should not be generalized to teachers who have not undergone professional development as prescribed by BIE (p. 49). However, the reality of the classroom is that most teachers will implement a program or curriculum without necessarily attending or ascribing to the available or suggested professional development.

To test knowledge acquisition, the researchers used Forms A and B of the Test of Economic Literacy (TEL), a 40 item, nationally normed measure developed by the National Council on Economic Education. Problem-solving skills were assessed on a performance task assessment developed by the National Center for Research on

Education, Standards, and Student Testing at UCLA. Of the five problem-solving tasks developed for the exam, students were randomly assigned two performance tasks contained within testing booklets. Results indicated that students in the PBE conditions outperformed students on the TEL and performance tasks. However, it should be noted that participation in the posttest was voluntary. Three hierarchical linear modeling (HLM) clusters were reported: Models A, B, and C; each model accounted for a different set of covariates. Finkelstein et al. (2011) reported Model C as the primary effect of PBE on the TEL (knowledge acquisition), $d = 0.32$, and performance task (application of skills), $d = 0.31$, a statistically and practically significant effect size on both measures. Model C included covariates that accounted for "randomization strata, baseline student TEL scores (pre-test), an indicator variable for missing data on the baseline TEL," and "student and teacher-level covariates [e.g., demographics and other survey-related information regarding interest in and experience with economics]" (p. 86). Although Model C is statistically "cleaner" in terms of eliminating the noise associated with covariates, Model B, which accounted for all but the student and teacher-level economics experience related covariates, represents a closer reality to the "messiness" of a classroom. Model A accounted for randomization strata only. Thus, for this meta-analysis, the researcher used the statistics reported in Model B: TEL $d = 0.21$, $p < 0.05$, and performance task $d = 0.15$. It should be noted that there was a statistically non-significant difference between PBL and LD students on the performance task in Model B. A comparison of ES to actual achievement based on mean scores (PBE (intervention) Model B: TEL $m = 22.24$, performance task $m = 6.60$; and LD (comparison) Model B: TEL $m = 20.50$, performance task $m = 6.30$, pp. 87, 89) indicate that students in both conditions scored below the national mean for the TEL (*m* = 25.74, p. 52) and scored at the low end of a composite score range of 4 to 12 for the performance task.

Parker et al. (2011) studied the effects of PBL and traditional (LD) instruction on Advanced Placement (AP) Government and Politics achievement on both the AP Government exam (composite scores range from 1-5) and a researcher developed complex-scenarios test (composite scores range from 1-6). The AP Exam was used to measure general government knowledge (identification and description) and the complex-scenarios test was used to assess students' ability to apply knowledge to a novel situation (p. 545). Three schools were included in the study, two in the PBL condition (School A, high achieving; School B, medium achieving) and one in the LD condition (School C, high achieving). The authors used hierarchical linear modeling to control for students' prior achievement, measured by GPA, PSAT scores, Washington Assessment of Student Learning-Reading scores, and any prior AP scores.

The researchers employed a design experiment (Brown, 1992) in which aspects of the experiment were modified and revised "for the sake of improving learning outcomes," thus "[g]eneralizability is not primarily the goal" (Parker et al., 2011, p. 556). These modifications were not identified by the authors. Thus, further quasi-experimental research is needed to confirm results. However, it should also be noted that responsive teachers are continually "manipulating" the learning environment to address the learning needs of students through use of formative assessment. Therefore it can be argued that a "design experiment" replicates the natural response(s) of a reflective teacher.

Results indicated that students in Schools A and B scored more 5's on the AP Government Exam and outperformed School C on the researcher designed, complex-scenarios test. In terms of overall, mean achievement on the AP exam, students in School A (*m* = 3.46) outperformed School C (*m* = 2.58), and School B (*m* = 2.40) performed comparably with School C. Compared to School C, the combined mean of Schools A and B (2.94) on the AP exam was statistically significant, $p < 0.05$ in regard to achievement, and calculates to an effect size of $d = 0.28$. On the complex-scenarios test, both Schools A (*M* = 2.34, *SD* = .091) and B (*M* = 2.07, *SD* = 0.83) outperformed School C (*M* = 1.61, *SD* = 0.75), $p < .05$. The combined mean of Schools A and B (*m* = 2.22) on the complex-scenarios test was statistically significant compared to School C (*m* = 1.61), $p < 0.05$, and calculates to an effect size of $d = 0.73$. Although the effect sizes indicate medium to high effect (Cohen, 1988), a comparison of mean academic achievement to ES is necessary. A (3) on the AP Exam is recognized by most colleges as a “passing” grade and earns college credit in some institutions. Only School A had a mean at/above a 3. In regard to the CST, scores were on the low end of the composite range for all three conditions (Parker et al., 2011).

Sungar et al. (2011) compared the effects of PBL and LD on achievement gains of 61 Grade 10 Turkish biology students in a quasi-experimental, pre- posttest design expanding a four week unit taught by the same teacher. Students in the PBL condition worked in small, collaborative groups of six to investigate ill-structured problems. In addition to role-playing, time was provided at the end of each session for self-reflection. Students monitored their learning and set goals for extended investigation. Students were expected to conduct independent research not only to meet personal goals, but also to contribute to their collaborative group. The role of the teacher in the PBL condition included monitoring group collaboration, fostering critical thinking, and directing students with open ended questions "when guidance was needed" (p. 157).

Results indicated that PBL students (*M* = 21.03, *SD* = 1.81) outperformed LD students (*M* = 17.75, *SD* = 2.43) on the 25-point multiple choice academic achievement test, which assessed students' general content knowledge and ability to apply that knowledge, ES $d = 1.53$. This ES is remarkably high. PBL students also outperformed the LD students on a 5-point performance skills (PS) essay, $p < 0.00$, which assessed students' ability to "use relevant information in addressing [a given] problem, articulate uncertainties, organize concepts; and interpret information" (p. 156). Again, the ES of this difference is remarkable, $d = 1.09$. However, a careful examination of PS mean scores for treatment and comparison show that PBL students averaged less than the mid-range 3 (*M* = 2.39, *SD* = 0.95), and LD students less than 2 (*M* = 1.49, *SD* = .68). Thus, scores should be interpreted cautiously. Critics of PBL might find these results evidence of the ineffectiveness of PBL to support the development of essential skills such as writing with evidence. Post treatment evaluation of students' perspectives on problem-based learning suggests that students had difficulty adapting to their increased role in the learning process (Sungar et al., 2011). Students expressed some desire for more structure and direct instruction. This resistance to a shift in the responsibility of learning from teacher to student by the students suggests that students may not readily adapt to PBL or the demands of self-directed learning.

Effects of reflective assessment and PBL on academic achievement. White and Frederiksen (1998) investigated the effects of reflective assessment and inquiry-based learning (IBL) on science achievement in a 10.5 week curriculum unit developed by the authors. Though described as "inquiry-based," the design parallels that of PBL. Participants were three teachers randomly assigned to intervention (IBL with reflective assessment) or comparison (IBL only) and 343 Grades 7, 8¸and 9 students within those teachers. White and Frederiksen hypothesized that metacognitive skills taught in authentic contexts would diminish the achievement gap between low- and high-achieving students, because "higher-achieving students already have implicit metacognitive skills for reflection, whereas low-achieving students lack such implicit skills" (p. 43).

Results indicated that low-achieving students in the experimental IBL with reflective assessment condition performed at par with or superior to high-achieving students in the comparison, IBL only condition. White and Frederiksen (1998) concluded that reflective assessment is particularly effective for low-achieving students. The authors also concluded that because reflective assessment in this experiment was practiced in conjunction with students' progress toward completing projects and evaluating their inquiry skills, not the acquisition of content/concept knowledge or solving problems, "reflective assessment [had] a direct effect on learning how to carry out inquiry, [but] only an indirect effect on the physics that was learned" (p. 67). Therefore, White and Frederiksen suggested that reflective assessment in conjunction with concept development is equally important to reflective assessment used to develop higher order thinking skills and monitor task completion.

Zohar and Ben David (2008) investigated the effects of explicitly teaching meta-strategic knowledge "mediated by verbal discussion" (p. 76) on achievement and retention in an inquiry-based science curriculum expanding 12 lessons. Meta-strategic knowledge is defined by Zohar and Ben David as a sub-component of metacognition that makes students explicitly aware of why, when, and how to use higher order thinking skills to solve problems (pp. 59, 62). Similar to the White and Frederiksen (1998) study, Zohar and Ben David hypothesized that explicit training would benefit low-achieving students more than high-achieving students, because high-achieving students “manage to construct elements of metacognitive knowledge by themselves” (p. 63). Participants were 119 (n = 45 boys, n = 64 girls) Grade 8 Israeli students in a public school. Six classes were randomly assigned to either the comparison (inquiry-based curriculum only) or experimental (inquiry-based curriculum with MSK instruction) condition. Results indicated that both high- and low-achieving students in the reflective assessment group outperformed both high- and low-achieving students in the comparison group on both the post- and retention-tests measuring basic concept knowledge and inference and application skills.

White and Frederiksen (1998) and Zohar and Ben David (2008) each discussed that explicitly teaching reflective assessment within contextual experiences aids the internalization and subsequent utilization of metacognitive strategies. Their studies also demonstrated that collaborative, reflective assessment and meta-strategic knowledge training may be particularly effective for low-achieving students. Zohar and Ben David noted that their implementation of explicitly teaching meta-strategic knowledge was moderated by verbal discussion and feedback; thus, research is needed to “discern the components of the educational intervention [discussion] and to evaluate their relative contribution to students’ reasoning gains” (p. 79).

Effects of discussion and PBL on academic achievement. Wirkala and Kuhn (2011) compared the effects of PBL on long term retention compared to traditional lecture discussion (LD) on Grade 6 social studies students in a quasi-experimental, within- and between-group, repeated measures design. To single out the effect of social interaction on PBL, Wirkala and Kuhn devised a team problem-based learning (TPBL) and an individual PBL (IPBL) condition. Results indicated that the TPBL and IPBL conditions outperformed the LD conditions on both the post- and retention-tests, and that TPBL and IPBL conditions performed comparably. Wirkala and Kuhn used *t*-tests with planned post hocs, which could increase Type I error, but the authors limited the scope of the comparisons to PBL combined v. LD, and TPBL v. IPBL. While effect sizes were not reported, calculation of effect sizes indicate that Cohen's *d* was between 0.80 and 0.95 depending on the test. Using Bonferroni's, a conservative post hoc measure (Field, 2009), the statistical significance of achievement between PBL and LD was well within acceptable terms (p = 0.001 to 0.006). These results suggested that PBL demonstrated greater effects on learning and retention than LD, but the social aspect of typical PBL practice did not increase student achievement compared to individual PBL.

Jewett and Kuhn (2015) investigated the effect of PBL on the inquiry skill "control of variables" on 79 Grade 6-7 low-achieving students in an urban charter school. Control of variables is used to generate valid inferences about various information (extraneous variables) on a "single focal variable" (p. 5). In this study, the problem-scenario was a real-world, contextual issue on juvenile crime and it causes. Jewett and Kuhn also investigated the active and social components of PBL on acquisition and application of the control of variables skill to determine whether it is "the problem or the social context that accounts for outcomes" (p. 6). Thus, there were four conditions: individual PBL condition (IPBL), a team PBL condition (TPBL), and observer condition, and a control condition (who did not participate in the intervention). Control of variables achievement was assessed using a post-intervention interview (videotape was used for scoring) and a 16-question, written test. Results indicated that students in the PBL conditions outperformed the observer and control conditions, but there were statistically non-significant differences between IPBL and TPBL students. In regard to the statistically non-significant difference between IPBL and TPBL students, the authors concluded that it is working with the problem, not the social context that promotes PBL effectiveness (p. 19).

The studies by Wirkala and Kuhn (2011) and Jewett and Kuhn (2015) suggest that the social component of PBL, a component that is often regarded as central to PBL success (Hmelo-Silver, 2004) and critical for success in the 21st Century (Larmer & Mergendoller, 2012a; Partnership for 21st Century Skills, 2011), does not account for PBL effectiveness. Kuhn (2015), however, questioned the notion that it is the problem only and not the collaboration that "provid[es] the benefit" (p. 48) in PBL. Kuhn, thus, postulated that in many PBL groups, talk is often centered on successful task completion rather than challenging learning. Thus, whether working in a group or individually, "goal failure" can be a motivating factor for learning. Another argument is also worth consideration and elaboration.

In both the Wirkala and Kuhn (2011) and Jewett and Kuhn (2015) studies the role of the coach (facilitator, tutor, teacher) was not intellectually disengaged from the learning process or the students. In the Wirkala and Kuhn study, if a student in the individual PBL (IPBL) condition had a procedural or content question, the coach responded in a number of ways depending on the question: redirecting the student to redefine the problem or re-read the problem-scenario, helping students understand the problem, and encouraging students to underline key ideas and facts necessary to solve the problem. As was the case in the Wirkala and Kuhn study, the coach in the Jewett and Kuhn study did not "give" answers but encouraged IPBL and team PBL (TPBL) students to review evidence when questions were asked. When both IPBL and TPBL students claimed they were done solving the problem, the coach listened to the presentations and if "the evidence the students presented was inadequate to justify a conclusion…it was the role of the coach to heighten students' recognition of the inadequacy of their evidence by suggesting alternative explanations" (p. 11). The coach did this by asking probing questions that challenged the students' thinking (IPBL and TPBL students). Thus, a social component did in fact exist in both studies.

As was discussed earlier, Piaget (1967) suggested that peer-to-peer interactions have important implications for collaborative learning methods. Namely, Piaget asserted that adult-to-child relationships create situations in which the child (student) is more likely to comply with the adult's (teacher's) thinking and ways of doing. However, in peer-to-peer interactions, students are "more likely to develop cognitively in contexts in which peers have equal power and all have opportunities to influence one another" (O'Donnell & Hmelo-Silver, 2013, p. 8). In the case of these two studies, since the coach acted as one who did not know the answers, but genuinely sought the ideas and conclusions of the student(s), the coach became a "peer," so to speak, and encouraged self-directed learning and cognitive growth through questioning, challenging ideas, and re-directing as necessary. These studies have important implications for the PBL environment: the role of the teacher as a co-learner and "additional peer" who challenges thinking is an equally important social component of PBL efficacy.

Summary

Meta-analyses of the effect of PBL on academic achievement in post-secondary (higher) education indicates that PBL students perform comparably or superiorly to LD instructed students, particularly on application of skills measures (Dochy et al., 2003; Vernon & Blake, 1993; Walker & Leary, 2009). However, there is great variance among these studies, particularly on achievement related to acquisition of knowledge (declarative knowledge) (Dochy et al., 2003; Vernon & Blake, 1993; Walker & Leary, 2009). Research in secondary education (Grades 6-12), on the other hand, indicates overwhelmingly positive effects of PBL on academic achievement compared to post-secondary education, although there is also variability among these studies. An investigation of what characteristics and/or moderating variables may contribute to these differences is valuable. Further, an exploration of the "conditions and practices associated with differences in effectiveness" (U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, 2010, p. 2) will provide context for implementation considerations.

Mean scores on knowledge acquisition (content) and application of knowledge (skills) tests indicate that students are not academically achieving despite medium to large effect sizes. The research in this review demonstrates the necessity for careful consideration of designing effective and aligned outcome measures (Dochy et al., 2003; Gijbels, Dochy, Van den Bossche, & Segers, 2005; Walker & Leary, 2009). White and Frederiksen (1998) analyzed the effects of reflective assessment on four types of outcomes: research projects, acquisition of inquiry skills, general content and concept knowledge acquisition, and concept application. Their study suggests that reflective assessment that targets only a specific element of the curriculum (e.g., inquiry skills) may demonstrate a direct impact on the projects and tests that emphasize this element, but only an indirect, if not negligible impact on outcome measures that emphasize content and/or extended application.

In each of the primary studies reviewed here and studies included in the meta-analysis (e.g., Parker et al., 2013) PBL students performed better on multiple choice measures of general knowledge, but struggled with the performance skills assessed on pen-and-pencil tests, as did their LD instructed counterparts, which required a high degree of technical writing. In light of these studies, it is important to remember that while project- and problem-based learning, reflective assessment, and academic discussion may promote content and concept formation and internalization, translation onto a test is a different matter. Jewett and Kuhn (2015) suggested that "further research is necessary to determine the extent to which failure to achieve full mastery on the written task should be attributed to cognitive challenges versus the challenges of the test format" (p. 18). It could also be argued that relative "failure" (low means) on these tests cannot necessarily be attributed to failure of PBL as an instructional approach either. Rather, explicitly teaching students how to access and articulate thinking into writing, applying concepts to a novel situation, and choosing among potential distractors (multiple-choice tests) is as important to the learning process as the problem-scenario itself.

Chapter 3: Research Methods

Although primary research within secondary (6-12) contexts (Mergendoller et al., 2006; Wirkala & Kuhn, 2011) indicates that PBL is often superior to traditional, lecture-based instruction and meta-analyses at the post-secondary level indicate that PBL is at par with or superior to traditional instruction (Dochy et al., 2003; Vernon & Blake, 1993; Walker & Leary, 2009), a synthesized and quantified exploration of the strength of relationship between PBL and academic achievement within middle school, junior high, and high school student populations (Grades 6-12) is needed.

Meta-analysis is one of many approaches to synthesize research literature within a particular domain attempting to answer the same question (Borenstein, Hedges, Higgins, & Rothstein, 2009; Lipsey & Wilson, 1993, 2001). Unlike qualitative research synthesis approaches, such as narrative reviews, meta-analysis is a quantitative synthesis of the selected research studies, which permits the "statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass, 1976, p. 3). Thus, meta-analytic procedures assist an investigator in quantitatively analyzing the context of a given effect, examine variance among different studies, and test the effects of moderator variables (covariates) (Borenstein et al., 2009; Field & Gillett, 2010; Lipsey & Wilson, 2001; Rosenthal & DiMatteo, 2001). Meta-analysis also provides a statistical means to combine the sample and effect sizes of several studies, which improves the power to detect a significant effect and "allows researchers to arrive at conclusions that are more accurate and more credible than can be presented in any one primary study or in a nonquantitative [sic], narrative review (Rosenthal & DiMatteo, 2001, p. 61).

Meta-analysis is similar to primary research in that the investigator develops a hypothesis (or hypotheses), provides a theoretical framework for the phenomenon under investigation, determines inclusion and exclusion criteria, reviews relevant research literature, collects data, conducts statistical analyses, and reports and discusses the results (Borenstein et al., 2009; Lipsey & Wilson, 2001; Rosenthal 1991; Rosenthal & DiMatteo, 2001). The main difference between meta-analysis [secondary research, cf. Glass, 1976] and primary research is that research studies are the focus of analysis, not participants (Lipsey & Wilson, 2001).

Rosenthal and DiMatteo (2001) wrote that there is no singular or correct way to conduct a meta-analysis; however, Lipsey and Wilson (2001) argued that "meta-analysis is conducted as a structured research technique in its own right and hence requires that each step be documented and open to scrutiny" (p. 5). Therefore, they suggested that a meta-analyst be transparent in the process and procedures for conducting the meta-analysis. In the following sections, the investigator describes and provides a rationale for the literature search methods, inclusion/exclusions criteria, study characteristics and coding measures, research synthesis methods (statistical procedures), and limitations and delimitations of this meta-analysis.

Literature Search Methods

The investigator conducted an extensive search for research literature using a variety of techniques suggested by Lipsey and Wilson (2001) and Slavin (2008), and demonstrated in meta-analyses conducted by Hass (2005), Igel and Apthorp (2015), Rasmussen (2013), and Walker and Leary (2009). The search for relevant literature began by identifying subject terms for *problem-based learning* and *project-based* *learning* in five prominent database systems: Academic Search Premier, Education Full Text, Education Resource Information Center (ERIC), PsychInfo, and JSTOR. The resulting subject terms were *problem based learning* (with and without a hyphen), *project based learning* (with and without a hyphen), *project method in teaching*, *active learning, inquiry learning*, and *collaborative learning.*

Active learning, inquiry learning, and collaborative learning as subject terms in relation to PBL are each somewhat problematic in that the operational definitions of these instructional methods are similar to or are components of PBL, but are distinct methods. Thus, the inclusion of these terms as part of the search criteria added a significant number of articles to review. Therefore, these three subject terms were combined with the terms problem- and project-based learning and project method in teaching to narrow the results to eligible articles. The intervention type subject terms were then combined with research type subject terms (e.g., research, effect(s)/effectiveness, and experiment/ experimental) and various sample type terms (e.g., high school, adolescents, junior high, etc.) (Lipsey & Wilson, 2001, p. 28). A similar search was conducted in ProQuest Dissertations and Theses and two online, university library catalogues for books or other sources with possible research on PBL. Titles and abstracts were then screened for inclusion criteria characteristics (Hass, 2005). If the abstract or title was unclear, the investigator retained the article or book for further review. Reference lists of potential inclusion articles and dissertations, as well as descriptive articles, were then searched for further research literature.

The investigator hand searched specific journals: *Interdisciplinary Journal of Problem-based Learning, Learning and Instruction, Instructional Science, The Journal of*

Learning Science, Journal of Experimental Education, Journal of Science Education and Technology, International Journal of Science Education, and *Research in Middle Level Education.* Online searching was conducted in Google Scholar, Google, Microsoft Academic Search, American Education Research Association, and other PBL dedicated associations. Further, the investigator used social media, such as Facebook, LinkedIn, and email to connect with researchers in PBL to solicit unpublished and/or submitted research reports.

Inclusion Criteria

The investigator determined inclusion criteria with the following considerations in mind: source characteristics, study characteristics and design, and methodological characteristics. The screening form the investigator used to collect data is available for review in Appendix C.

Source characteristics. The first three inclusion criteria relate to language range, time frame, and publication type. Studies were included if they were available in English. The abstract and extended summary of one study (Kuşdemir, Ay, & Tüysüz, 2013) was available in English and indicated high potential for inclusion. Google Chrome Translate was used to translate the rest of the document from Turkish to English. The translation was clear enough to determine final study eligibility and obtain statistics to calculate effect sizes. Studies conducted between 1985 and 2015 were included. The choice of this date range was based on three major educational reform movements in the past twenty years:

- 1985-2001: Post publication of *A Nation at Risk* and inclusion of 1990s standards' movement;

- 2001-2009: *No Child Left Behind Act of 2001*and high stakes testing to pre-implementation of *Race to the Top* (2009) and Common Core State Standards (2010) initiatives;
- 2010-2015: Current state: *No Child Left Behind, Race to the Top,* Common Core State Standards, and high stakes testing. (Ravitch, 2010)

Published and unpublished literature were included (see the "Publication bias analysis" section for further description).

Study characteristics and design. Study characteristics and design were further inclusion limiters. Studies had to include an operational definition of PBL consistent with PBL literature (cf. Gallagher & Stepien, 1996; Hmelo-Silver, 2004; Hmelo-Silver & DeSimone, 2013; Mergendoller et al., 2006; Wirkala & Kuhn, 2011). Included research had to compare PBL to traditional, lecture-based instruction. If the PBL intervention was compared to another form of PBL (e.g., PBL with and without scaffolds, cf. Belland, Glazewski, & Richardson, 2011) or a student-centered technique (e.g., inquiry-based learning) it was excluded. If a study compared two types of PBL to traditional, lecture-based instruction (e.g., individual and team PBL, cf. Wirkala & Kuhn, 2011), the study was included and the means and standard deviations for the types of PBL were averaged for a composite mean and standard deviation. PBL designs that included technology enhancements as part of the PBL interface were included, but online courses utilizing PBL were excluded. Research conducted in afterschool or summer programs were excluded. Further, studies were excluded if the PBL intervention participants were compared to a "true" control that did not receive the same content under study. Research had to be conducted in Grades 6-12, independently or in mixed grade classrooms.

The investigator included only quantitative studies in which the outcome measure was individual academic achievement via a content acquisition, skills application, or combination of content and skills test (i.e., no group tests). Studies were included if the posttest was immediate or delayed (retention). Quantitative designs were limited to quasi-experimental and experimental studies conducted in classroom settings based on ecological validity and logical/statistical considerations (Dochy et al., 2003); thus, experimental studies conducted in non-classroom settings, ex post facto, and single group designs were excluded.

Ecological validity considerations. Classroom based experimental and quasi-experimental designs are ecologically valid because they occur in a natural setting. A problem with using a pure experiment in which the treatment is not carried out in a classroom to investigate the effectiveness of an instructional strategy is that students develop within and react because of their environment (context) (Berliner, 2002). Slavin (1986) also argued that many experimental studies are "highly artificial" (p. 7) and do not mimic the reality of the classroom; namely because experiments conducted in laboratory settings are designed to eliminate any systematic error that might confound the intervention. Thus, the elimination of this "noise" in an artificial setting may render the intervention invalid in an actual classroom. The main difference between quasi-experimental and experimental designs is that quasi-experimental designs lack random selection and assignment of participants to conditions (Campbell & Stanley, 1963). Although students cannot be randomly selected or assigned to conditions, intact classes can be randomly assigned to conditions. Randomization of participants to conditions, a requisite of experimental designs, does not guarantee that the conditions are similar;

rather they are probabilistically similar (Campbell & Stanley, 1963). Even though groups within quasi-experimental designs technically lack randomization, they can approximate experimental conditions by holding constant internal validity confounds such as "history, maturation, testing, and instrumentation" and confirming similarity using pretest scores (Campbell & Stanley, 1963, p. 48). In a meta-analysis of meta-analyses of psychological, behavioral, and educational treatments, Lipsey and Wilson (1993) found that quasi-experimental, nonequivalent comparison group designs tend to underestimate (or "suppress") an overall effect size (p. 1193).

Logical and statistical considerations. Although single group designs can be ecologically valid, they are problematic. In relation to testing an intervention, an underlying assumption should be that regardless of the instructional method students are likely to demonstrate growth from pre- to posttest. Thus, one cannot attribute posttest gains to the intervention alone, even if a counterbalanced design is employed (Campbell & Stanley, 1963). When the same students receive a PBL treatment with one unit of study and then another unit of study with traditional instruction, one cannot assume that it is the treatment that made a difference; rival hypotheses abound. It could be that the unit of study in one could be more interesting/boring or easier/harder than the other, or a host of other reasons not controlled for in the study. Statistically speaking, Lipsey and Wilson (1993) found that single group designs tend to overestimate effect sizes by as much as 61% (p. 1193) and suggested that they not be included in meta-analyses. Lipsey and Wilson (2001) later concluded that "the standardized mean gain effect size statistic is different from that of the standardized mean *difference* effect size…It follows that these two effect size statistics should not be mixed in the same meta-analysis" (p. 45).

Ex post facto designs are often used when it is impossible, unethical, or not feasible to manipulate the independent variable (McMillan & Schumacher, 1997). Data analyzed through ex post facto designs derive from an event that occurred in the past (Creswell, 2013). Ex post facto studies are also problematic. Gall, Gall, and Borg (2007) explained that the "disadvantage" of this design is that "inferences about causality on the bases of the collected data are necessarily tentative" (p. 18), because other factors could explain the phenomenon under investigation. For example, students who self-select to a PBL program or school based on interest or other mitigating factors, might make the aforementioned group motivationally or fundamentally different than the comparison group who attends a traditional high school (cf. Simmers & Dickinson, 2012). This potentially confounding difference could explain differences in academic achievement irrespective of program type (i.e., a PBL school versus a traditional school).

Methodological characteristics. Studies were included if they had enough statistics to calculate effect sizes. Studies that did not test for group differences, an issue that can overestimate an effect size if the experimental group is inherently more able than the control (Campbell & Stanley, 1963), were included. The rationale for this decision was based on intention to code for pretest equivalency (See Appendix D) and that quasi-experimental, nonrandomized studies do not necessarily inflate effect sizes (Lipsey & Wilson, 1993; see discussion in the subsection "Ecological validity"). Reporting of test-retest reliability statistics or some sort of content validity discussion was also an inclusion criterion; however, studies that did not report these statistics were coded and included in the meta-analysis (see Appendix D). See "Methodological characteristics" related to moderator coding in the next section for explanation of how the investigator dealt with non-reporting of group equivalency and test-retest reliability in the main analysis.

Study Characteristics and Coding: Possible Moderators

In addition to inclusion criteria, the investigator coded each study for possible covariant (moderator) features. Similar to the inclusion criteria, studies were coded by source, study, and methodological characteristics. However, after careful consideration, the investigator had to limit the number of covariates examined in the meta-analysis because of the small number of studies. It is recommended that for regression, there are at least ten studies per covariate (Borenstein et al., 2009; Field, 2009). The coding schematic is available in Appendix D, which includes variables that were used for descriptive statistics only.

Source characteristics. Studies were coded for publication date range and publication type.

Study characteristics. Studies were coded for grade level (middle school: 6-8, high school: 9-12) and academic subject (science, math, social studies/history, English/language arts (ELA), and elective/other). In the Wright (2009) study, the researcher assessed academic achievement in ELA based on the Florida Comprehensive Assessment Test in both middle school and high school populations. In this study, the number of students per grade level (i.e. middle or high school) was not delineated. Therefore the investigator coded the study as "high school" based on the following reasons: (a) more teachers (n = 11) at the high school level were trained in PBL (and therefore in the PBL condition) compared to teachers at the middle school level (n = 5); (b) 17 teachers at the high school level compared to 15 teachers at the middle school level participated in the study (while these numbers are nearly equal, the number of sections each teacher taught of their respective intervention was not specified; (c) the researcher reported high correlation between the grade level versions of the Florida Comprehensive Assessment Test; and (d) there were more students in the PBL group who were likely high school students based on the higher proportion of high school teachers in the PBL condition (pp. 55, 67, 68). Sample size coding was based on Lipsey and Wilson's (1993) recommendations of "less than 50, 51 to 100, and more than 100" (p. 1195). Studies were coded for duration of treatment, as the length (shorter or longer) of the PBL experience may either suppress or magnify the effectiveness.

Studies were coded for "Role/Experimental facilitator" to account for the effect a researcher or the classroom teacher who facilitates the experimental intervention might have on the summary outcome. The facilitator's role, meaning the interactive nature of that facilitator, was coded as "highly interactive," "active," and "passive." These labels are somewhat subjective in that what one considers active another may interpret as highly active, or vice versa. It is also problematic that some authors omit key information; thus, highly interactive facilitator roles may be described in the definition of PBL, but not in the procedures. When coding, the investigator attributed facilitator activity based on what was described in the methods (procedure) section of the report rather than the theory described. After reconsideration, however, facilitator role (level of interaction with learning process) was omitted as a moderator variable due to the subjectivity of the category.

As the investigator analyzed each study, it became apparent that the inclusion (or not) of professional development/training in PBL might have a significant impact on the intervention effect, and therefore implications for implementation practices. Lipsey and Wilson (2001) argued that a meta-analyst should be flexible in their coding in that as one becomes more intimate with studies, other salient features or distinguishing characteristics begin to emerge. This moderating variable was not conceived *a priori* when designing the screening form. Thus, the investigator amended the coding schematic (Appendix D) to delineate between studies in which the teacher was the facilitator, but no professional development/training was indicated, the teacher was the facilitator and received professional development (brief or extensive) or was experienced in PBL, or the researcher facilitated the intervention (assumed familiarity or experience in PBL).

Lastly, studies were coded for the inclusion (or not) of reflective assessment and academic discussion, two salient features of PBL. Studies were coded as having one, both, or none of these features based on what the author explained in the procedures section of the study.

Methodological characteristics. Studies were coded for methodological characteristics. These characteristics included use of equivalency measures, outcome measure development (teacher, researcher, or standardized), and reporting of test-retest reliability. According to Jacobs (1991) and Wells and Womack (2003), a Cronbach's alpha (α) of 0.70 or higher is acceptable on teacher created assessments. For standardized, low-stakes tests in the social sciences a reliability α of 0.80 or 0.85 is acceptable; however, for high-stakes, standardized and placement tests, 0.90 or higher is required. In general, a higher α coefficient (e.g., 0.80) is preferable (Field, 2009). These considerations were taken into account when coding studies for reliability of the achievements tests. Since there is relatively small number of studies in the meta-analysis, the investigator decided to run the main meta-analysis with and without the studies that (a) did not report pre-intervention equivalency of conditions and (b) did not report test-retest reliability or the test-retest alpha was less than 0.70 to determine whether the summary mean effect changed significantly as a result of these studies rather than treat pre-intervention equivalence and test-retest reliability as moderators.

Research Synthesis Methods

The investigator used Comprehensive Meta-Analysis, Version 3 (Biostat, 2015) to analyze the effects of PBL on the following measures of academic achievement: content acquisition tests, skills application tests, and combined content and skills tests. Retention tests were also included in the analysis. The investigator also used Comprehensive Meta-Analysis, Version 3, to investigate the relationship between moderator variables and the resulting summary mean effects. Data from Comprehensive Meta-Analysis was exported to Microsoft Excel to conduct further descriptive statistics about study features.

Effect Size Calculation

Assessment of the dependent variable *academic achievement* in PBL literature is commonly measured using content acquisition (e.g., recall or comprehension questions) skills application (e.g., problem-solving, analysis, or synthesis problems), or combined content and skills tests. These results are then typically reported as mean scores, or change in score from pretest to posttest. Both Cohen's *d* and Hedge's *g* are standardized effect sizes calculated by dividing the difference of two conditions' means by the pooled standard deviation of each condition ($\bar{X}_1 - \bar{X}_2 / \sqrt{} \, (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 / n_1 + n_2 - 2)$, and are appropriate effect size estimates for determining differences between groups.

Borenstein et al. (2009) noted that Cohen's *d* slightly overestimates the effect size of small samples. Hedge's *g* corrects for this slight bias by multiplying *d* by a factor called *J* (*d* x *J*). *J* is calculated by the following formula 1 – (3 / 4*df* – 1), where *df* is calculated using the same *df* formula for estimating within groups standard deviation ($n^1 + n^2 - 2$) (Borenstein et al., 2009, p. 27). For the purposes of this meta-analysis, the investigator reported Hedge's *g* because there were a moderate number of studies within the meta-analysis with small sample sizes and because its use renders a slightly smaller, and perhaps a more accurate, summary effect size than Cohen's *d* (Borenstein et al., 2009). In cases in which standard deviation and/or mean scores were not reported, the investigator used *t*-statistics, *F*-scores, and *p*-values to compute Hedge's *g* (Lipsey & Wilson, 2001; Rosenthal, 1991). In cases in which the authors reported non-significant differences, an effect size of 0.00, a one-tailed *p*-value of 0.50, and a *Z*-value of 0.00 was entered (Rosenthal, 1995).

Cohen's *d* is reported more often than Hedge's *g* in educational research; however, the conceptualization of *d* and *g* in regard to magnitude and direction are the same. Cohen (1992) suggested that small, medium, and large effects sizes are *d* equals 0.20, 0.50, and 0.80 respectively (p. 157); therefore, *g* can be interpreted similarly. Hattie (2009) suggested that for educational interventions one consider small, medium, and large effect sizes as 0.20, 0.40, and 0.60 respectively (p. 9). Hattie warned, however, that such adjectives should be ascribed tentatively, because a small effect size does not mean an intervention is not worth pursuing nor does a large effect mean that an intervention is worth pursuing. Other considerations are important, including feasibility and cost.

Independence of Effect Sizes

In this meta-analysis, the effects of PBL were explored on the following measures of academic achievement: content acquisition tests, skills application tests, combined content and skills tests, retention of knowledge, retention of skills, and retention of knowledge and skills. Borenstein et al. (2009) suggested that when the intention is to examine the effects of an intervention on separate outcome types, it is appropriate to perform separate analyses for each outcome type. For the six outcome types, separate meta-analyses were conducted. However, a summary mean effect was also desired. In this case, it was important to consider how to deal with studies that reported more than one outcome (e.g., reported knowledge and skills separately) or more than one time point (e.g., posttest and retention).

When information comes from the same participants (i.e., two outcomes from the same students) to simply treat each outcome as if it is independent overestimates "the precision of the summary effect" (Borenstein et al., 2009, p. 226; cf. Lipsey & Wilson, 2001). Additionally, Borenstein et al. (2009) emphasized that treating outcome measures independently in an overall analysis contributes too much weight "to studies with more than two outcomes than studies with one outcome" (p. 226). Combining outcomes is one approach to this issue (Borenstein et al., 2009; Lipsey & Wilson, 2001). However, the investigator needed to determine the appropriateness of (a) combining content and skills assessments and/or (b) combining posttest and retention effects in the *same* analysis.

In regard to integrating content acquisition and skills application constructs, it was determined that combining the two was appropriate namely because it takes "knowledge" of general concepts and/or principals in order to solve problems (Gijbels et al., 2005). In this regard, the idea of applying "skills" is akin to applying knowledge, whether it is knowledge of how to solve problems or how to apply acquired (or even prior) knowledge to novel situations. Secondly, while testing of immediate and delayed knowledge (content) and skills (application of knowledge/content) may tap into different aspects of learning, the general idea of testing (typically) is to determine whether and to what extent one learned. In regard to intervention testing, post- and retention tests are used to determine the effectiveness of the intervention in general, both in the immediate and the long-term. The underlying assumption by advocates of PBL is that doing projects and solving problems (another form of "doing") have both an immediate and delayed benefit; thus, it was deemed appropriate to combining post- and retention tests to examine the effects of PBL "in general" on academic achievements across grade levels and academic subjects, rather than running separate analyses for each type of outcome.

Statistical Model

Fixed-effect and random-effects are two types of statistical models used to conduct meta-analyses (Borenstein et al., 2009; Hunter & Schmidt, 2000). In the fixed-effect model, the underlying true effect within each model is assumed to be the same, or "fixed." Thus, any slight variance within in studies is attributed to sampling error only (Borenstein et al., 2009). The generated common effect applies to the studies included in the analysis only and, therefore, cannot be generalized beyond those studies (Borenstein et al., 2009; Field, 2003; Hunter & Schmidt, 2000).

In contrast, the assumption in the random-effects model is that variance exists not only *within* each study because of sampling error, but also *between* studies due to actual design, participant, and methodological differences (Borenstein et al., 2009). These differences include, but are not limited to grade level, socio-economic status, and ability level(s) of the participants, year of publication (such as before or after major legislation regarding educational practices), and whether professional development was provided (or not) to the teachers carrying out an intervention. The generated summary effect size in a random effects model is a *mean effect,* meaning that the summary effect is a mean of the parameters of the included studies (Borenstein et al., 2009); therefore, the summary effect(s) are generalizable beyond the scope of studies included in the meta-analysis (Borenstein et al., 2009; Field, 2003; Hunter & Schmidt, 2000). Thus, for research related to educational settings, a random effects model is more appropriate because variance between studies is attributed to more than just sampling error. School policy-makers need to make decisions on interventions that will be effective within their school, a context beyond the scope of any given body of literature included in a meta-analysis (Borenstein et al., 2000; Field, 2003; Hunter & Schmidt, 2000).

Reporting of the main effect analysis includes the summary effects, reported in Hedge's *g* (see "Calculating effect sizes" in the next section); the test of whether the mean effect is not null (Z-value, $p < 0.05$); and whether the true effect varies from study-to-study (Q, $p < 0.05$). The percentage of variation not explained by the model is reported in *I*-squared (I^2), which indicates how much of the effect size variation might be explained by moderators. Lastly, tau (T) is the estimate of the standardized parameter about the true mean, which explains "the distribution of effect sizes about the mean effect" (Borenstein et al., 2009, p. 116) on the same scale as the selected effect size (e.g. Hedge's *g*). To determine this distribution, T is multiplied by 1.96 (95% confidence interval).

Forest plots were generated to visually analyze the variation within and between studies. A forest plot is a visual representation of each study's observed effect and confidence interval, each study's effect in relationship to other studies within the analysis, the precision of each study and therefore its relative weight within the analysis (represented by the size (large or small) of the plotted square that represents the effect size), and the summary effect (represented as a diamond) and its confidence interval (Borenstein et al., 2009). If the confidence interval of any one study or summary effect includes the null (0 on the horizontal line), then one cannot rule out that the true effect is actually null (Borenstein et al., 2009). Forest plots were generated for the summary mean effect, which includes outcome types, academic subjects that had two or more studies, and grade level (middle school and high school).

Stem and leaf plots were created to show the distribution of effect sizes in a concise manner (Rosenthal, 1995). The digits of the one and tenth places are listed as the stem (e.g., 1.0 or 0.8) and the hundredth places are listed as the leaf. Seven stem and leaf plot were created: one to demonstrate the distribution of effect sizes of all outcome measures, and six individual plots to demonstrate the distribution of effect sizes within each outcome type.

Analysis of Moderator Variables: Meta-regression

Meta-regression is the meta-analysis version of multiple regression in primary studies (Borenstein et al., 2009). Meta-regression is used to analyze the relationship and impact of moderators on the summary mean effect(s). Meta-regression can be conducted with a fixed- or random-effects model, based on the same principles as a main effect analysis. A random effects model was selected for the meta-regression analysis, because it was assumed that true effects within studies would vary from study-to-study due to inherent differences in school, student, and study characteristics. Reported statistics include whether at least one of the covariates is related to the summary mean effect (model fit: *Z*-value and *Q*-value, $p < 0.05$), whether there is variance among the study's true effects related to a covariate (goodness of fit: *Q*-value, $p < 0.05$), how much variance exists between effects (tau-squared, T^2), the percentage of variance not explained by the model without the covariates (I^2), and the proportion of variance explained by the covariates (R^2) (Borenstein et al., 2009; Field & Gillett, 2010).

Borenstein et al. (2009) and Field (2009) suggested a minimum of 10 studies per covariate when conducting meta-regression. A categorical moderator in meta-regression typically describes a "set" of covariates (Borenstein et al., 2009). That is, a categorical moderator such as *inclusion of reflective assessment (RA) and/or academic discussion (AD)* contains three covariates ($m - 1$, where *m* is the number of categorical variables within the moderator): neither RA nor AD mentioned in procedures; RA only mentioned in procedures; AD only mentioned in procedures; RA and AD mentioned in procedures. Therefore, a moderator variable with four variables ($4 – 1 = 3$) would require a minimum of 30 studies to conduct a reliable meta-regression; however, each covariate within that set would require 10 studies in order to render a statistically stable and meaningful analysis.

Publication Bias Analysis

Publication bias is a potential, but serious threat to meta-analysis results (Field & Gillett, 2010; Lipsey & Wilson, 1993; Rosenthal, 1979; Rosenthal & DiMatteo, 2001). Rosenthal and DiMatteo (2001) explained that the act of establishing inclusion criteria (and therefore exclusion) creates a certain amount of both subjectivity and bias in a meta-analysis (p. 66). An additional issue is that as rigorous as one might attempt to obtain all of the literature on a particular topic, this attempt is stymied by incomplete computer searches, lack of access, and limitations in translated materials (Lipsey & Wilson, 2001; Rosenthal, 1979; Rosenthal & DiMatteo, 2001). Added to the issue is that researchers are more likely to submit and editors are more likely to publish studies that demonstrate a statistically significant and large effect. The result of this "publication bias" is that what is most available to meta-analysts and primary researchers are overestimated effects of a particular intervention on target populations (Borenstein et al., 2009; Lipsey & Wilson, 1993, 2001).

To reduce publication bias, Lipsey and Wilson (1993, 2001) and Rosenthal and DiMatteo (2001) recommend that a meta-analyst search for unpublished research as rigorously as published works. This task was completed by searching ProQuest dissertations and theses, contacting researchers in PBL, and other online searches described in the "Literature Search Methods" section of this chapter. In addition, calculations of publication bias are recommended (Borenstein et al., 2009; Field & Gillet, 2010). Orwin's fail-safe *N* (Orwin, 1983) is one method to account for publication bias. In addition, funnel plot analysis and Duval and Tweedie's (2000) *Trim and Fill* method were employed to statistically analyze the existence of and adjust for publication bias.

Orwin's fail-safe *N* is an analytic tool used to determine the number of additional studies (typically unpublished or non-retrieved studies) needed to render the summary mean effect size trivial (Orwin, 1983). Orwin's fail-safe *N* is computed as $N_{fs} = N_0 (\bar{d}_0 - \bar{d}_c) / \bar{d}_c - \bar{d}_{fs}$, where N_{fs} is the resulting number of fail-safe studies, N_0 is the number of observed studies in the meta-analysis, $\bar{d}_0$ is the observed summary mean effect in the meta-analysis, $\bar{d}_c$ is the criterion effect size that will render the mean effect trivial, and d_{fs} is the assumed effect size reported in the unpublished or non-retrieved studies. Comprehensive Meta-Analysis, Version 3, calculates Orwin's fail-safe *N* using the fixed effect grand mean rather than the random effects summary mean. Field and Gillett (2010) argued, however, that Orwin's fail-safe *N* addresses the wrong issue in that the method calculated the number of studies needed to "reverse a conclusion" (p. 686). Rather, the issue at hand is whether there is bias and what can be done to adjust for this bias (Borenstein et al., 2009; Field & Gillett, 2010).

A funnel plot provides a visual mechanism to examine the potential existence of publication bias. A funnel plot juxtaposes each study's effect size along the horizontal axis against its precision (measured in standard error) along the vertical axis (Field & Gillett, 2010). The resulting shape is a funnel, with smaller studies typically dispersing along the bottom and larger studies funneling and condensing to the middle top (Higgins & Green, 2011; Terrin, Schmid, & Lau, 2005). A symmetrical shape indicates the lack of publication bias. The problem with relying on funnel plots alone to determine whether publication bias exists is that other confounds may explain asymmetry, such as true heterogeneity between studies, inadequate analysis, and methodological designs (Higgins & Green, 2011; Terrin et al., 2005). Further, interpretation of funnels plots are often subjective (Duval & Tweetie, 2000) and misinterpretation of funnel plot symmetry is not uncommon (Terrin et al., 2005). Therefore, Duval and Tweedie's (2000) *Trim and Fill* method was included as a secondary analysis.

Duval and Tweedie's (2000) *Trim and Fill* method is based on the funnel plot method in which asymmetric studies are trimmed and the remaining "symmetric" studies are used to calculate a new "true center of the funnel and then replace the trimmed studies and their missing counterparts around the center" (Duval & Tweedie, 2000, p. 457). The resulting effect size is an adjusted summary effect that accounts for publications bias. Duval and Tweedie's (2000) *Trim and Fill* method was also conducted to determine an adjusted effect size for the summary mean effect across outcome types, grade levels, and academic subjects in the event of potential publication bias depicted in the funnel plot.

Limitations and Delimitations of the Study

There are several limitations related to selecting only quasi-experimental and classroom-based experimental studies to include in this meta-analysis. The first is that given the nature of classrooms as dynamic environments and variation among teaching styles and enthusiasm, other confounds could explain the results (Finkelstein et al., 2011; Slavin, 2008). Further, since teachers volunteer to open their classrooms for such studies, one cannot rule out that the types of teachers who volunteer for a study or are interested in implementing PBL are not somehow fundamentally different than those who do not volunteer (either for the study or to try PBL) (Slavin, 2008).

Another limitation is the focus on the effectiveness of PBL in Grades 6-12. The decision to limit the scope of research to Grades 6-12 is that secondary teachers often want research related to their grade levels, arguing that what works in elementary school does not necessarily work in middle and high school due to the assumed expectation of content coverage over exploration and extended learning (Gallagher & Stepien, 1996). Thus, generalizability of results is limited to Grades 6-12 populations.

The limited number of studies included in the meta-analysis created another limitation in that exploration of possible covariates was limited. Field (2009) and Borenstein et al. (2009) recommend at least 10 studies per covariate, which was unattainable in most cases and restricted moderator analyses to primarily the entire model (all studies across academic subjects and outcome types).

Delimitations are deliberate limitations implemented by the investigator to limit the scope of the research or make it manageable. Delimitations of this study were the focus on Grades 6-12; inclusion of quasi-experimental and classroom based experimental studies for ecological validity; limiting studies conducted between 1985 and 2015; and limiting studies to only those that compare PBL to traditional, lecture-based instruction. English translation was an additional delimitation in order to properly analyze the operationalized definition of PBL, procedures, and statistics reported in the study. This delimitation, simultaneously created a limitation. The meta-analysis conducted by Batdi (2014) provided 17 potential secondary level studies for inclusion; however, only two of those studies were in 6-12 populations and translated into English. The investigator also delimited the analysis to academic achievement, rather than investigate other outcomes of PBL such as attitude, self-regulation, self-efficacy, attendance, discipline referrals, or motivation.

Summary

The investigator conducted a meta-analysis using a random-effects model to investigate the effects of PBL on academic achievement after an extensive search for published and unpublished literature using a variety of search techniques. Moderators were coded and explored through separate meta-analyses, and use of meta-regression was considered if each covariate within the moderator set contained a minimum of ten studies each (Borenstein et al., 2009). These moderators included publication date, grade level, academic subject, duration of intervention, ability level, outcome type, sample size, inclusion (or not) of reflective assessment and/or academic discussion, and impact of the type of facilitator and to the extent that facilitator was trained and/or experienced in PBL. Publication bias was addressed by calculating Orwin's fail-safe *N*, visually inspecting a funnel plot, and running a *Trim and Fill* analysis. In the following chapters, the investigator reports the results of the main analysis and meta-regression, and then discusses these results related to hypotheses, theory, and research.

Chapter 4: Results

Search Results and Study Characteristics

The investigator acquired 72 studies that appeared to meet source, study, and methodological inclusion criteria after screening titles and abstracts of studies that resulted from the initial searches. After a second, extensive review, 38 studies were excluded and 34 -were retained for inclusion in the meta-analysis. Table 2 shows the reasons and rates for exclusion.

Table 2

Excluded Studies: Reasons and Rates

Primary Reason for Exclusion	*f*	Rate
Research design	**11**	**28.9**
Experimental, non-classroom setting	1	2.6
Ex Post Facto	3	7.9
Single group, pre-posttest	4	10.5
Non-quantitative: Case study, ethnography, etc.	3	7.9
Comparison condition not traditional, lecture-discussion instruction	**11**	**28.9**
True control condition: Did not receive same content under study	**2**	**5.3**
Operational definition inconsistent with or not PBL	**5**	**13.2**
Insufficient/appropriate data to calculate effects sizes	**4**	**10.5**
Grade/academic level outside scope of study	**1**	**2.6**
Different outcome measures than variables under study	**2**	**5.3**
Intervention (PBL) conflated with other instructional techniques	**1**	**2.6**
Possible duplicate study published in two journals	**1**	**2.6**

The following description provides a summary of the 34 retained studies. Among the 34 retained studies, one study (Ridlon, 2009) included two independent samples. A second study (Wirkala & Kuhn, 2011) used a crossed within samples design in which the same participants experienced different treatments with different unit topics separated by a summer break. The data within these two studies were treated as independent sample and effect sizes for the main analyses. For descriptive purposes the Wirkala and Kuhn (2011) study was treated as one sample. Therefore, the demographics of 35 studies are described herein. Detailed information can be retrieved in Tables E1, *Overview of Retained Studies* (Appendix E), and F1, *Summary of Study Characteristics* (Appendix F). References for all studies included in the meta-analysis can be reviewed in Appendix G.

The total sample size was 9,998 participants (PBL, *n* = 5,519; Traditional, LD, *n* = 4,479) and the majority of studies (*n* = 21) had samples sizes larger than 100. Ten countries are represented with the highest proportion in the United States (*n* = 18) and Turkey (*n* = 8). Two of the studies were conducted between 1985 and 2001; the remaining 33 studies were split nearly evenly with 17 conducted between 2001 and 2009 and the other 16 studies conducted between 2010 and 2015. A majority of the authors operationalized the intervention as problem-based learning (*n* = 23). The most frequent treatment duration was 4-6 weeks (*n* = 12). There were 16 studies conducted in science, seven in math, nine in social studies, one in ELA, and two coded as elective/other. Twelve studies were conducted at the middle school level (Grades 6-8) and 23 at the high school level (Grades 9-12). Three studies did not indicate pretest/group equivalency and 11 did not report test-rest reliability of outcome measures. In all, 48 outcome effect sizes were generated. Academic achievement was most commonly assessed with content (knowledge) tests (*n* = 26). Table 3 shows the number of effect sizes generated for each outcome type.

Table 3

Number of Effect Sizes by Outcome Type

Outcome Type	Number
Posttests	
Acquisition of Content	26
Application of Skills	9
Combination Content & Skills	6
Retention Tests	
Retention of Content	3
Retention of Skills	2
Retention of Content and Skills	2

Main Analysis Results

Summary Mean Effect

The summary mean effect across outcomes, academic subjects, and grade level was generated by conducting separate analyses with and without studies that (a) did not report pre-intervention equivalency of conditions and (b) did not report test-retest reliability, or the test-retest alpha was less than Cronbach's alpha of 0.70, to determine whether the summary mean effect changed significantly as a result of these studies. The main analysis with the inclusion of all study outcomes yielded a summary mean effect of g = 0.59 with a 95% confidence interval (CI) of g = 0.44 at the lower limit (LL) and g = 0.75 at the upper limit (UL). The statistically significant Z-value of 7.695, $p < .001$, disconfirms that the effect is zero. Tau, 0.42, indicates that most effects (95%) are distributed $g = \pm 0.82$ about the mean. A second analysis was then conducted to test the influence of non-report of pre-intervention equivalency. Three studies were extracted: Araz and Sungar (2007), Cicchino (2015), and Elshefei (1998). The adjusted analysis generated a summary mean effect of g = 0.69, $p < .001$. A third analysis was conducted to test the impact of studies with non-report of test-retest reliability or reporting of a Cronbach's alpha less than 0.70. Fourteen outcome measures were extracted from the model (see Appendix G, studies marked with an asterisks had one or more outcome measures extracted). This analysis yielded an effect size of $g = 0.66$, $p < .001$. These results suggest that the extracted studies suppressed, not inflated, the summary mean effect, and corroborate Lipsey and Wilson's (1995) assertion that nonequivalent comparison group designs as well as lower quality methodological designs are "almost as likely to be an underestimate as an overestimate" (p. 1193) of the effect size. Therefore, all studies were retained for further analyses.

Study Name	Outcome Type	Grade Level	Subject	Hedge's *g*	Relative weight	Lower limit	Upper limit
Cicchino (2015)	Posttest Content	Middle School	Social Studies	-0.216	3.02	-0.525	0.092
Tan et al. (2007)	Posttest Content & Skills	Middle School	Social Studies	-0.154	3.14	-0.408	0.101
Araz & Sungar (2007)	Combined	Middle School	Science	0.243	3.11	-0.027	0.513
Roesch, Nerb, & Riess (2015)	Posttest Skills	Middle School	Science	0.270	3.10	-0.001	0.541
Yancy (2012)	Posttest Content & Skills	Middle School	Math	0.345	2.85	-0.027	0.717
Ridlon (2009b)	Posttest Content	Middle School	Math	0.490	2.38	-0.054	1.034
Wong (2012)	Combined	Middle School	Science	0.598	2.89	0.242	0.955
Ankinoglu & Tandogan (2007)	Posttest Content	Middle School	Science	0.633	2.34	0.073	1.192
Wirkala & Kuhn (2011b)	Combined	Middle School	Social Studies	0.833	2.61	0.372	1.294
Wirkala & Kuhn (2011a)	Combined	Middle School	Social Studies	1.083	2.48	0.573	1.593
Hernandez-Ramos & De La Paz (2009)	Combined	Middle School	Social Studies	1.342	2.91	0.990	1.694
Ridlon (2009a)	Posttest Content	Middle School	Math	1.659	2.18	1.036	2.282
Bayrak & Bayram (2011)	Posttest Content	Middle School	Science	1.792	2.20	1.178	2.405
Visser (2003)	Posttest Content & Skills	High School	Science	-0.767	2.41	-1.303	-0.230
Elshafei (1998)	Combined	High School	Math	-0.148	3.20	-0.372	0.076
Brokes (2010)	Combined	High School	Math	-0.059	2.95	-0.392	0.274
Mergendoller et al. (2000)	Posttest Content	High School	Social Studies	0.000	2.90	-0.355	0.355
Wright (2009)	Posttest Content & Skills	High School	ELA	0.117	3.39	0.012	0.221
Mergendoller et al. (2006)	Posttest Content	High School	Social Studies	0.150	3.15	-0.101	0.402
Chang (2001)	Combined	High School	Science	0.151	3.01	-0.160	0.462
van Loggerenberg-Hattingh (2003)	Retention: Content & Skills	High School	Science	0.168	2.96	-0.163	0.498
Finkelstein et al. (2011)	Combined	High School	Social Studies	0.179	3.42	0.113	0.246
Parker et al. (2011)	Combined	High School	Social Studies	0.504	3.11	0.235	0.774
Parker et al. (2013)	Combined	High School	Social Studies	0.511	3.15	0.258	0.763
Worry (2011)	Posttest Content	High School	Math	0.605	2.53	0.113	1.096
Mioduser & Betzer (2007)	Posttest Content	High School	Other/Elective	0.697	2.77	0.295	1.099
Vilardi (2013)	Posttest Content & Skills	High School	Science	0.705	3.01	0.392	1.017
Bas (2011)	Posttest Content	High School	Other/Elective	0.786	2.45	0.267	1.305
Anyafulude (2013)	Posttest Content	High School	Science	0.798	3.14	0.542	1.055
Maree & Molepo (2005)	Posttest Content	High School	Math	0.834	3.26	0.636	1.031
Tarhan et al. (2008)	Posttest Content	High School	Science	0.992	2.60	0.526	1.459
Kusdemir, Ay, & Tuysuz (2013)	Posttest Content	High School	Science	1.209	2.28	0.625	1.793
Sungar et al. (2006)	Combined	High School	Science	1.293	2.38	0.746	1.841
Afolabi & Akinbobola (2009)	Posttest Content	High School	Science	1.744	2.65	1.297	2.191
Nafees (2012)	Posttest Content	High School	Science	1.867	2.32	1.297	2.436
Tarhan & Acar (2007)	Posttest Content	High School	Science	2.292	1.78	1.504	3.080
				0.594		0.443	0.745

Forest Plot

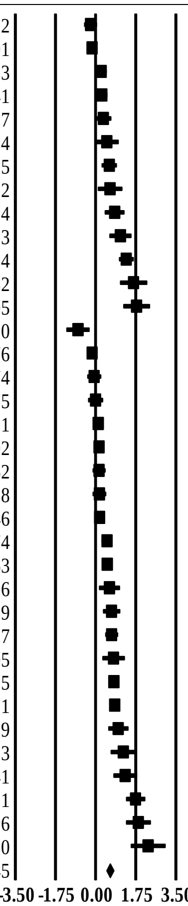


Figure 1. Forest plot of observed and mean effects across outcome types, academic subject, and grade level.

Figure 1 shows the forest plot of the range of individual, observed effects and the summary mean effect. A review of the forest plot indicates that there is variability among the individual studies ranging in effect size from *g* = -0.77 to 2.29. This plot accounts for independence of effect size; thus, same sample effect sizes are combined. The stem and leaf plot (Figure 2) shows the full distribution of the 48 outcomes generated from the 36 studies. The majority of individual, observed effect sizes cluster between *g* = -0.22 and 1.00. Nine effect sizes then disperse between *g* = 1.20 and 2.30.

Stem	**Leaf**
-0.7	7
-0.6	
-0.5	
-0.4	
-0.3	
-0.2	2
-0.1	8, 6, 5, 4
-0.0	4
0.0	0, 6
0.1	2, 5, 5, 6, 7
0.2	1, 7, 8
0.3	2, 3, 4, 4, 7
0.4	9, 9
0.5	
0.6	0, 3, 4
0.7	0, 0, 0, 1, 3, 9
0.8	0, 3, 8
0.9	9
1.0	3, 8
1.1	
1.2	1, 9
1.3	
1.4	
1.5	1
1.6	6
1.7	4, 9
1.8	9
1.9	
2.0	
2.1	
2.2	9
2.3	1

Figure 2. Stem and leaf display of 48 academic achievement outcome effect sizes.

Effects of PBL by Outcome Measure

The independent variable, academic achievement, was measured in the various studies with the following post- and retention tests: acquisition of content, application of skills, combination of content skills, retention of content, retention of skills, and retention of content and skills. Separate meta-analyses were conducted for posttest achievement (content and skills) and retention test achievement (content and skills), as well as for each outcome type independently. In each case, analyses included all grade levels and academic subject areas.

Pooled Posttest Results. The summary mean effect for posttest achievement measuring both content and skills indicates that PBL conditions outperformed traditionally instructed conditions, g = 0.62, CI g = 0.45 (LL) to 0.73 (UL), 24Z =7.11, p < .001,. A tau value of 0.46 indicates that 95% of the observed effects are expected to vary g = ± 0.90 about the mean effect size.

Acquisition of Content Tests. The summary mean effect for acquisition of content (n = 26) is g = 0.78, Z = 6.98, p < .001, disconfirms the null hypothesis that the effect is zero. The CI of this summary effect is g = 0.56 (LL) to 1.00 (UL). A tau of 0.53 indicates that most effects (95%) distributed g = ±1.04 about the mean. Figure 3 shows the stem and leaf distribution of observed effects for acquisition of knowledge tests. Effect sizes range from g = -0.22 to 2.31, with three clustered areas: the first from -0.22 to 0.49, the second from 0.60 to 1.20, and third from 1.50 to 1.80.

Stem	Leaf
-0.3	
-0.2	2
-0.1	6
0.0	0, 6
0.1	5, 6
0.2	8
0.3	4
0.4	9
0.5	
0.6	0, 3
0.7	0, 09
0.8	0, 3
0.9	9
1.0	3, 8
1.1	
1.2	1
1.3	
1.4	
1.5	1
1.6	6
1.7	4, 9
1.8	9
1.9	
2.0	
2.1	
2.2	9
2.3	1

Figure 3. Stem and leaf display of 26 acquisition of content effect sizes

Application of skills tests. Analysis of studies that reported use of an application of skills test (n = 9) generated an effect size of g = 0.24, CI of g = 0.06 (LL) to 0.43 (UL), Z = 2.6, p = .009, disconfirming the null hypothesis that the effect size is zero. Tau, 0.24, indicates that most effects (95%) are distributed g = ±0.47 about the mean. Figure 4 shows the stem and leaf distribution of observed effects for acquisition of knowledge tests. Effect sizes range from g = -0.18 to 1.00, with the majority of the effects sizes ranging between g = -0.18 and 0.33.

Stem	Leaf
-0.1	8, 4
-.0.0	4
0.0	
0.1	5
0.2	7
0.3	2, 3
0.4	
0.5	
0.6	
0.7	3
0.8	
0.9	
1.0	8

Figure 4. Stem and leaf display of nine application of skills effect sizes

Combination of content and skills tests. Six studies reported use of an outcome measure that combined content acquisition and skills application in a single test. Content and skills as a combined measure generated an effect size of *g* = 0.16 and CI of *g* = -0.14 (LL) to 0.45 (UL). However, a statistically non-significant *Z*- value of 1.04, *p* = .301, indicates that the true mean effect could be zero. Figure 5 shows the forest plot of the six included studies for combined content and skills assessments by grade level and academic subject.

Study Name	Outcome Type	Grade Level	Subject	Hedge's *g*	Relative weight	Lower limit	Upper limit	Forest Plot
Visser (2003)	Posttest Content & Skills	High School	Science	-0.767	12.36	-1.303	-0.230	
Tan et al. (2007)	Posttest Content & Skills	Middle School	Social Studies	-0.154	18.20	-0.408	0.101	
Wright (2009)	Posttest Content & Skills	High School	ELA	0.117	20.55	0.012	0.221	
Yancy (2012)	Posttest Content & Skills	Middle School	Math	0.345	15.74	-0.027	0.717	
Wong (2012)	Posttest Content & Skills	Middle School	Science	0.491	16.13	0.138	0.845	
Vilardi (2013)	Posttest Content & Skills	High School	Science	0.705	17.02	0.392	1.017	
				0.155		-0.138	0.448	

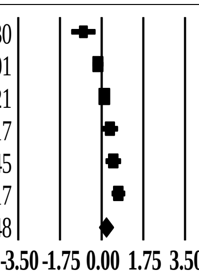


Figure 5. Forest plot of content and skills, assessment developer, and academic subject.

A visual inspection of the stem and leaf display for the combined content and skills outcome measure (Figure 6) shows the absence of any clustering, demonstrating a range of singleton studies dispersed between *g* = -0.70 and 0.70.

Stem	Leaf
-0.7	7
-0.6	
-0.5	
-0.4	
-0.3	
-0.2	
-0.1	5
0.0	
0.1	2
0.2	
0.3	4
0.4	9
0.5	
0.6	
0.7	0

Figure 6. Stem and leaf display of six content and skills outcome effect sizes.

Retention tests. Seven studies reported the use of retention tests to measure academic achievement in PBL. The summary mean effect for retention test achievement of g = 0.60, CI of g = 0.29 (LL) to 0.90 (UL), Z = 3.82, p < .001, disconfirms the null hypothesis that the mean effect is zero. A tau value of 0.29 indicates that 95% of the effect sizes distributed g = ±0.56 about the mean effect size. Retention of content (N = 3) yielded an effect size g = 0.87, CI of g = 0.29 (LL) to 1.45 (UL), Z = 2.93, p = .003. Tau, 0.46, indicates that most true effects (95%) are distributed g = ±0.91 about the mean. Observed effect sizes within content retention range from g = 0.37 to 1.29. Two studies reported use of a retention of skills test, both of which came from the same study (Wirkala & Kuhn, 2011), g = 0.75, CI of g = 0.41 (LL) to 1.08 (UL), Z =4.36, p < .001. There was a statistically non-significant result for retention of content and skills (N = 2), g = 0.43, CI of g = -0.10 (LL) to 0.96 (UL), Z (1.61), p = .109, indicating that the mean effect could be zero. This statistically non-significant result is confirmed by examining the confidence interval, which includes the null, -0.10 (lower limit) to 0.96 (upper limit).

Effects of PBL by Academic Subject Area

Five academic subjects were coded for the meta-analysis: science, math, social studies, ELA, and elective/other. Only one study was conducted in ELA (Wright, 2009); thus a meta-analysis was not conducted. Descriptively, the effect size of this large (N = 1,423), multi-grade study is relatively small, g = 0.12. There were two studies in the "elective/other" designation (Baş, 2011; Mioduser & Betzer, 2007) generating a mean effect size of g = 0.73, Z = 4.51, p < .001. The following reports the effects of PBL on science, math, and social studies.

Effect of PBL on science learning. Meta-analysis of science achievement (n = 16) generated a summary mean effect of g = 0.83, CI of g = 0.53 (LL) to 1.14 (UL), Z = 5.33, p < .001, disconfirming the null hypothesis that the effect size is zero. The forest plot of the effect of PBL on science learning (Figure 7) shows a wide distribution of effects from g = -0.77 to 2.29, which subsequently bound the range of effects for all studies included in the meta-analysis.

Study Name	Outcome Type	Grade Level	Subject	Hedge's *g*	Relative weight	Lower limit	Upper limit
Visser (2003)	Posttest Content & Skills	High School	Science	-0.767	5.95	-1.303	-0.230
Chang (2001)	Combined	High School	Science	0.151	6.78	-0.160	0.462
van Loggerenberg-Hattingh (2003)	Retention: Content & Skills	High School	Science	0.168	6.72	-0.163	0.498
Araz & Sungar (2007)	Combined	Middle School	Science	0.243	6.90	-0.027	0.513
Roesch, Nerb, & Riess (2015)	Posttest Skills	Middle School	Science	0.270	6.89	-0.001	0.541
Wong (2012)	Combined	Middle School	Science	0.598	6.63	0.242	0.955
Ankinoglu & Tandogan (2007)	Posttest Content	Middle School	Science	0.633	5.86	0.073	1.192
Vilardi (2013)	Posttest Content & Skills	High School	Science	0.705	6.77	0.392	1.017
Anyafulude (2013)	Posttest Content	High School	Science	0.798	6.93	0.542	1.055
Tarhan et al. (2008)	Posttest Content	High School	Science	0.992	6.23	0.526	1.459
Kusdemir, Ay, & Tuysuz (2013)	Posttest Content	High School	Science	1.209	5.76	0.625	1.793
Sungar et al. (2006)	Combined	High School	Science	1.293	5.91	0.746	1.841
Afolabi & Akinbobola (2009)	Posttest Content	High School	Science	1.744	6.31	1.297	2.191
Bayrak & Bayram (2011)	Posttest Content	Middle School	Science	1.792	5.63	1.178	2.405
Nafees (2012)	Posttest Content	High School	Science	1.867	5.82	1.297	2.436
Tarhan & Acar (2007)	Posttest Content	High School	Science	2.292	4.91	1.504	3.080
				0.831		0.525	1.137

Forest Plot

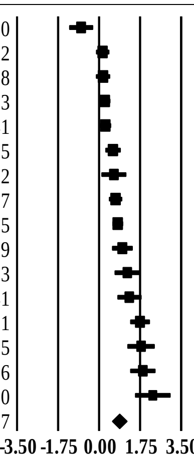


Figure 7. Forest plot of observed effects and the summary mean effect within science.

Further, the lower and upper boundary effects generated from studies conducted at the high school level. The distribution of effects, *T* = 0.58, indicates that most (95%) are distributed *g* = ±1.53 about the mean. The stem and leaf display (Figure 8) provides a visual display of this dispersion: observed effects of PBL in science range from *g* = -0.77 to 2.29. Nine of 16 studies clustered between *g* = 0.16 and 0.37, which are typically associated with a moderately small to medium effect size.

Stem	Leaf
-0.7	7
-0.6	
-0.5	
-0.4	
-0.3	
-0.2	
-0.1	
-0.0	
0.0	
0.1	6, 7
0.2	4, 7
0.3	2, 3, 4, 4, 7
0.4	
0.5	
0.6	0
0.7	1
0.8	0
0.9	9
1.0	
1.1	
1.2	1, 9
1.3	
1.4	
1.5	
1.6	
1.7	4, 9
1.8	8
1.9	
2.0	
2.1	
2.2	9

Figure 8. Stem and leaf display of the effect of PBL on science learning.

Effect of PBL on math learning. Results for math achievement (*n* = 7) generated a summary mean effect of *g* = 0.50, CI of *g* = 0.08 (LL) to 0.92 (UL), *Z* = 2.31, *p* = 0.02. The estimate of the distribution about the mean (95% of studies), *T* = 0.53, was calculated as g = ±1.04. The forest plot (Figure 9) indicates that observed effect sizes range from g = -0.15 to 1.66 and that each of the studies within the model contributed similar relative weights.

Study Name	Outcome Type	Grade Level	Subject	Hedge's g	Relative weight	Lower limit	Upper limit	Forest Plot
Elshafei (1998)	Combined	High School	Math	-0.148	15.79	-0.372	0.076	
Brokes (2010)	Combined	High School	Math	-0.059	14.99	-0.392	0.274	
Yancy (2012)	Posttest Content & Skills	Middle School	Math	0.345	14.65	-0.027	0.717	
Ridlon (2009b)	Posttest Content	Middle School	Math	0.490	12.97	-0.054	1.034	
Worry (2011)	Posttest Content	High School	Math	0.605	13.50	0.113	1.096	
Maree & Molepo (2005)	Posttest Content	High School	Math	0.834	15.95	0.636	1.031	
Ridlon (2009a)	Posttest Content	Middle School	Math	1.659	12.16	1.036	2.282	
				0.498		0.076	0.921	

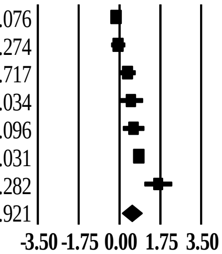


Figure 9. Forest plot of observed effects and the summary mean effect within math

The singleton dispersion of studies did not necessitate the creation of a stem and leaf display as the forest plot indicates absence of clustering.

Effect of PBL on social studies learning. Analysis of studies conducted in social studies (n = 10) generated a mean effect of g = 0.39, CI of g = 0.15 (LL) to 0.63 (UL), Z =3.18, p = .001. The forest plot of the effects of PBL on social studies achievement (Figure 10) shows a range of effects g = -0.22 to 1.34. A tau of 0.35 indicates that most effect sizes distributed g = ±0.69 about the mean effect.

Study Name	Outcome Type	Grade Level	Subject	Hedge's g	Relative weight	Lower limit	Upper limit	Forest Plot
Cicchino (2015)	Posttest Content	Middle School	Social Studies	-0.216	10.09	-0.525	0.092	
Tan et al. (2007)	Posttest Content & Skills	Middle School	Social Studies	-0.154	10.65	-0.408	0.101	
Mergendoller et al. (2000)	Posttest Content	High School	Social Studies	0.000	9.57	-0.355	0.355	
Mergendoller et al. (2006)	Posttest Content	High School	Social Studies	0.150	10.68	-0.101	0.402	
Finkelstein et al. (2011)	Combined	High School	Social Studies	0.179	11.98	0.113	0.246	
Parker et al. (2011)	Combined	High School	Social Studies	0.504	10.50	0.235	0.774	
Parker et al. (2013)	Combined	High School	Social Studies	0.511	10.67	0.258	0.763	
Wirkala & Kuhn (2011b)	Combined	Middle School	Social Studies	0.833	8.38	0.372	1.294	
Wirkala & Kuhn (2011a)	Combined	Middle School	Social Studies	1.083	7.84	0.573	1.593	
Hernandez-Ramos & De La Paz (2009)	Combined	Middle School	Social Studies	1.342	9.61	0.990	1.694	
				0.391		0.150	0.631	

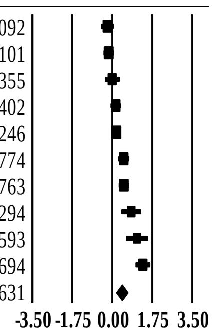


Figure 10. Forest plot of observed effects and the summary mean effect within social studies

Effects of PBL on Academic Achievement by Grade Level

An analysis of PBL effectiveness by grade level was conducted to determine the extent to which PBL differed in middle school and high school. Analysis at the middle school level (*n* = generated a mean effect of g = 0.65, CI of g = 0.33 (LL) to 0.97 (UL), Z = 3.98, p < .001. Observed effects range from g = -0.22 to 1.80, and T = 0.54, indicates that the distribution about the mean is g = ±1.07. Outcomes at the high school level generated a mean effect of g = 0.57, CI of g = 0.39 (LL) to 0.75 (UL), Z = 6.23, p < .001. Observed effects range from g = -0.77 to 2.29. A tau of 0.39 indicates that the effects (95%) distribute g = ±0.77 about the mean.

Effects of PBL on Academic Achievement by Location

Studies were categorized by location to investigate whether there were differences in effect sizes by location. Ten countries were represented in the sample. Two countries, Turkey (n = 8) and the United States (n = 18), had three or more studies. Analysis by location generated an effect size of g = 1.11, CI of g = 0.65 (LL) to 1.57 (UL), Z = 4.75, p < .001 for Turkey. A tau of 0.599 indicates that the effect sizes of the studies (95%) conducted in Turkey distribute g = ±1.17 about the mean. Studies conducted in the United States rendered a mean effect of g = 0.37, CI of g = 0.20 (LL) to 0.54 (UL), Z = 4.30, p < .001. A tau of 0.314 indicates that 95% of effect sizes disperse g = ±0.62 about the mean.

Effects of PBL on Academic Achievement by Ability Level

Five categories were coded for ability level, one of which was “not-specified” due to author non-report: low achieving (n = 4), average achieving (n = 3), high achieving (n = 2), mixed (low, average, and high, n = 14), and not specified (n = 13). Results indicate that the "not specified" category had a significantly higher summary mean, g = 1.01, CI of g = 0.64 (LL) to 1.37 (UL), Z = 5.39, $p < .001$, then the other four categories. Studies coded as mixed ability generated a statistically significant mean effect of g = 0.31, CI of g = 0.17 (LL) to 0.45 (UL), Z = 4.31, $p < .001$. Studies coded as low achieving also generated a statistically significant mean effect of g = 0.71, CI of g = 0.13 (LL) to 1.31 (UL), Z = 2.40, p = .02. The summary mean effects for high achieving (g = -0.02, p = .97) and average achieving (g = 0.47, p = .50) were all statistically non-significant, indicating that the true effect could be zero.

Effects of PBL by Inclusion of Reflective Assessment and/or Academic Discussion

Four categories of inclusion of reflective assessment (RA) and/or academic discussion (AD) were coded based on indication in the procedures/methods section of the study: neither RA nor AD, RA only, AD only, both RA and AD. The studies coded as "neither RA nor AD" (n = 7) rendered a summary mean effect of g = 0.68, CI of g = 0.25 (LL) to 1.16 (UL), Z = 3.08, p = .002. One study was coded as "RA only;" descriptively the study generated a statistically non-significant g = 0.46, $p < .341$. Results of the analysis of "AD only" (n = 13) rendered a mean effect of g = 0.87, CI of g = 0.48 (LL) to 1.25 (UL), Z = 4.42, $p < .001$. Results of the combination group, "RA and AD" (n = 10) generated a mean effect of g = 0.40, CI of g = 0.23 (LL) to 0.57 (UL), Z = 4.57, $p < .001$.

Effects of PBL by Facilitator Type and Experience

Studies were coded into four categories: teacher facilitator with no mention of training provided, teacher facilitator with brief training (or an instructional manual provided); teacher facilitator with extensive training and/or prior experience; or researcher facilitated (experience presumed). Results for "teacher facilitation with no mention of training provided" (n = 8) generated a statistically significant mean effect of g = 0.68, CI of g = 0.39 (LL) to 0.98 (UL), Z = 4.56, $p < .001$, disconfirming the null hypothesis that the true effect is zero. A tau of 0.34 indicates that most effect sizes distributed $g = \pm 0.73$ about the mean effect. Studies coded as "teacher facilitator with brief training (or an instructional manual provided)" (n = 7) rendered a statistically non-significant mean effect of g = 0.35, Z = 01.92, p = 0.60, indicating that the true mean could be zero. Studies coded as "teacher facilitator with extensive training and/or prior experience" (n = 13) rendered a mean effect of g = 0.50, CI of g = 0.25 (LL) to 0.76 (UL), Z = 03.87, $p < .001$, and T = 0.42 indicates that the distribution of these effect sizes about the mean is $g = \pm 0.82$. Studies coded as "researcher facilitated (experience presumed)" (n = 8) rendered a mean effect of g = 0.93, CI of g = 0.66 (LL) to 1.20 (UL), Z = 6.77, $p < .001$. A tau of 0.29 indicates that the effect sizes distributed about the mean $g = \pm 0.57$.

Test of Homogeneity

The test of heterogeneity for the summary mean effect across outcome types, academic subjects, and grade levels indicates that there is significant variance between studies, $Q(35)$ = 368.03, $p < .001$. I-squared indicates that 90.49% of the variance is unexplained in the model; thus warranting an investigation of moderator variables.

Borenstein et al. (2009) suggested a minimum of 10 studies per covariate and do not recommend meta-regression there is a small number of studies The moderators in this meta-analysis were categorical and a categorical variable typically describes a "set" of covariates (Borenstein et al., 2009). Thus, each covariate within a categorical variable set would require a minimum of ten studies each for the most reliable results. Majority of the moderators coded for this meta-analysis contained three of more covariates. After examining the number of cases per covariate for moderators with three or more covariates in the set, it was determined that meta-regression was not feasible, because one or more covariates within a moderator set had less than 10 studies. Although grade level contained two covariates, this moderator was eliminated as an option for meta-regression analysis because the summary mean effects for middle school $g = 0.65$ and high school $g = 0.57$ were relatively similar. Further, although the dispersion of observed effects in high school demonstrate a broader range, $g = -0.77$ to 2.29, compared to middle school, $g = -0.22$ to 1.80, this range represents the lower and upper boundary effect sizes of all included studies.

Publication Bias Analyses

Funnel Plot Analysis

A funnel plot analysis was conducted to visually analyze the possibility of publication bias. Figure 11 shows the funnel plot generated for the 48 effect sizes generated from 34 studies across six types of outcome measures, five academic subjects, and two grade levels (middle school and high school), accounting for independence of effect sizes. The high concentration of studies located in the upper left and lower right of the funnel suggests asymmetry; therefore, publication bias. However, the funnel plot also indicates that the imputed studies have a fairly high level of precision, likely due to the large number of studies that had sample sizes greater than 100 ($n = 20$ of 36 samples). Therefore, the asymmetry may also suggest true heterogeneity among studies (Higgins & Green, 2011; Terrin et al., 2005)

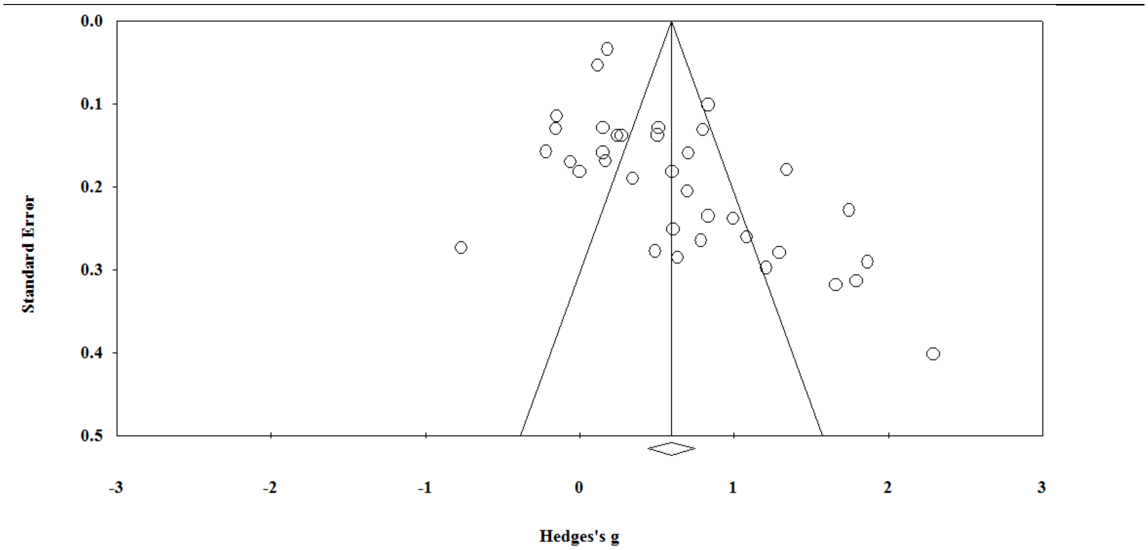


Figure 11. Funnel plot of potential publication bias.

Orwin's fail-safe *N*

As explained in Chapter 3, Orwin's fail-safe *N* is used to determine the number of additional studies, typically unpublished or non-retrieved studies, needed to render the summary mean effect size trivial (Orwin, 1983). The investigator determined that a small effect size, g = 0.20, (Cohen, 1992; Hattie, 2009) would be needed to render the summary mean effect trivial and assumed a g = 0.00 effect size in the unpublished or non-retrieved studies. Using the formula for Orwin's fail-safe *N* (see Chapter 3, "Publication bias analysis"), 21 additional studies would be needed to render the effect size trivial. An additional 21 studies is just over half of the total number of studies in the meta-analysis; thus it is possible that the summary mean effect is an overestimation of the true population effect.

Duval and Tweedie's *Trim and Fill*

Duval and Tweedie's (2000) *Trim and Fill* method (see Figure 12) calculated an adjusted summary mean, g = 0.50, to account for any publication bias. This adjusted effect size is medium-to-large effect for an educational intervention (Hattie, 2009) and a narrow 0.09 reduction from the primary summary mean effect of $g = 0.59$. Thus, based on the *Trim and Fill* method, the summary mean effect reported in the main analysis is relatively robust.

Chapter 5: Discussion

The purpose of this meta-analytic investigation was to explore the effects of PBL on academic achievement compared to traditional instruction in Grades 6-12 populations. Although primary and secondary research within secondary (6-12) contexts indicates that PBL is often superior to traditional, lecture-based instruction (Batdi, 2014; Mergendoller et al., 2006; Wirkala & Kuhn, 2011) and meta-analyses at the post-secondary level indicate that PBL is at par with or superior to traditional, lecture-based instruction (Dochy et al., 2003; Vernon & Blake, 1993; Walker & Leary, 2009), a synthesized and quantified exploration of the strength of relationship between PBL and academic achievement within middle school, junior high, and high school student populations was needed. After an extensive search for literature, 34 research articles were retained for inclusion in the meta-analysis, yielding 36 independent samples and 48 outcome effects. The purpose of this chapter is to summarize the seven guiding questions and two hypotheses that defined the scope of this study, discuss these findings in relationship to theory and previous research, identify and discuss limitations, and provide suggestions for further research.

Summary of Findings and Connection to Previous Research

Null Hypotheses Testing: Confirmative and Tentative Findings

Two null hypotheses were generated for this study in regard to the effect of PBL on academic achievement. Null Hypothesis 1 stated that there would be non-significant differences between PBL and traditional instruction on academic achievement on immediate, posttest achievement for both content acquisition and skills application tests. Null Hypothesis 2 stated that there would be non-significant differences between PBL and traditional instruction on academic achievement on retention achievement for content and skills application tests. First an overall, summary mean effect across outcome types, grade levels, and subject areas was generated to understand the general effect of PBL on academic achievement compared to traditional instruction. Then, separate analyses were conducted to identify the effect of PBL at post- and retention test. These effects were further explored by subject area, grade level, location of study, and ability level.

Overall mean effect. The investigator conducted a random effects meta-analysis to determine the overall effect of PBL on academic achievement compared to traditional, lecture-based instruction in Grades 6-12. The overall effect included five broad academic subject categories (science, math, social studies, ELA, and other/ electives) and was measured by content and/or skills posttests and retention tests. Results indicate that PBL has an overall statistically significant mean effect of $g = 0.59$ and narrow confidence interval of $g = 0.44$ at the lower limit (LL) and $g = 0.75$ at the upper limit (UL), meaning that the effect size is precise and is a large effect size for educational interventions (Hattie, 2009). By comparison, Vernon and Blake (1993) reported an overall effect of $d_w = -0.18$ in medical fields, slightly favoring traditional instruction. Walker and Leary (2009) reported an overall effect of $d_w = 0.13$ across academic disciplines in primarily post-secondary contexts favoring PBL. In a meta-analysis of Turkish master's theses and doctoral dissertations on the effects of PBL on academic achievement across academic disciplines, Batdi (2014) reported a very large effect size favoring PBL, $d = 1.30$.

An effect size of $g = 0.59$ is equivalent to a 22% percentile gain (Marzano Research, 2015). Put differently, students who learn in PBL classrooms will on average exceed (or outperform) 72% students who learn via traditional, lecture-based instruction.

The confidence interval of this mean, g = 0.44 at the lower limit and 0.75 at the upper limit, suggests that even at the low end, PBL students may outperform 67% of their peers in traditionally instructed classes. Publication bias analysis using the funnel plot and Orwin's (1983) fail-safe N indicate that there could be publication bias and that only 21 studies would be needed to render the effect size small and possibly trivial. However, Duval and Tweedie's (2000) *Trim and Fill* calculated an adjusted mean effect of g = 0.50, suggesting that the main analysis effect size of g = 0.59 is fairly robust.

Although an overall, summary mean effect is interesting it provides little information. Analysis of the effects of PBL on academic achievement by outcome type, academic subject, and grade level provides additional information.

Null hypothesis 1. There are statistically non-significant differences in individual academic achievement among PBL and traditionally instructed conditions in Grades 6-12 populations as measured by immediate content and/or skills posttests.

Results of the meta-analysis indicate that, overall, PBL students outperformed traditionally taught students on posttest achievement measuring content and skills, g = 0.62, $p < .001$, suggesting disconfirmation of the null hypothesis. Results for each of the outcome types were mixed. On acquisition of content posttests, g = 0.78, $p < .001$, and application of skills posttests, g = 0.24, p = .009, PBL students outperformed traditionally instructed students; thus, disconfirming the null hypothesis. By comparison, Dochy et al. (2003) reported an overall mean effect for acquisition knowledge tests of $d_w = -0.223$, indicating traditionally instructed students outperformed PBL students. On application of knowledge tests, Dochy et al. reported a mean effect of $d_w = +0.46$, a significantly larger effect than the result of this meta-analysis. There was a statistically non-significant

Z-value on tests that combined content and skills, $g = 0.16$, $p = .301$, indicating that the mean effect size could be zero. This statistically non-significant result for combined content and skills on a single measure is interesting in light of the statistically significant difference between conditions when content and skills were tested separately. Thus, disconfirmation of the null hypothesis is tentative.

Null hypothesis 2. There are statistically non-significant differences in individual academic achievement among PBL and traditionally instructed conditions in Grades 6-12 populations as measured by content and/or skills retention tests.

The summary mean effect for retention test achievement measuring both content and skills indicates that PBL conditions outperformed traditionally instructed conditions, $g = 0.60$, $p < .001$. However, the wide confidence interval of $g = 0.29$ (LL) to 0.90 (UL) indicates that this effect size is imprecise. Viewed individually, retention of content tests and retention of skills test were statistically significant and generated mean effects of $g = 0.87$, $p = .003$, $g = 0.75$, $p < .001$, respectively. However, there was a statistically non-significant result for retention of content and skills tests, $g = 0.43$, $p = .109$. Results of the effectiveness of PBL on retention of content and skills should be interpreted with caution, as there were no more than three studies per construct. Although Rosenthal (1995) remarked that a meta-analysis can "be applied to as few as two studies," he warned that doing so leads to "unstable" results (p. 185). Thus, Null Hypothesis 2 cannot be definitively disconfirmed. More research on retention of knowledge and application of skills in

Effects of PBL by academic subject. Five academic subjects were coded for the meta-analysis: science ($n = 16$), math ($n = 7$), social studies ($n = 10$), ELA ($n = 1$), and elective/other (n = 2). Main analyses were conducted on science, math, and social studies due to the insufficient number of studies to conduct analyses in ELA and the elective/other category.

Science. The analysis on science achievement yielded a statistically significant and larger effect size than math or social studies, g = 0.83, indicating that PBL students outperformed traditionally instructed students. By comparison, Walker and Leary (2009) reported nearly negligible effects of PBL on science achievement, d_w = 0.06, in post-secondary, non-medical subjects and Badti (2014) reported a very large effect size of d = 1.32. Based on the recommendations of Cohen (1992) and Hattie (2009), this effect size is large, especially for educational interventions. This effect is equivalent to a 30% percentile gain, meaning that on average, students who learn science in a PBL context will outperform 80% of traditionally instructed students. However, results also indicated that the distribution about the mean was g = ±1.53, a large range. Of note, seven of the 16 studies were conducted in Turkey; as will be discussed, the studies conducted in Turkey yielded the highest mean effects.

Math. The analysis on math achievement resulted in a statistically significant mean effect of g = 0.50, a medium-high effect size (Hattie, 2009), indicating that PBL students outperformed traditionally instructed students. The distribution of mean effects, similar to science, was wide g = ±1.04. By comparison, the master's theses and dissertations meta-analyzed by Batdi (2014) yielded a large effect size and somewhat comparable effect size of d = 0.79. An effect size of g = 0.50 translates to 19% percentile gain or that PBL students will on average outperform students instructed in a traditional classroom by 69%.

Social studies. Half of the studies conducted in the social studies, which consisted of geography, history, and economics courses, indicated small effects. The analysis for all studies generated a statistically significant summary mean effect of g = 0.39, a medium effect (Hattie, 2009). Although the mean effect for social studies was smaller than science and math, the distribution of means was narrower, though significantly varied, g = ±0.69. An effect size of g = 0.39 translates to a 15% percentile gain, or that PBL students will on average outperform students instructed in a traditional classroom by 65%. The results of the analysis of PBL on achievement in social studies rendered a similar result to that reported by Walker and Leary (2009), d_w = 0.30 (social sciences), but significantly smaller than Batdi's (2014) reporting of a very large effect size in social sciences of d = 1.88.

In analyzing the studies that comprised the analysis of social studies achievement in this meta-analysis, there were noteworthy observations. The Finkelstein et al. (2009) and Mergendoller et al. (2000, 2006) studies were conducted using a program curriculum created by Buck Institute for Education (BIE). In particular, the Finkelstein et al. study was conducted specifically for the purpose of testing the efficacy of the BIE Problem Based Economics curriculum in a large scale implementation. In all three studies, the researchers had little control over the fidelity of teacher implementation (implementation fidelity was "confirmed" by teacher report), despite extensive professional development. The three aforementioned studies rendered small effect sizes. Another study, conducted by Cicchino (2015), used a teacher and researcher designed outcome measure with only five questions (no test-retest reliability was reported). Thus, it is possible that the effect size is suppressed by these studies. However, the absence of teacher interference demonstrated in the Finkelstein et al. and Mergendoller et al. studies may more realistically produce the true effects of PBL when teachers are left to implement a program with little supervision (though extensive support was offered the intervention teachers throughout the course of each study, Finkelstein et al. reported that few teacher took advantage of it).

Effects of PBL by grade level. The analysis of the effect of PBL on academic achievement by grade level was statistically significant for both middle and high school. The effect of PBL was larger in middle school, $g = 0.65$, than high school, $g = 0.57$; however, both effect sizes are considered large for an educational intervention (Hattie, 2009). The range of effects in both middle ($g = -0.22$ to 1.80) and high school ($g = -0.76$ to 2.29) were wide.

Effects of PBL by location. Ten countries were represented in the included studies. Turkey and the United States were the only two countries with three or more studies, providing a meaningful analysis. Studies conducted in Turkey produced the largest summary mean effect of $g = 1.11$, $p < .001$ compared to those conducted in the United States, $g = 0.37$, $p < .001$. Although the summary mean effect for the studies conducted in the United States is smaller, it should be highlighted that the dispersion of effects about the mean for these studies was calculated as $g = \pm 0.62$, compared to a dispersion in the Turkish studies of $g = \pm 1.17$. This difference in dispersion may suggest that there was more consistency in the studies conducted in the U.S. than in Turkey.

Effect of PBL by ability level. One of the questions under exploration was to what extent, if any, is the effect of PBL moderated by ability level? Five categories were coded for ability level: low achieving, average achieving, high achieving, mixed (low, average, and high), and not-specified. While this question was sound both theoretically and practically, as teachers and administrators are charged with implementing instructional practices that benefit all students (Banks, 1995, 2008; Ridlon, 2009), exploration of ability level was hindered by primary authors' lack of reporting the samples' ability or achievement level. Of the retained studies, one-third (n = 13) were marked as "not specified" during the coding process due to non-report. Results indicated that the non-specified category had a significantly higher summary mean, $g = 1.01$, $p < .001$ than the other four categories. Descriptively, of the 13 studies that comprised the "not specified" category, six were conducted in Turkey, four in the United States, two in Nigeria, and one in Pakistan. Turkish studies comprised of eight total studies in the meta-analysis, and with a large mean effect by location, $g = 1.11$, it is likely that the Turkish studies influenced the effect size of the "not specified" category. The effect size of PBL achievement for conditions reported as mixed ability (n = 14) was of medium-small effect and statistically significant, $g = 0.31$, $p < .001$. The effect of PBL on low achieving students was larger, $g = 0.71$, $p = .02$, but a wide confidence interval renders this effect size imprecise. The summary mean effects for high achieving ($g = -0.02$, $p = .97$) and average achieving ($g = 0.47$, $p = .50$) were all statistically non-significant, indicating that the true effect could be zero. However, these results should be interpreted with caution due to the limited number of studies (2 and 3) in each of these categories.

Effect of PBL by inclusion of reflective assessment and/or academic discussion. Four categories of inclusion of reflective assessment (RA) and/or academic discussion (AD) were coded based on indication in the procedures/methods section of the study. Many studies mentioned the use of reflection and/or discussion in the theory sections of the research report, but not in the procedures. The absence or inclusion of RA and/or AD may be informative in terms of distinguishing more effective implementation of and practices within PBL. Meta-analysi of this moderator indicated variability among studies. The studies coded as "neither RA nor AD" (n = 7) rendered a summary mean effect of g = 0.68, Z = 3.08, p = .002. One study was coded as "RA only." Descriptively the study generated a statistically non-significant g = 0.46, p < .341. Results of the analysis of "AD only" (n = 13) rendered a mean effect of g = 0.87, Z = 4.42, p < .001. Results of the combination group, "RA and AD" (n = 10) generated a mean effect of g = 0.40, Z = 4.57, p < .001.

Reflective assessment and academic discussion are considered salient features of PBL (Barron et al., 1998; Hmelo-Silver, 2004; Hmelo-Silver & DeSimone, 2013; Mergendoller et al., 2006; Schmidt et al., 2007). All but five of the studies included in the meta-analysis (Afolabi & Akinbobola, 2009; Baş,, 2011; Maree & Molepo, 2005; Vilardi, 2013; Worry, 2011) mentioned either reflective assessment or academic discussion in either the theory section of the report or in the conclusion, acknowledging the role of both RA and/or AD in PBL effectiveness. Two additional studies (Elshafei, 1998; Yancy, 2012) were coded as "neither RA nor AD" due to lack of explicit connection of use within the intervention group, although both of these studies identified reflection and discussion as salient features of PBL. The differences in effect sizes between use of AD (g = 0.87) and RA and AD (g = 0.40) are interesting. Both elicit forms of reflective practice, one in an individual manner and the other socially (Brown, 1977; Costa & O'Leary, 1992, Flavell, 1979; Schraw, 1998). Thus, the smaller effect size of the combination of RA and AD compared to AD only may be worth further exploration. However, it should also be highlighted that of the 13 studies that comprised the “AD only” category, six of those studies were conducted in Turkey. As mentioned, there were a total of eight Turkish studies in the meta-analysis as a whole and these studies rendered a very large effect size.

Effect of PBL by facilitator type and experience. Studies were coded into four categories by facilitator type and experience: teacher facilitator with no mention of training provided, teacher facilitator with brief training (or an instructional manual provided); teacher facilitator with extensive training and/or prior experience; or researcher facilitated (experience presumed). Meta-analysis by this moderator indicated some variance among effect sizes. Results for “teacher facilitation with no mention of training provided” (n = 8) generated a mean effect of g = 0.68, Z = 4.56, p < .001. Studies coded as “teacher facilitator with brief training (or an instructional manual provided)” (n = 7) rendered a statistically non-significant mean effect of g = 0.35, Z = 01.92, p = 0.60, indicating that the true mean could be zero. Notably, there was only one study for this variable, therefore any conclusions regarding the effect of RA on PBL achievement is unwarranted. Studies coded as “teacher facilitator with extensive training and/or prior experience” (n = 13) rendered a mean effect of g = 0.50, Z = 03.87, p < .001. Studies coded as “researcher facilitated (experience presumed)” (n = 8) rendered a mean effect of g = 0.93, Z = 6.77, p < .001.

Although inexperience with PBL may create initial frustrations for teacher(s) and students (Rogers, Cross, Gresalfi, Trauth-Nare, & Buck, 2009; Scott, 1994; Parker et al., 2011; Tan et al., 2007), the implementation process may create a learning environment conducive to learning. The role of the facilitator is to guide the learning (Barron &

Darling-Hammond, 2008; Hmelo-Silver, 2004; Walker & Leary, 2009; Wirkala & Kuhn, 2011; Zohar & Ben David, 2008) and even with initial struggles, teacher and students can learn together. Piaget (1967) posited that peer-to-peer interactions have important implications for collaborative learning methods. Namely, Piaget asserted that adult-to-child relationships create situations in which the child (student) is more likely to comply with the adult's (teacher's) thinking and ways of doing. However, in peer-to-peer interactions, students are "more likely to develop cognitively in contexts in which peers have equal power and all have opportunities to influence one another" (O'Donnell & Hmelo-Silver, 2013, p. 8).

When a novice PBL teacher enters into the PBL experience as a learner not only of the process, but also the problem or project, it is possible that a "peer-to-peer" relationship is established, and cognitive conflict, resolution, and growth are encouraged. Likewise, experienced teachers may bring to each new PBL experience insights on how to not only encourage peer-to-peer conflict so as to stimulate cognitive conflict, but also rediscover anew (from each new group of students) approaches to the problem or project.

It is noteworthy that the "researcher as facilitator (experience assumed)" category yielded the largest effect size of $g = 0.93$, $Z = 6.77$, $p < .001$, suggesting that either the novelty of an outside expert/facilitator or the skills that individual brings to the experience may impact the learning experience of the students. Further, in each of the studies included in this meta-analysis, the PBL curriculum was either created by the researcher, co-created with the teachers with the guidance of university professionals, or developed by a professional organization (such as Buck Institute for Education). None of the studies investigated the effects of teacher created, especially novice-teacher created PBL units.

Limitations of the Study

There were several limitations in this study. One limitation was selecting only quasi-experimental and classroom-based experimental studies to include in this meta-analysis. Given the nature of classrooms as dynamic environments and variation among teaching styles and enthusiasm, other confounds could explain the results (Finkelstein et al., 2011; Slavin, 2008). Further, since teachers volunteer to open their classrooms for such studies, one cannot rule out that the types of teachers who volunteer for a study or are interested in implementing PBL are not somehow fundamentally different than those who do not volunteer (either for the study or to try PBL) (Slavin, 2008).

Another limitation is the focus on the effectiveness of PBL in Grades 6-12. The decision to limit the scope of research to Grades 6-12 is that secondary teachers often want research related to their grade levels, arguing that what works in elementary school does not necessarily work in middle and high school due to the assumed expectation of content coverage over exploration and extended learning (Gallagher & Stepien, 1996). Thus, generalizability of results is limited to Grades 6-12 populations.

The limited number of studies included in the meta-analysis created another limitation in that exploration of possible covariates was unfeasible. Field (2009) and Borenstein et al. (2009) recommend at least 10 studies per covariate, which was unattainable and restricted moderator analyses.

Inclusion of only English translated reports was a delimitation in order to properly analyze the operationalized definition of PBL, procedures, and statistics reported in the study. However, this delimitation, simultaneously created a limitation. The meta-analysis conducted by Batdi (2014) provided 17 potential secondary level studies for inclusion; however, only two of those studies were translated into English *and* conducted in a 6-12 population.

The subjective nature of meta-analysis could also be a limitation (Rosenthal & DiMatteo, 2001). Much of the coding, particularly of the inclusion of reflective assessment and/or academic discussion, but also ability/achievement levels and facilitator experience, had to be gleaned from sometimes vague information. Although the investigator could have made logical assumptions about inclusion of reflective assessment or academic discussion based on references in the literature review portion of the report, or possibly recoded "not specified" ability type to "mixed ability," such assumptions and designations had the potential for error and subsequent inaccurate conclusions.

Recommendations for Further Research

The effectiveness of PBL in science, math, and social studies are reported in both this meta-analysis and in primary studies in Grades 6-12 contexts (Jewett & Kuhn, 2015; Ridlon, 2009; Wirkala & Kuhn, 2011; Wong, 2012). However, more research on the effectiveness of PBL on academic achievement is needed in general and in the humanities, specifically ELA, in particular. Conclusions drawn in this meta-analysis were made based a small amount of literature. To substantiate results, more research is needed across disciplines. Further, the lack of studies in ELA contexts limits any definitive conclusions about the appropriateness and effectiveness of PBL in a heavily literature and writing based discipline. More research is also needed in the social sciences (social studies). There were 10 studies in this meta-analysis categorized as "social studies," an umbrella term for an array of disciplines (history, geography, sociology, government/ politics, and economics). Majority were conducted applied social sciences (economics, $n = 3$; government, $n = 2$; and sociology, $n = 2$). The use of PBL in social science disciplines that tend to be more application-based may be more easily adaptable to PBL. Therefore, more research on the effectiveness of PBL is needed in disciplines such as history where the assumed expectation of content coverage over exploration and extended learning (process-oriented) is typical (Gallagher & Stepien, 1996), in order to demonstrate to history teachers that PBL can address both content and process (Parker et al., 2011).

Research is also needed to explore the differences in PBL group dynamics, teacher practices, and testing within Turkish contexts compared to the United States. The very large effect sizes demonstrated in the Turkish studies provides an impetus to identify and possibly replicate, if possible, PBL practices used in Turkey.

Further, research is needed to better understand the effects of reflective assessment and academic discussion on PBL effectiveness. Although reflective assessment and academic discussion are considered salient features of PBL (Barron et al., 1998; Hmelo-Silver, 2004; Hmelo-Silver & DeSimone, 2013; Mergendoller et al., 2006; Schmidt et al., 2007) there is a lack of research quantifying the specific contributions of reflective assessment and academic discussion on PBL outcome achievement. Wirkala and Kuhn (2011), Pease and Kuhn (2011), and Jewett and Kuhn (2015) studied the effects of discussion on PBL effectiveness and suggested that it may not be discussion, but the problem itself that accounts for PBL effectiveness. Wirkala and Kuhn, however, attributed reflection as a possible explanation for students' ability to "transfer learning to new settings and events" (p. 1184). Thus, research combining and isolating reflective assessment and academic discussion may further understanding of the extent to which, if any, each impacts PBL outcome achievement.

Summary

Problem- and project-based learning appear to be viable and effective instructional approaches in Grades 6-12 contexts across academic subjects. The results in this meta-analysis indicate that overall, PBL students outperformed traditionally instructed students, $g = 0.54$, on content and skills exams across academic subject types and grade levels. Analysis of the funnel plot suggests publication bias; however, an adjustment of the mean effect using Duval and Tweedie's (2000) *Trim and Fill* rendered a similar summary effect of $g = 0.50$. Although the mean summary effect is verily robust, effect sizes varied depending on subject area and specific types of outcome measures. The test of homogeneity indicated that 90.49% of the variance between studies was unexplained. An insufficient number of studies rendered meta-regression unfeasible; therefore, hindering exploration of possible explanations for this variance.

Since the enactment of NCLB in 2001 and the subsequent high stakes testing movement, standards' movements, and now teacher evaluation systems there has been an increasing impetus for teachers to use instructional "best practices" that promote the learning of all students. Contemporary iterations of PBL, especially in K-12 contexts, promote scaffolding, mini-lectures, formative, reflective assessment opportunities, discussion, and explicit reference to outcome objectives (making learning targets clear) in order to connect doing with learning (Barron et al., 1998; Hmelo-Silver, 2004;

Mergendoller et al., 2006; Parker et al., 2011, 2013; Wirkala & Kuhn, 2011). Further, coupled with these "best practices," teachers at the secondary level are charged with both providing students reasons for learning (application) and ensuring that they learned (high stakes testing) (NCLB, 2003; Common Core State Standards Initiative, 2012). These expectations may account for the large effect sizes in content mastery as measured by content posttests in secondary compared to those at the post-secondary level.

References

Afolabi, F., & Akinbobola, A. O. (2009). Constructivist problem based learning technique and the academic achievement of physics students with low ability level in Nigerian secondary school. *Eurasian Journal of Physics and Chemistry Education, 1*(1), 45-51. Retrieved from http://eurasianjournals.com/ index.php/ejpce/article/view/34/59

Anderson, K. T., Zuiker, S. J, Taasoobshirazi, G., & Hickey, D. T. (2007). Classroom discourse as a tool to enhance formative assessment and practise in Science. *International Journal of Science Education, 29*, 1721-1744. http://dx.doi.org//10.1080/09500690701217295

Anyafulude, J. C. (2013). Effects of problem-based and discovery-based instructional strategies on students' academic achievement in chemistry. *Journal of Educational and Social Research*, *3*(6), 105-112. http://www.doi.org/10.5901/ JESR.2013.v3n6p105

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685-730. http://dx.doi.org/10.3102/00028312040003685

Applefield, J. M., Huber, R., & Moallem, M. (2001). Constructivism in theory and practice: Toward a better understanding. *The High School Journal, 84*(2), 35-53. Retrieved from http://muse.jhu.edu/journals/high_school_journal/

Araz, G., & Sungur, S. (2007). Effectiveness of problem-based learning on academic performance in genetics. *Biochemistry and Molecular Biology Education, 35*(6), 448-451. doi:10.1002/bmb.97

Arends, R. I., & Kilcher, A. (2010). *Teaching for student learning: Becoming an accomplished teacher.* New York, NY: Routledge.

Bagley, W. C. (1921). Dangers and difficulties of the project method and how to overcome them—A symposium, II. Projects and purposes in teaching and in learning. *Teachers College Record*, *22*(4). 288-297. Retrieved from http://www.books.google.com/

Bagley, W. C. (1939). The significance of the essentialist movement in educational theory. *The Classical Journal*, *34*(4), 326-344. Retrieved from http://www.jstor.org/stable/3290875

Bandura, A. (1977). *Social learning theory. Englewood Cliffs,* NJ: Prentice Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*(1), 1-26. http://dx.doi.org/10.1146/annurev.psych.52.1.1

Banks, J. A. (1995). Multicultural education and curriculum transformation. *The Journal of Negro Education*, *64*, 390-400. http://dx.doi.org/10.2307/2967262

Banks, J. A. (2008). Diversity, group identity, and citizenship education in a global age. *Educational Researcher, 37*, 129-139. http://dx.doi.org/10.3102/0013189 X08317501

Barron, B., & Darling-Hammond, L. (2008). *Teaching for meaningful learning: A review of research on inquiry-based and cooperative learning,* [PDF]. R. Furger (Ed.), The George Lucas Educational Foundation, Edutopia. Retrieved from http://www.edutopia.org/pdfs/edutopia-teaching-for-meaningful-learning.pdf

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., … the Cognition and Technology Group at Vanderbilt. (1998). Doing with understanding: Lessons from research on problem- and project- based learning. *The Journal of Learning Sciences, 7*(3/4), 271-311.

Baş, G. (2011). Investigating the effects of project-based learning on students' academic achievement and attitudes towards English lessons. *The Online Journal of New Horizons in Education, 1*(4), 2011. Retrieved from http://www.tojned.net/pdf/tojnedv01i04-01.pdf

Batdi, V. (2014). A meta-analysis study comparing problem based learning with traditional instruction. *Electronic Journal of Social Sciences, 13*(51), 346-364. Retrieved from http://dergipark.ulakbim.gov.tr/esosder/article/viewFile/5000076864/5000070920

Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2011). Problem-based learning and argumentation: testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science, 39*(5), 667-694. http://dx.doi.org/ 10.1007/s11251-010-9148-z

Berliner, D. C. (2002). Comment: Educational research: The hardest science of all. *Educational Researcher, 31*(8), 18-20. http://dx.doi.org/10.3102/0013189X031008018

Biostat. (2015). *Comprehensive Meta-Analysis, Version* 3. [Software]. Englewood, NJ. Retrieved from http://www.meta-analysis.com/index.php

Black, P., & William, D. (2009). Developing a theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. http://dx.doi.org/10.1007/s11092-008-9068-5

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Hoboken, NJ: John Wiley & Sons.

Boud, D., Keogh, R., & Walker, D. (1985). *Reflection: Turning experience into learning.* New York, NY: Nichols Publishing Company.

Brooks, J. G., & Brooks, M. G. (1993). *In search of understanding: The case for constructivist classrooms.* Alexandria, VA: Association of Supervision and Curriculum Development.

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brown, A. L. (1977). *Knowing when, where, and how to remember: A problem of metacognition. Technical Report No. 47.* Washington, DC: National Institute of Education. Retrieved from ERIC database (ED146562).

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*(2), 141-178. http://dx.doi.org/10.1207/s15327809jls0202_2

Bruffee, K. A. (1999). *Collaborative learning: Higher education, interdependence, and the authority of knowledge* (2nd ed.). Baltimore, MD: John Hopkins University Press.

Bruner, J. (1985). Vygotsky: A historical and conceptual perspective. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (pp. 21-34). Cambridge, UK: Cambridge University Press.

Bruner, J. (1996). *The culture of education.* Cambridge, MA: Harvard University Press.

Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Bruner, J. S. (1966). *Toward a theory of instruction.* Cambridge, MA: Belknap Press of Harvard University Press.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston, MA: Houghton Mifflin Company.

Chang, C. -Y. (2001). Comparing the impact of a problem-based computer-assisted instruction and the direct-interactive teaching method on student science achievement. *Journal of Science Education and Technology, 10*(2), 147-153. http://dx.doi.org/10.1023/A:1009469014218

Cicchino, M. I. (2015). Using game-based learning to foster critical thinking in student discourse. *Interdisciplinary Journal of Problem-based Learning, 9*(2), http://dx.doi.org/10.7771/1541-5015.1481

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24*(2), 205-249. http://dx.doi.org/10.1007/ s10648-011-9191-6

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. http://dx.doi.org/10.1037/0033-2009.112.1.155

Common Core State Standards Initiative. (2012). *Frequently asked questions: Implementation and future work.* Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers. Retrieved from http://www.corestandards.org/resources/frequently-asked-questions

Costa, A. L., & O'Leary, P. W. (1992). Co-cognition: The cooperative development of the intellect. In N. Davidson & T. Worsham (Eds.), *Enhancing thinking through cooperative learning* (pp. 41-65). New York, NY: Teachers College Press.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Dewey, J. (1899). *My pedagogic creed.* Teachers Manuals, *25*, 3-18. Retrieved from http://www.books.google.com/

Dewey, J. (1922). *Democracy and education: An introduction to the philosophy of education.* New York, NY: The McMillan Company. Retrieved from http://www.books. google.com

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process.* Boston, MA: D.C. Heath and Company.

Dewey, J. (1938). *Experience and education.* New York, NY: Simon & Schuster.

Dewey, J. (1991). The challenge of democracy to education. In J. A. Boydston, (Ed.), *John Dewey: The later works*, *1925-1953* (Vol. 11, pp. 181-190). Carbondale, IL: Southern Illinois University Press. (Reprinted from *Progressive Education*, *14,* February 1937, pp. 79-85].

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, *20*(4), 391-409. http://dx.doi.org/10.1007/s10648-008-9083-6

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, *13*(5), 533-568. http://dx.doi.org/10.1016/S0959-4752(02)00025-7

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-465.

Elizabeth, T., Ross Anderson, T. L., Snow, E., & Selman, R. L. (2012). Academic discussions: An analysis of instructional discourse and an argument for an integrative assessment framework. *American Educational Research Journal, 49*(6), 1214-1250. http://dx.doi.org/103102/0002831212456066

Ellis, A., Bond, J., & Denton, D. (2012). An analytical literature review of the effects of metacognitive teaching strategies in primary and secondary student populations. *Asia Pacific Journal of Educational Development, 1*(1), 9-23. http://dx.doi.org/ 10.6228/APJED.01.01.02

Ellis, A. K. (2001). *Teaching, learning, & assessment together: The reflective classroom.* New York, NY: Eye on Education.

Elshafei, D. L. (1998). *A comparison of problem-based and traditional learning in Algebra II*. (Unpublished doctoral dissertations). Indiana University, Bloomington, IN. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 9919416)

Ertmer, P. A., & Simons, K. D. (2006). Jumping the PBL implementation hurdle: Supporting the efforts of K-12 teachers. *Interdisciplinary Journal of Problem-Based Learning, 1*(1), 40-54. http://dx.doi.org/10.7771/1541-5015.1005

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles, CA: Sage Publications.

Field, A. P. (2003). The problems in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, *2*(2), 105-124. http://dx.doi.org/10.1207/S15328031US0202_02

Field, A. P., & Gillett, R. (2010). How to do meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*(3), 665-694. http://dx.doi.org/10.1348/000711010X502733

Finkelstein, N., Hanson, T., Huang, C. W., Hirschman, B., & Huang, M. (2011). *Effects of Problem Based Economics on high school economics instruction.* (NCEE 2010-4022rev) [PDF]. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/edlabs/regions/west/pdf/REL_20104022.pdf

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-236). Hillsdale, NJ: Lawrence Erlbaum Associates.

Flavell, J. H. (1979). Metacognition and cognitive monitory: A new area of cognitive developmental inquiry. *American Psychologist, 34*(10), 905-911. http://dx.doi.org/10.1037/0003-066X.34.10.906

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 21-29). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Forman, E. A. & Cazden, C. B. (1985). Exploring Vygotskian perspectives in education: The cognitive value of peer interaction. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (pp. 323-347). Cambridge, UK: Cambridge University Press.

Gall, M. D., Gall. J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Boston, MA: Pearson Education, Inc.

Gallagher, S. A., & Stepien, W. J. (1996). Content acquisition in problem-based learning: Depth versus breadth in American Studies: *Journal for the Education of the Gifted*, *19*(3), 257-275. http://dx.doi.org/ 10.1177/016235329601900302

Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*(1), 27-61. http://dx.doi.org/ 10.3102/00346543075 001027

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8. http://dx.doi.org/10.3102/0013189X005010003

Graaff, E. de, & Kolmos, A. (2007). History of problem-based and project-based learning. In E. de Graff & A. Kolmos (Eds.), *Management of change: Implementation of problem-based and project-based learning in engineering* (pp. 1-8). Rotterdam, The Netherlands: Sense Publishers.

Grant, M. M., & Hill, J. R. (2006). Weighing the risks with the rewards: Implementing student-centered pedagogy within high stakes testing. In R. Lambert & C. McCarthy (Eds.), *Understanding teacher stress in an age of accountability* (pp. 19-42). Greenwich, CT: Information Age Publishing.

Hanney, R., & Savin-Baden, M. (2013). The problem of projects: Understanding the theoretical underpinnings of project-led PBL. *London Review of Education, 11*(1), 7-19. http://dx.doi.org/10.1080/14748460.2012.761816

Hass, M. (2005). Teaching methods for secondary algebra: A meta-analysis of finding. *NASSP Bulletin, 89*(642), 24-48. http://dx.doi.org/10.1177/019263650508964204

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Routledge.

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions* [Version 5.1.0]. Retrieved from http://handbook.cochrane.org/

Hirsch, Jr., E. D. (1996). *The schools we need and why we don't have them.* New York, NY: Anchor Books.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235-266. http://dx.doi.org/10.1023/

B:EDPR.0000034022.16470.f31

Hmelo-Silver, C. E., & DeSimone, C. (2013). Problem-based learning: An instructional model of collaborative learning. In C. E. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, & A. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 370-385). New York, NY: Routledge.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*(4), 275-292. http://dx.doi.org/10.1111/1468-2389.00156

Igel, C., & Apthorp, H. (2015). *The effect of cooperative instruction on academic performance: A meta-analysis of findings from the past fifteen years of primary research on K-12 populations.* Manuscript submitted for publication.

Jacobs, L. C. (1991). *Test reliability*. IUB Evaluation Services & Testing, Indiana University. Retrieved from http://www.indiana.edu/~best/test_reliability.shtml.

Jewett, E., & Kuhn, D. (2015). *Social science problem-based learning as a tool in developing scientific thinking skills in low-achieving students.* Manuscript submitted for publication.

Johnson, D. W., & Johnson, R. T. (1992). Encouraging thinking through constructive controversy. In N. Davidson & T. Worsham (Eds.), *Enhancing thinking through cooperative learning* (pp. 120-137). New York, NY: Teachers College Press.

Johnson, D. W., Johnson, R. T., Roy, P., & Zaidman, B. (1985). Oral interaction in cooperative learning groups: Speaking, listening, and the nature of statements made by high-, medium-, and low-achieving students. *The Journal of Psychology, 119*(4), 303-321. http://dx.doi.org/10.1080/00223980.1985.10542903

Katz, L., & Chard, S. C. (2000). *Engaging children's minds: The project approach* (2nd ed.). Stamford, CT: Ablex Publishing Corporation.

Kilpatrick, W. H. (1918). *The project method: The uses of the purposeful act in the educative process.* New York, NY: Teachers College, Columbia University.

Kilpatrick, W. H. (1921). Dangers and difficulties of the project method and how to overcome them—A symposium, I. Introductory statement: Definition of terms. *Teachers College Record, 22*(4), 283-288. Retrieved from http://www.books. google.com/

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*(2), 75-86. http://dx.doi.org/10.1207/s15326985ep4102_1

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In T. Lickona (Ed.), *Moral development and behavior: Theory, research, and social issues* (31-53). New York, NY: Holt, Rinehart and Winston.

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher, 44*(1), 46-53. http://dx.doi.org/10.3102/0013189X15569530

Kuşdemir, M., Ay, Y., & Tüysüz, C. (2013). An analysis of the effect of problem based learning model on the 10th grade students' achievement, attitude and motivation in the unit of "mixtures." *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education, 7*(2), 195-224. Retrieved from http://www.nef.balikesir.edu.tr/~dergi/makaleler/yayinda/15/EFMED_KME141.pdf (Google Translate)

Larmer, J. (2013). Project based learning vs. problem based vs. XBL. [Blog]. *Experts & NewBIEs, Bloggers on Project Based Learning.* Retrieved from http://biepbl.blogspot.com/2013/11/problem-based-vs-project-based-learning.html

Larmer, J., & Mergendoller, J. R. (2012). *8 essentials for project-based learning* [PDF]. Buck Institute for Education. Retrieved from http://images.bie. org/ uploads/useful_stuff/8_Essentials_article_small_file_size_ Oct2012version.pdf

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*(12), 1181-1209. http://dx.doi.org/ 10.1037/0003-066X.48.12.1181

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

Maree, J. G., & Molepo, J. M. (2005). Achievement in mathematics in grades 9 and 11 in Limpopo Province South Africa: Introduction of a problem based approach. *Psychological Reports*, *97*(3), 732-736. http://dx.doi.org/10.2466/PR0.97.7.732-736

Marzano Research. (2015). *Conversion of effect size to percentile gain.* [PDF]. Retrieved from http://www.marzanoresearch.com/reproducibles/highly_engaged

McCombs, B. (2010). Learner-centered practices: Providing the context for positive learner development, motivation, and achievement. In J. L. Meece, & J. S. Eccles (Eds.), *Handbook of research on schools, schooling, and human development* (pp. 60-74). New York, NY: Routledge.

McMillan, J. H., & Schumacher, S. (1997). *Research in education: A conceptual introduction* (2nd ed.). New York, NY: Longman, Addison-Wesley Educational Publishers Inc.

Mergendoller, J. R., Maxwell, N. L., & Bellisimo, Y. (2000). Comparing problem-based learning and traditional instruction in high school economics. *The Journal of Educational Research, 93*(6), 374-382. http://dx.doi.org/10.1080/0022067 0009598732

Mergendoller, J. R., Maxwell, N. L., & Bellisimo, Y. (2006). The effectiveness of problem-based instruction: A comparative study of instructional methods and student characteristics. *Interdisciplinary Journal of Problem-based Learning, 1*(2), 49-69. http://dx.doi.org/10.7771/1541-5015.1026

Mioduser, D., & Betzer, N. (2007). The contribution of project-based-learning to high-achievers' acquisition of technological knowledge and skills. *International Journal of Technology and Design Education, 18*(1), 59-77. http://dx.doi.org/ 10.1007/s10798-006-9010-4

National Governors Association Center for Best Practices, Council of Chief State School Officers (NGACCSSO). (2010*). Common core standards for English language arts and literacy in history/social studies, science, and technical subjects.* [PDF]. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

No Child Left Behind Act of 2001 (NCLB). 20 U.S.C.A. § 6301 *et seq.* (West 2003).

Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes*, *35*(2), 135-198. http://dx.doi.org/10.1207/S15326950 DP3502_3

O'Donnell, A., & Hmelo-Silver, C. A. (2013). Introduction: What is Collaborative learning? In C. E. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, & A. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 1-15). New York, NY: Routledge.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics, 8*(2), 157-159. http://dx.doi.org/10.3102/ 10769986008002157

Palmer, P. J. (1993). *To know as we are known: Education as a spiritual journey*. New York, NY: Harper Collins.

Parker, W., Mosborg, S., Bransford, J., Vye, N., Wilkerson, J., & Abbott, R. (2011). Rethinking advanced high school coursework: Tackling the depth/breadth tension in the AP US Government and Politics course. *Journal of Curriculum Studies, 43*(4), 533-559. http://dx.doi.org/10.1080/00220272.2011.584561.

Parker, W. C., Lo, J., Yeo, A. J., Valencia, S. W., Nguyen, D., Abbott, R. D.,…Vye, N. J. (2013). Beyond breadth-speed test: Toward deeper knowing and engagement in and advanced placement course. *American Educational Research Journal, 50*(6), 1424-1459. http://dx.doi.org/ 10.3102/0002831213504237

Partnership for 21st Century Skills. (2011). *Framework for 21st century skills.* Retrieved from http://www.p21.org/our-work/p21-framework

Pease, M. A., & Kuhn, D. (2011). Experimental investigation of the effectiveness of problem-based learning. *Science Education, 95*(1), 57-86. http://dx.doi.org/10.1002/sce.20412

Phillips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Researcher*, *24*(7), 5-12. http://dx.doi.org/10.3102/0013189X024007005

Piaget, J. (1967). *Six psychological studies*. (A. Tenzer, Trans., D. Elkind, Ed.). New York, NY: Vintage Books.

Piaget, J. (1977). *The essential Piaget*. (H. E. Gruber & J. J. Vonèche, Eds.). New York, NY: Basic Books.

Rasmussen, T. (2013). *A meta-analysis of academic success courses in postsecondary institutions.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3572090).

Ravitch, D. (2000). *Left back: A century of failed school reforms.* New York, NY: Simon & Schuster.

Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education.* New York, NY: Basic Books.

Ravitz, J. (2009). Introduction: Summarizing findings and looking ahead to a new generation of PBL research. *Interdisciplinary Journal of Problem-Based Learning, 3*(1), 4-11. http://dx.doi.org/10.7771/1541-5015.1088

Resnick, L. B., & Glaser, R. (1976). Problem solving and intelligence. In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ridlon, C. L. (2009a/b). Learning mathematics via a problem-centered approach: A two-year study. *Mathematical Thinking and Learning, 11*(4), 188-225.

http://dx.doi.org/10.1080/10986060903225614

Rogers, M. A. P., Cross, D. I., Gresalfi, M. S., Trauth-Nare, A. E., & Buck, G. A. (2009). First year implementation of a project-based learning approach: The need for addressing teachers' orientations in the era of reform. *International Journal of Science and Mathematics Education, 9*(4), 893-917. http://dx.doi.org/ 10.1007/s10763-010-9248-x

Rosenshine, B. (1987). Explicit teaching and teacher training. *Journal of Teacher Education, 38*(3), 31-36. http://dx.doi.org/10.1177/002248718703800308

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641. http://dx.doi.org/10.1037/0033-2909.86.3.638

Rosenthal. R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage Publications.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*(2), 183-192. http://dx.doi.org/10.1037//0033-2909.118.2.183

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*(1), 59-82. http://dx.doi.org/10.1146/annurev.psych.52.1.59

Ross, E. W., & Marker, P. M. (2005). Social studies: Wrong, right, or left?, A critical response to Fordham Institutes' *Where did social studies go wrong? Social Studies, 96*(4), 139-142. http://dx.doi.org/10.3200/TSSS.96.4.139-142

Savery, J. R. (2006). Overview of problem-based learning: Definitions and distinctions. *Interdisciplinary Journal of Problem-Based Learning, 1*(1), 9-20.

http://dx.doi.org/10.7771/1541-5015.1002

Schmidt, H. G., Loyens, S. M. M., Van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 91-97. http://dx.doi.org/10.1080/00461520701263350

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*(1), 113-125. http://dx.doi.org/10.1023/A:1003044231033

Schunk, D. H. (1999). Social-self interaction and achievement behavior. *Educational Psychologist, 54(*4), 219-227. http://dx.doi.org/10.1207/s15326985ep3404_3

Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In B. Zimmerman & D. H. Schunk, *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.) (pp. 125-151). Mahwah, NJ: Lawrence Erlbaum Associates.

Scott, C. A. (1994). Project-based science: Reflections of a middle school teacher. *The Elementary School Journal, 95*(1), 75-94. Retrieved from http://www.jstor.org/stable/1001774

Simmers, E. J., & Dickinson, G. (2012). A longitudinal investigation of project-based instruction and student achievement in high school social studies. *Interdisciplinary Journal of Problem-based Learning, 6*(1), 82-103. http://dx.doi.org/10.7771/1541-5015.1313

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5-11. http://dx.doi.org/10.3102/0013189X015009005

Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, *37*(1), 5-14. http://dx.doi.org/10.3102/ 0013189X08314117

Stanley, W. B. (2005). Social studies and the social order: Transmission or transformation. *Social Education, 69,* 282-286. Retrieved from http://publications.socialstudies.org/se/ 6905/6905282.pdf

Stepien, W., & Gallagher, S. (1993). Problem-based learning: As authentic as it gets. *Educational Leadership*, *51*(7), 25-28.

Sungar, S., Tekkaya, C., & Geban, O. (2011). Improving achievement through problem-based learning. *American Educational Research Journal, 40*(4), 155-160. http://dx.doi.org/10.1080/00219266.2006.9656037

Tan, I. G., Sharan, S., & Lee, C. K. (2007. Group investigation effects on achievement, motivation, and perceptions of students in Singapore. *The Journal of Educational Research, 40*(4), 142-154. http://dx.doi.org/10.3200/JOER.100.3.142-154

Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology, 58*(9), 894-901. http://dx.doi.org/ 10.1016/j.jclinepi.2005.01.006

Thomas, J. W. (2000). *A review of research on project-based learning* [PDF]. San Rafael, CA: The Autodesk Foundation. Retrieved from http://bie.org/object/ document/a_review_of_research_on_project_based_learning

U.S. Department of Education. (2009). *Race to the Top program executive summary* [PDF]. Washington, DC: U.S. Department of Education. Retrieved from http://www2.ed.gov/programs/ racetothetop/executive-summary.pdf

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies* (ED-04-CO-0040) [PDF]. Washington, DC: U.S. Department of Education. Retrieved from http://www2.ed.gov/rschstat/eval/tech/ evidence-based-practices/finalreport.pdf

van Loggerenberg-Hattingh, A. (2003). Examining learning achievement and experience of science learners in a problem-based learning environment. *South African Journal of Education, 23*(1), 52-57. Retrieved from http://www.ajol.info/ index.php/saje/article/view/24846

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3-14. http://dx.doi.org/10.1007/s11409-006-6893-0

Vernon, D. T., & Blake, R. L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine, 68*(7), 550-563. Retrieved from http://journals.lww.com/academicmedicine/pages/default.aspx

Vilardi, V. A. (2013). *An examination of the influence of inquiry-based laboratory activities and success on standards based achievement tests in a suburban high school* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3562470)

Visser, Y. L. (2003). *The effect of problem-based and lecture-based instructional strategies on learner problem solving performance, problem solving process, and* *attitudes* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3109316)

von Glasersfeld, E. (1989). Cognition, construction of knowledge, and teaching. *Synthese, 80*(1), 121-140. http://dx.doi.org/10.1007/BF00869951

von Glasersfeld, E. (2006). Radical constructivism and teaching. [PDF]. Originally published in French in *Perspectives, 31*(3), 191-204. Retrieved from http://www.vonglasersfeld.com/244.2 (Original work published 2001).

Vygotsky, L. S. (1962). *Thought and language.* (E. Hanfmann & G. Vakar, Trans.). New York, NY: John Wiley and Sons. (Original work published 1934).

Vygotsky, L. S. (1978). *Mind in society.* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman Trans.). Cambridge, MA: Harvard University Press.

Walker, A., & Leary, H. (2009). A problem based learning meta analysis: Difference across problem types, implementation types, disciplines, and assessment levels. *International Journal of Problem-based Learning*, *3*(1), 12-43. http://dx.doi.org/ 10.7771/1541–5015.1061

Washington State Teacher/Principal Evaluation Project (TPEP) (2013). *Instructional Frameworks.* Olympia, WA: Office of Superintendent of Public Instruction. Retrieved from http://tpep-wa.org/the-model/framework-and-rubrics/instructional-frameworks/

Webb, N. M. (1982). Peer interaction and learning in cooperative small groups. *Journal of Educational Psychology, 74*(5), 642-655. http://dx.doi.org/10.1037/0022-0663.74.5.642

Webb, N. M., & Kenderski, C. M. (1984). Student interaction and learning in small-group and whole-class settings. In P. L. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction: Group organization and group processes,* (pp. 153-170). Orlando, FL: Academic Press, Inc.

Wells, C. S., & Womack, J. A. (2003). *An instructor's guide to understanding test reliability.* Testing & Evaluation Services, University of Wisconsin. Retrieved from http://testing.wisc.edu/Reliability.pdf

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3-118. http://dx.doi.org/10.1207/s1532690xci1601_2

Wirkala, C., & Kuhn, D. (2011). Problem-based learning in K-12 education: Is it effective and how does it achieve its effects? *American Educational Research Journal, 48*(5), 1157-1186. http://dx.doi.org/10.3102/0002831211419491

Wolters, C. A. (2010). *Self-regulated learning and the 21st century competencies* [PDF]. Unpublished manuscript, Department of Psychology, University of Houston, Houston, TX. Retrieved from The William and Flora Hewlett Foundation at http://www.hewlett.org/uploads/Self_Regulated_Learning__21st_Century_Competencies.pdf

Wong, K. H. (2012). *Implementation of problem-based learning in junior secondary science curriculum*. (Unpublished doctoral dissertation). University of Hong Kong, Hong Kong. Retrieved from http://hub.hku.hk/bitstream/10722/193087/1/FullText.pdf?accept=1

Worry, V. (2011). *A comparison of high school geometry student performance and motivation between traditional and project-based instruction techniques*

(Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3478823)

Worsham, T. (1992). The inclusion process: Cooperative metacognition for discovering, describing, and assessing thinking skills. In N. Davidson & T. Worsham (Eds.), *Enhancing thinking through cooperative learning* (pp. 66-83). New York, NY: Teachers College Press.

Wright, J. R. (2009). *An examination of the implementation of the intel essentials project-based learning model on middle and secondary reading and language arts FCAT student achievement and engagement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3388797)

Yager, S., Johnson, D. W., & Johnson, R. T. (1985). Oral discussion, group-to-individual transfer, and achievement in cooperative learning groups. *Journal of Educational Psychology*, *77*(1), 60-66. http://dx.doi.org/10.1037/0022-0663.77.1.60

Yancy, Y. G. (2012). *The effects of project-based learning activities on intrinsic motivation and skill acquisition of rural middle school math students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3514801)

Zimmerman, B. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. Zimmerman & D. H. Schunk, *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.) (pp. 1-37). Mahwah, NJ: Lawrence Erlbaum Associates.

Zimmerman, B., & Schunk, D. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum

Associates.

Zohar, A., & Ben David, A. B. (2008). Explicit teaching of meta-strategic knowledge in authentic classroom situations. *Metacognition and Learning*, *3*(1), 59-82. http://dx.doi.org/10.1007/s11409-007-9019-4

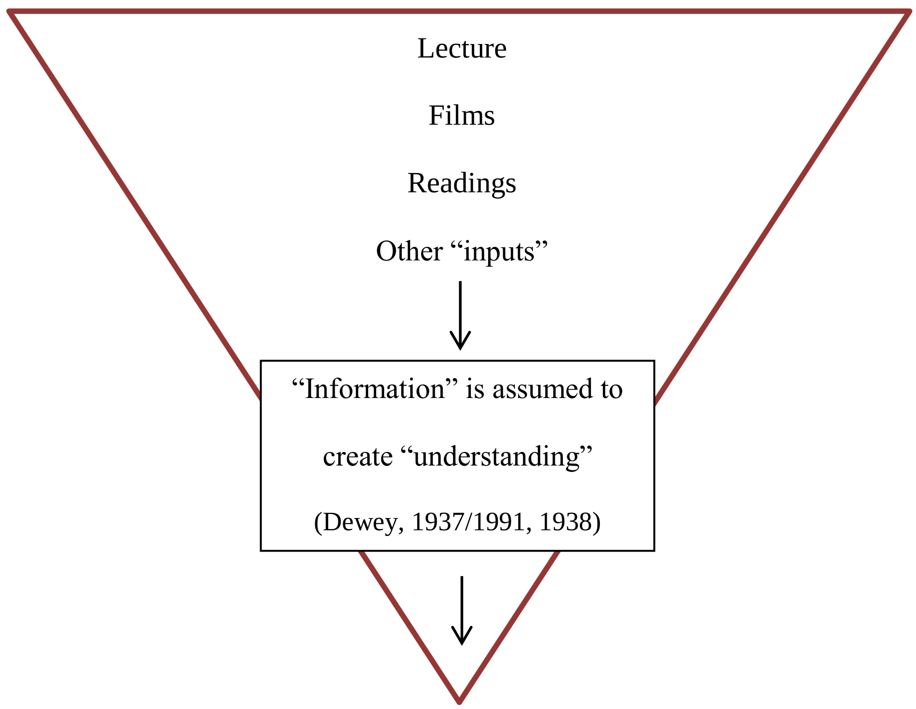
Appendix A

Traditional/Universal Direct Instruction Paradigm

Acquire knowledge and/or skills: "Information"

(Dewey, 1937/1991; Kilpatrick, 1918: Parker et al., 2011)

Teacher Initiated and Driven



End Unit Assessment/Project

The "project" is used as an "end result" to assess "learning." Some refer to projects in this sense as "authentic assessment"

(Parker et al., 2011; Thomas, 2000)

Problem- and Project-Based (PBL) Learning Paradigm

Begin with the Problem or Project:

Teacher-Student Initiated and Driven

(Hmelo-Silver, 2004; Kilpatrick, 1918; Parker et al., 2011; Thomas, 2001)

The problem/project drives knowledge/skill acquisition

Understanding

requires acquisition of knowledge

(Bruner, 1993; Dewey, 1933, 1937/1991;

Kilpatrick, 1918; Schmidt et al., 2007)

and

Experiential Learning:

"Understanding has to be in terms of how

things work and how to do things"

(Dewey, 1937/1991, p. 184; see also

Bruner, 1996; Kilpatrick, 1918;

Hmelo-Silver & DeSimone, 2013)

The process drives understanding and internalization

End with the Problem Solution or Project

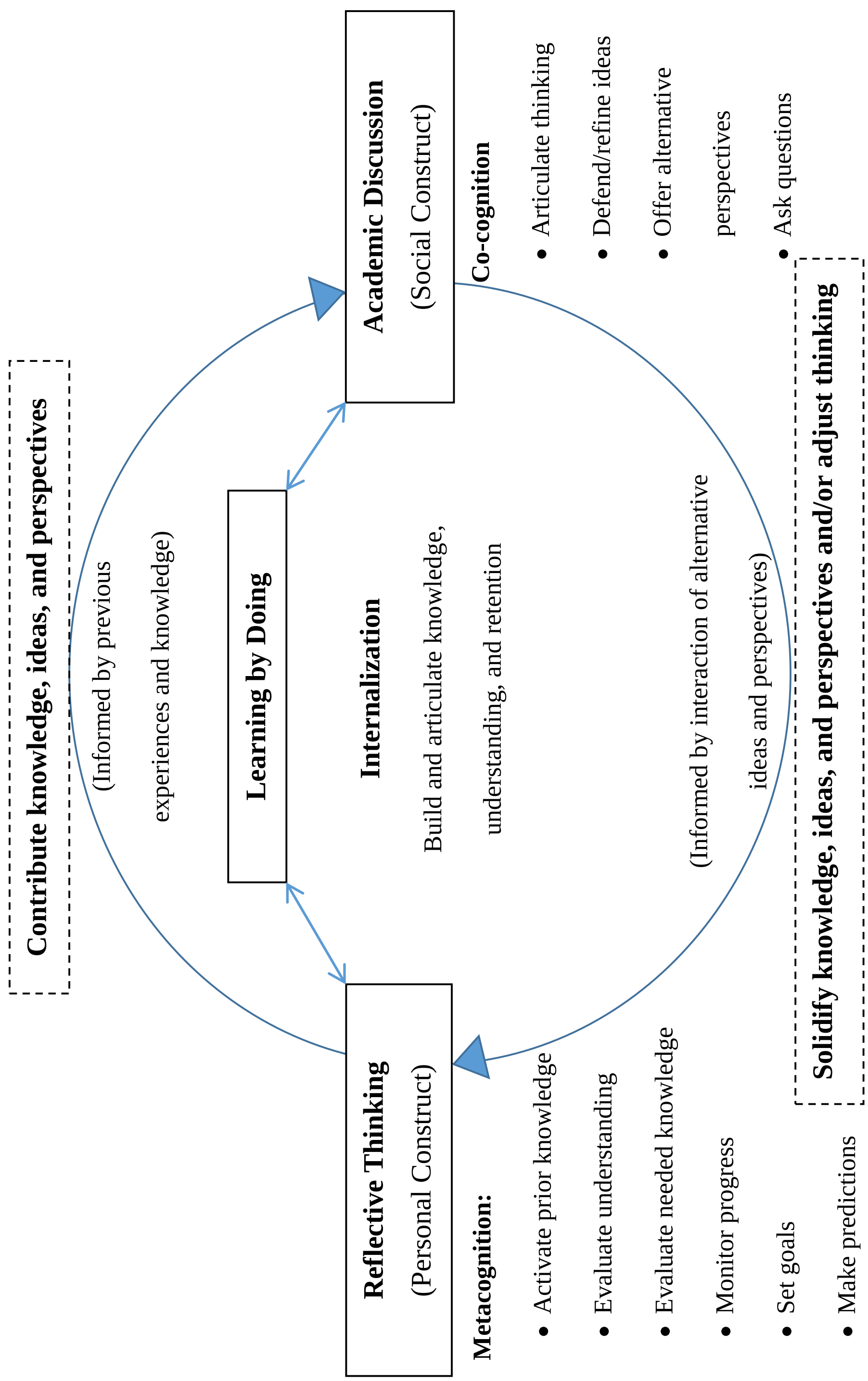
Demonstration of student's amassed experience and acquired knowledge

(Dewey, 1938; Kilpatrick, 1918; Parker et al., 2011; Thomas, 2000)

Appendix B

The Reflection-Discussion Paradigm:

Reciprocal Interaction of Personal and Social Constructs



Appendix C

Screening and Study Characteristics Form

Study Name		**Study#**	
Study author(s) classified as ☐ Problem-based learning ☐ Project-based learning ☐ Problem/Project based ☐ Inquiry based			

SOURCE CHARACTERISTICS
Year of Publication:

Type of Publication:
- ☐ Unpublished article, report, working paper, or conference paper/proceeding
- ☐ Unpublished doctoral dissertation/master's thesis
- ☐ Submitted for publication
- ☐ Published: Peer-reviewed journal or report
- ☐ Published: Non-peer reviewed (self-published article, university website, etc.)
- ☐ Book/monograph

STUDY CHARACTERISTICS
Operationalized Definition of PBL? ☐ YES ☐ NO ☐ Technology enhanced
Reflective assessment used: ☐ Yes ☐ NS Academic discussion used: ☐ Yes ☐ NS

QUASI? ☐ YES ☐ NO PRE-TEST/EQUIVALANCY Measures: ☐ YES ☐ NO

Exclusion: ☐ Not PBL ☐ No LD comp. ☐ Ex Post Facto ☐ Experimental ☐ Single group
☐ Qualitative ☐ Other:

Grade level: ☐ Middle School/JH (mixed grades) ☐ High School (mixed grades

Academic subject: ☐ Science ☐ Math ☐ Social Studies/History ☐ ELA ☐ Other:

Sample size (Lipsey & Wilson, 1993, p. 1195)
- ☐ Less than 50 (includes 50)
- ☐ 51 to 100
- ☐ More than 100

Duration of treatment:
- ☐ Less than one (1) week
- ☐ 1-3 weeks
- ☐ 4-6 weeks
- ☐ 7-10 weeks (≈one academic quarter)
- ☐ 10-20 weeks (≈one academic semester)
- ☐ 20-40 weeks (≈one academic year)
- ☐ More than one academic year

ROLE/Experimental facilitator: ☐ Teacher ☐ Researcher ☐ Both
- ☐ Highly interactive (scaffolding/mini-lecture, monitor group progress, conducts reflective assessment strategies, conducts discussion, encourages thinking)
- ☐ Active (facilitates some of the features described above)
- ☐ Passive (students left to "learn" on their own: the teacher is mostly just a supervisor)
- ☐ Not specified

STUDENT/SCHOOL DEMOGRAPHICS

Country/Region of Study:

Ethnicity: ☐ Not specified
White %: Black %: Hispanic/Latino %: Asian %: Other%:

SES % FLR: ☐ Low SES ☐ Middle SES ☐ High SES ☐ Not specified

Gender: ☐ Not Specified Male%: Female%:

Ability level: ☐ Average achieving ☐ Low achieving ☐ High achieving ☐ Mixed ability
☐ Not specified

Demographic Setting: ☐ Urban ☐ Suburban ☐ Rural ☐ Mixed ☐ Not specified

School Setting: ☐ Public ☐ Alternative ☐ Charter/Magnet ☐ Private ☐ Not specified

METHODOLOGICAL CONSIDERATIONS

Assessments

☐ Teacher developed assessment
☐ Researcher developed assessment
☐ Standardized: state, national, or professional organization developed assessment

Type of Assessment

☐ Acquisition of knowledge (content)

Cronbach's alpha: Inter-rater reliability: Other:

PBL N $\bar{X}$ SD F t p df

LD N $\bar{X}$ SD F t p df

☐ Application of knowledge (skills)

Cronbach's alpha: Inter-rater reliability: Other:

PBL N $\bar{X}$ SD F t p df

LD N $\bar{X}$ SD F t p df

☐ Combined content and skills: Weighed the means ☐ Yes ☐ No

Appendix D

Coding Schematic

Source Characteristics

Publication Date	1: 1985-2001 (Post *A Nation at Risk* and inclusion of 1990s standards' movement) 2: 2001-2009 (*No Child Left Behind Act of 2001*(NCLB) and high stakes testing (HST) to pre-implementation of *Race to the Top* (RTTT) and Common Core State Standards (CCSS) initiatives) 3: 2010-2015 (Current state with NCLB, RTT, CCSS, and HST)
Publication Type	1: Unpublished article, report, working paper, or conference paper 2: Unpublished doctoral dissertation/master's thesis 3: Submitted for publication 4: Published: Peer reviewed journal or report 5: Published: Non-peer reviewed (self-published article, university website, other academic/paper upload site) 6: Book/monograph

Study Characteristics

PBL Classification	1: Problem-based learning 2: Project-based learning 3: Problem- and project-based learning 4: Inquiry-based and/or Problem-based inquiry
Use of Reflective Assessment (RA) & Academic Discussion (AD)	1: Neither RA nor AD specified 2: RA specified in procedures 3: AD specified in procedures 4: RA and AD specified in procedures
Grade Level	1: Middle School (Grades 6-8) 2: High School (Grades 11-12)
Academic Subject	1: Science (Lab and non-lab) 2: Math 3: Social Studies (History, Economics, Geography, Government) 4: ELA

	5: Other/Elective
Sample Size (Lipsey & Wilson, 1993, p. 1195)	1: Less than 50 (includes 50) 2: 51-100 3: More than 100
Study Duration	1: Less than 1 week 2: 1-3 weeks 3: 4-6 weeks 4: 7-10 weeks (approx. one academic quarter) 5: 11-20 weeks (approx. one academic semester) 6: 21-40 weeks (approx. one academic year) 7: More than one academic year
Experimental facilitator	1: Classroom Teacher: Professional development in PBL not specified 2: Classroom Teacher: Professional development in PBL (brief or unspecified duration); training manual/guide, but not instructed 3: Classroom Teacher: Professional development in PBL (extensive) or experienced in PBL 4: Researcher (Experience with PBL assumed)
Facilitator role	1: Highly interactive (scaffolding/mini-lecture, monitor group progress, conducts reflective assessment strategies, conducts discussion, encourages thinking) 2: Active (facilitates some of the features above) 3: Passive (begins project/problem, then student left to “learn” on their own: facilitator is mostly just a supervisor) 99: Not specified
Computer/Technology Enhanced	1: Non-computer/technology enhanced (traditional) 2: Computer/technology enhanced
Student, School, and Regional Demographics	
Country/Region	Categorical list
Predominant Ethnicity/Race by %	1: Greater than 60% White 2: Greater than 60% Black 3: Greater than 60% Hispanic/Latino 4: Greater than 60% Asian/Pacific Islander

	5: Greater than 60% Native American 6: No category greater than 60% 7: Approximately equal proportions of two or more categories 8: Ethnic composition other than U.S. designated category 99: Not specified
Socioeconomic Status % Free-Reduced Lunch	1: FLR: Less than 10% 2: FLR: 10-25% 3: FLR: 26-50% 4: FLR: 51-75% 5: FLR: 76-100% 6: Low SES 7: Middle SES 8: High SES 9: Mixed SES 99: Not specified
Gender	1: Greater than 60% female 2: Greater than 60% male 3: Approx. equal female/male 99: Not specified
Ability level	1: Average achieving 2: Low achieving 3: High achieving 4: Talented and gifted (TAG) 5: Mixed ability 99: Not specified
Demographic Setting	1: Urban 2: Suburban 3: Rural 4: Mixed 99: Not specified
School Setting	1: Public 2: Alternative

	3: Charter/magnet
	4: Private
	99: Not specified
Methodological Characteristics	
Pre-test/Equivalency	1: Pre-test equivalency and/or posttest scores adjusted
	2: Other equivalency
	3: No pretest/equivalency or unspecified
Assessment Development	1: Teacher developed assessment
	2: Researcher developed assessment
	3: Standardized assessment: State, national, or professional org.
Test-retest reliability	1: Stated: greater than Cronbach's alpha 0.70 (or equivalent)
	2: Stated: less than Cronbach's alpha 0.70 (or equivalent)
	99: Not specified

Appendix E

Studies Included in the Meta-Analysis: Overview

Table E1

Overview of Included Studies

Study Name	***N***	**Location**	**Subject**	**Grade Level**	**Duration**
Afolabi & Akinbobola (2009)	105	Nigeria	Science	High School	4 - 6 weeks
Ankinoglu & Tandogan (2007)	50	Turkey	Science	Middle School	7 - 10 weeks
Anyafulude (2013)	255	Nigeria	Science	High School	4 - 6 weeks
Araz & Sungar (2007)	217	Turkey	Science	Middle School	4 - 6 weeks
Baş (2011)	60	Turkey	Other/Elective	High School	4 - 6 weeks
Bayrak & Bayram (2011)	56	Turkey	Science	Middle School	4 - 6 weeks
Brokes (2010)	133	United States	Math	High School	4 - 6 weeks
Chang (2001)	159	Taiwan	Science	High School	1 - 3 weeks
Cicchino (2015)	177	United States	Social Studies	Middle School	Less than 1 week
Elshafei (1998)	342	United States	Math	High School	4 - 6 weeks
Finkelstein, Hanson, Huang, Hirschman, & Huang (2011)	3752	United States	Social Studies	High School	11 - 20 weeks
Hernandez-Ramos & De La Paz (2009)	169	United States	Social Studies	Middle School	4 - 6 weeks
Kuşdemir, Ay, & Tüysüz (2013)	52	Turkey	Science	High School	7 - 10 weeks
Maree & Molepo (2005)	427	South Africa	Math	High School	21 - 40 weeks
Mergendoller, Maxwell, & Bellisimo (2000)	186	United States	Social Studies	High School	4 - 6 weeks
Mergendoller, Maxwell, & Bellisimo (2006)	246	United States	Social Studies	High School	1 - 3 weeks
Mioduser & Betzer (2007)	107	Israel	Other/Elective	High School	More than one academic year

Nafees, Farooq, Tahirkheli, & Akhtar (2012)	67	Pakistan	Science	High School	11 - 20 weeks
Parker et al. (2011)	269	United States	Social Studies	High School	21 - 40 weeks
Parker et al. (2013)	262	United States	Social Studies	High School	21 - 40 weeks
Ridlon (2009a)	52	United States	Math	Middle School	7 - 10 weeks
Ridlon (2009b)	52	United States	Math	Middle School	7 - 10 weeks
Roesch, Nerb, & Riess (2015)	213	Germany	Science	Middle School	4 - 6 weeks
Sungar, Tekkaya, & Geban (2006)	61	Turkey	Science	High School	4 - 6 weeks
Tan, Sharan, & Lee (2007)	241	Singapore	Social Studies	Middle School	4 - 6 weeks
Tarhan & Acar (2007)	40	Turkey	Science	High School	Less than 1 week
Tarhan, Ayar-Kayal, Urek, & Acar (2008)	78	Turkey	Science	High School	1 - 3 weeks
van Loggerenberg-Hattingh (2003)	140	South Africa	Science	High School	7 - 10 weeks
Vilardi (2013)	166	United States	Science	High School	11 - 20 weeks
Visser (2003)	60	United States	Science	High School	7 - 10 weeks
Wirkala & Kuhn (2011a/b)	89	United States	Social Studies	Middle School	Less than 1 week
Wong (2012)	125	Hong Kong	Science	Middle School	21 - 40 weeks
Worry (2011)	65	United States	Math	High School	1 - 3 weeks
Wright (2009)	1423	United States	ELA	High School	21 - 40 weeks
Yancy (2012)	111	United States	Math	Middle School	11 - 20 weeks

Note. Descriptive statistics are based on 34 sets of study participants. The Ridlon (2009a/b) studies are treated as two separate studies as different participants were used in each study. The Wirkala and Kuhn (2011) study is treated as a single study in the *Overview of Included Studies* and *Summary of Study Characteristics* tables, because the researchers used the same participants in a crossed within subjects design. In all other summary tables, the Wirkala and Kuhn study is treated as two studies, as these participants were put in different conditions over two topics.

Appendix F

Studies Included in the Meta-Analysis: Characteristics

Table F1

Summary of Study Characteristics of 35 Independent Samples

Source Characteristics		
Variable	***f***	**Rate**
Publication Date		
1985-2001: Post *A Nation at Risk*	2	5.7
2001-2009: Post *No Child Left Behind Act of 2001*	17	48.6
2010-2015: Post *Race to the Top* and CCSS	16	45.7
Publication Type		
Unpublished article, report, working paper, or conference paper	0	---
Doctoral dissertation/master's thesis	8	22.9
Submitted for publication	0	---
Published: Peer reviewed journal or report	27	77.1
Published: Non-peer reviewed (self-published article, university website, other academic/paper upload site)	0	---
Book/monograph	0	---

Study Characteristics		
Variable	***f***	**Rate**
PBL Classification		
Problem-based learning	23	65.7
Project-based learning	9	25.7
Problem- and project-based learning	1	2.9
Inquiry-based and/or Problem-based inquiry	2	5.7
Use of Reflective Assessment & Academic Discussion		
Neither RA nor AD specified	11	31.4
RA specified in procedures	4	11.4
AD specified in procedures	11	31.4
RA and AD specified in procedures	9	25.7

Grade Level		
Middle School (Grades 6-8)	12	34.3
High School (Grades 11-12)	23	65.7
Academic Subject		
Science (Lab and non-lab)	16	45.7
Math	7	20.0
Social Studies	9	25.7
English/Language Arts (ELA)	1	2.9
Other/elective	2	5.7
Sample size (Lipsey & Wilson, 1993, p. 1195)		
Less than 50 (includes 50)	2	5.7
51 - 100	11	31.4
More than 100	22	62.9
Study Duration		
Less than 1 week	3	8.6
1 - 3 weeks	4	11.4
4 - 6 weeks	12	34.3
7 - 10 weeks (approx. one academic quarter)	6	17.1
11 - 20 weeks (approx. one academic semester)	4	11.4
21 -4 0 weeks (approx. one academic year)	5	14.3
More than one academic year	1	2.9
Experimental Facilitator		
Classroom Teacher: Professional development in PBL not specified	9	25.7
Classroom Teacher: Professional development in PBL (brief or unspecified duration); training manual/guide, but not instructed	7	20.0
Classroom Teacher: Professional development in PBL (extensive) or experienced in PBL	13	37.1
Researcher (Experience with PBL assumed)	6	17.1
Computer/Technology Enhanced		
Non-computer/technology enhanced (traditional)	27	77.1
Computer/technology enhanced	8	22.9

Student, School, and Regional Demographics

Variable	*f*	Rate
Predominant Ethnicity/Race by %		
Greater than 60% White	5	14.3
Greater than 60% Black	3	8.6
Greater than 60% Hispanic/Latino	2	5.7
Greater than 60% Asian/Pacific Islander	3	8.6
Greater than 60% Native American	0	---
No category great than 60%; categories not equal	2	5.7
Approximately equal proportions of two or more categories	4	11.4
Ethnic composition other than U.S. designated category	10	28.6
99 Not specified	6	17.1
Socioeconomic Status % Free-Reduced Lunch		
FLR: Less than 10%	1	2.9
FLR: 10-25%	1	2.9
FLR: 26-50%	3	8.6
FLR: 51-75%	3	8.6
FLR: 76-100%	0	---
Low SES	1	2.9
Middle SES	5	14.3
High SES	0	---
Mixed SES	1	2.9
Not specified	20	57.1
Gender		
1: Greater than 60% female	2	5.7
2: Greater than 60% male	3	8.6
3: Approx. equal female/male	11	31.4
99: Not specified	19	54.3
Ability level		
Average achieving	3	8.6
Low achieving	3	8.6

High achieving	2	5.7
Talented and gifted (TAG)	0	---
Mixed ability	14	40.0
Not specified	13	37.1
Demographic Setting		
Urban	11	31.4
Suburban	7	20.0
Rural	4	11.4
Mixed	4	11.4
Not specified	9	25.7
School Setting		
Public (comprehensive)	24	68.6
Alternative	1	2.9
Charter/magnet	1	2.9
Private	6	17.1
Not specified	3	8.6

Methodological Characteristics

Variable	***f***	**Rate**
Pre-test/Equivalency		
Pretest equivalency and/or posttest scores adjusted	29	82.9
Other equivalency	3	8.6
No pretest/equivalency or unspecified	3	8.6
Assessment Development*		
Teacher developed	3	6.2
Researcher developed	37	77.1
Standardized assessment	8	16.7
Test-retest reliability		
Stated: greater than Cronbach's alpha 0.70	22	62.9
Stated: less than Cronbach's alpha 0.70	2	5.7
Not specified	11	31.4

Note. Descriptive statistics are based on 34 sets of study participants. The Ridlon (2009a/b) studies are treated as two separate studies as different participants were used in each study. The Wirkala and Kuhn (2011) study is treated as a single study in the *Overview of Included Studies* and *Summary of Study Characteristics* tables, because the researchers used the same participants in a crossed within subjects design. In all other summary tables, the Wirkala and Kuhn study is treated as two studies, as these participants were put in different conditions over two topics. *Assessment development frequency and rate is based on *N* = 48 outcomes.

Appendix G

Studies Included in the Meta-Analysis: References

Afolabi, F., & Akinbobola, A. O. (2009). Constructivist problem based learning technique and the academic achievement of physics students with low ability level in Nigerian secondary school. *Eurasian Journal of Physics and Chemistry Education, 1*(1), 45-51. Retrieved from http://eurasianjournals.com/ index.php/ejpce/article/view/34/59

Ankinoğlu, O., & Tandoğan, R. O. (2007). The effects of problem-based active learning in science education on students' academic achievement, attitude and concept learning. *Eurasia Journal of Mathematics, Science & Technology Education, 3*(1), 71-81. Retrieved from http://ejmste.com/v3n1/EJMSTEv3n1_Akinoglu.pdf

Anyafulude, J. C. (2013). Effects of problem-based and discovery-based instructional strategies on students' academic achievement in chemistry. *Journal of Educational and Social Research, 3*(6), 105-112. http://www.doi.org/10.5901/ JESR.2013.v3n6p105

*Araz, G., & Sungur, S. (2007). Effectiveness of problem-based learning on academic performance in genetics. *Biochemistry and Molecular Biology Education, 35*(6), 448-451. doi:10.1002/bmb.97

Baş, G. (2011). Investigating the effects of project-based learning on students' academic achievement and attitudes towards English lessons. *The Online Journal of New Horizons in Education, 1*(4), 2011. Retrieved from http://www.tojned.net/ pdf/tojnedv01i04-01.pdf

Bayrak, B. K., & Bayram, H. (2011). Effects of problem-based learning in a web environment on conceptual understanding: The subject of acids and bases. *International Online Journal of Educational Sciences, 3*(3), 831-848. Retrieved from http://www.iojes.net//userfiles/Article/IOJES_710.pdf

*Brokes, J. C. (2010). *The effect of using a problem/project based, document driven unit of instruction, on high school students' achievement on the data analysis cluster of the HSPA and on their attitude toward mathematics and data analysis* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3411512)

Chang, C. -Y. (2001). Comparing the impact of a problem-based computer-assisted instruction and the direct-interactive teaching method on student science achievement. *Journal of Science Education and Technology, 10*(2), 147-153. http://dx.doi.org/10.1023/A:1009469014218

*Cicchino, M. I. (2015). Using game-based learning to foster critical thinking in student discourse. *Interdisciplinary Journal of Problem-based Learning, 9*(2), http://dx.doi.org/10.7771/1541-5015.1481

*Elshafei, D. L. (1998). *A comparison of problem-based and traditional learning in Algebra II* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 9919416)

Finkelstein, N., Hanson, T., Huang, C. W., Hirschman, B., & Huang, M. (2011). *Effects of Problem Based Economics on high school economics instruction.* (NCEE 2010-4022rev) [PDF]. Washington, DC: National Center for Education

Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/edlabs/ regions/west/pdf/REL_20104022.pdf

*Hernández-Ramos, P., & De La Paz, S. (2009). Learning history in middle school by designing multimedia in a project-based learning experience. *Journal of Research on Technology in Education, 42*(2), 151-173. http://dx.doi.org/10.1080/ 15391523.2009.10782545

Kuşdemir, M., Ay, Y., & Tüysüz, C. (2013). An analysis of the effect of problem based learning model on the 10th grade students' achievement, attitude and motivation in the unit of "mixtures." *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, *7*(2), 195-224. Retrieved from http://www.nef.balikesir.edu.tr/~dergi/makaleler/yayinda/15/EFMED_KME141.pdf (Google Translate)

*Maree, J. G., & Molepo, J. M. (2005). Achievement in mathematics in grades 9 and 11 in Limpopo Province South Africa: Introduction of a problem based approach. *Psychological Reports*, *97*(3), 732-736. http://dx.doi.org/10.2466/PR0.97.7.732-736

*Mergendoller, J. R., Maxwell, N. L., & Bellisimo, Y. (2000). Comparing problem-based learning and traditional instruction in high school economics. *The Journal of Educational Research, 93*(6), 374-382. http://dx.doi.org/10.1080/0022067 0009598732

*Mergendoller, J. R., Maxwell, N. L., & Bellisimo, Y. (2006). The effectiveness of problem-based instruction: A comparative study of instructional methods and student characteristics. *Interdisciplinary Journal of Problem-based Learning, 1*(2), 49-69. http://dx.doi.org/10.7771/1541-5015.1026

Mioduser, D., & Betzer, N. (2007). The contribution of project-based-learning to high-achievers' acquisition of technological knowledge and skills. *International Journal of Technology and Design Education, 18*(1), 59-77. http://dx.doi.org/ 10.1007/s10798-006-9010-4

*Nafees, M., Farooq, G., Tahirkheli, S. A., Akhtar, M. (2012). Effects of instructional strategies on academic achievement in a high school general science class. *International Journal of Business and Social Science, 3*(5), 161-166. Retrieved from http://ijbssnet.com/journals/Vol_3_No_5_March_2012/18.pdf

Parker, W., Mosborg, S., Bransford, J., Vye, N., Wilkerson, J., & Abbott, R. (2011). Rethinking advanced high school coursework: Tackling the depth/breadth tension in the AP US Government and Politics course. *Journal of Curriculum Studies, 43*(4), 533-559. http://dx.doi.org/10.1080/00220272.2011.584561.

Parker, W. C., Lo, J., Yeo, A. J., Valencia, S. W., Nguyen, D., Abbott, R. D.,…Vye, N. J. (2013). Beyond breadth-speed test: Toward deeper knowing and engagement in and advanced placement course. *American Educational Research Journal, 50*(6), 1424-1459. http://dx.doi.org/ 10.3102/0002831213504237

Ridlon, C. L. (2009a/b). Learning mathematics via a problem-centered approach: A two-year study. *Mathematical Thinking and Learning*, *11*(4), 188-225. http://dx.doi.org/10.1080/10986060903225614

Roesch, F., Nerb, J., & Riess, W. (2015). Promoting experimental problem-solving ability in sixth-grade students through problem-oriented teaching of ecology:

Findings of an intervention study in a complex domain. *International Journal of Science Education*, *37*(4), 577-598. http://dx.doi.org/10.1080/09500693.2014. 1000427

*Sungar, S., Tekkaya, C., & Geban, O. (2011). Improving achievement through problem-based learning. *American Educational Research Journal*, *40*(4), 155-160. http://dx.doi.org/ 10.1080/00219266.2006.9656037

Tan, I. G., Sharan, S., & Lee, C. K. (2007. Group investigation effects on achievement, motivation, and perceptions of students in Singapore. *The Journal of Educational Research*, *40*(4), 142-154. http://dx.doi.org/10.3200/JOER.100.3.142-154

Tarhan, L., & Acar, B. (2007). Problem-based learning in an eleventh grade chemistry class: 'Factors affecting cell potential'. *Research in Science and Technological Education*, *25*(3), 351-369. http://dx.doi.org/10.1080/02635140701535299

Tarhan, L., Ayar-Kayali, H., Urek, R. O., & Acar, B. (2007). Problem-based learning in 9th grade chemistry class: 'Intermolecular forces.' *Research in Science Education*, *38*(3), 285-300. http://dx.doi.org/10.1007/s11165-007-9050-0

*van Loggerenberg-Hattingh, A. (2003). Examining learning achievement and experience of science learners in a problem-based learning environment. *South African Journal of Education*, *23*(1), 52-57. Retrieved from http://www.ajol.info/ index.php/saje/article/view/24846

*Vilardi, V. A. (2013). *An examination of the influence of inquiry-based laboratory activities and success on standards based achievement tests in a suburban high school* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3562470)

Visser, Y. L. (2003). *The effect of problem-based and lecture-based instructional strategies on learner problem solving performance, problem solving process, and attitudes* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3109316)

Whitehead, A. N. (1929). *The aims of education*. Presidential address to the Mathematical Association of England, 1916. New York, NY: Mentor Books.

Wirkala, C., & Kuhn, D. (2011a/b). Problem-based learning in K-12 education: Is it effective and how does it achieve its effects? *American Educational Research Journal, 48*(5), 1157-1186. http://dx.doi.org/10.3102/0002831211419491

Wong, K. H. (2012) *Implementation of problem-based learning in junior secondary science curriculum*. (Unpublished doctoral dissertation). University of Hong Kong, Hong Kong. Retrieved from http://hub.hku.hk/bitstream/10722/193087/1/FullText.pdf?accept=1

*Worry, V. (2011). *A comparison of high school geometry student performance and motivation between traditional and project-based instruction techniques* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3478823)

Wright, J. R. (2009). *An examination of the implementation of the intel essentials project-based learning model on middle and secondary reading and language arts FCAT student achievement and engagement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3388797)

*Yancy, Y. G. (2012). *The effects of project-based learning activities on intrinsic motivation and skill acquisition of rural middle school math students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global. (UMI No. 3514801)

Appendix H

Internal Review Board (IRB) Approval Documentation

Exempt Review

Subject: IRB Approval – IRB # 141506009(Exempt)

Dear Ms. Jensen,

Your research project "*A meta-analysis of the effect of problem- and project-based learning on academic achievement in 6-12 populations,*" has been approved under exempt IRB review. This study was approved under exempt review as it met the following criteria.

3. ___X_ Research uses survey or interview procedures or observations (including observations by participants) of public behavior AND at least one of the following conditions exist:
   a. __X_ Human participants cannot be identified directly or through identifiers code or numbers

   OR

   b. __x__ The participants' responses or the observations recorded, if they became known outside research, cannot reasonably place the participant at risk of criminal or civil liability or be damaging to the participant's financial standing or employment

   OR

   c. __x__ The research does not deal with sensitive aspects of the participant's own behavior, such as illegal conduct, drug use, sexual behavior, or use of alcohol

Your approval is in effect until **01/20/2016**. Your study has been assigned IRB number: **IRB # 141506009.**

To complete your documents please add your IRB # and expiration date to you study's written recruitment material and invitation to participate in the research project.

Please contact me when you have completed collecting data for your study so that I can close your file. If you need more than one year to complete data collection, you must file a request for an extension with me six weeks before the expiration date of this study. Your request for an extension can be written or communicated through e-mail and must include a report on the status of your study. Otherwise you will need to file a new IRB application to continue with data collection after the expiration date.

Use your study number in any further communication regarding this study.

This is the only documentation that you will receive regarding your study's approval. Please print it out and add to your study's documentation**.**

Best Wishes in the Completion of your Research

Thomas Alsbury, IRB Committee Member-SOE Rep.
Petersen, Room 401
Ph: 206-378-5099
Email: alsburyt@spu.edu