A Problem Based Learning Meta Analysis: Differences Across Problem Types, Implementation Types, Disciplines, and Assessment Levels

Andrew Walker and Heather Leary

Abstract

Problem based learning (PBL) in its most current form originated in Medical Education but has since been used in a variety of disciplines (Savery & Duffy, 1995) at a variety of educational levels (Savery, 2006). Although recent meta analyses have been conducted (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Gijbels, Dochy, Van den Bossche, & Segers, 2005) that attempted to go beyond medical education, they found only one study in economics and were unable to explain large portions of the variance across results. This work builds upon their efforts as a meta-analysis that crosses disciplines as well as categorizes the types of problems used (Jonassen, 2000), the PBL approach employed (Barrows, 1986), and the level of assessment (Gijbels et al., 2005; Sugrue, 1993, 1995). Across 82 studies and 201 outcomes the findings favor PBL (d_w = 0.13, +/- .025) with a lack of homogeneity (Q = 954.27) that warrants a closer examination of moderating factors.

Introduction

Problem based learning (PBL) is most widely associated with the large body of literature coming out of medical education. Briefly, PBL is characterized as an approach to learning in which students are given more control over their learning than a traditional approach, asked to work in small groups, and most importantly acquire new knowledge only as a necessary step in solving authentic, ill-structured, and cross-disciplinary problems representative of professional practice (Barrows, 1986, 1996, 2002; Barrows & Tamblyn, 1980). This approach to learning arose, in part, from a sharp contrast between experiences at the beginning and end of medical school. During the first two years, students were put off by learning vast amounts of factual information, unsure of its connection to their future practice. During their residency however, they tended to be highly motivated while engaging with patients and their problems (Spaulding, 1969). With over three decades of research going back to Neufeld & Barrows (1974), there is a robust collection of primary research and even several meta-analyses of PBL (Albanese & Mitchell, 1993; Dochy et al., 2003; Gijbels et al., 2005; Kalaian, Mullan, & Kasim, 1999; Vernon & Blake, 1993). These works are important and provide an invaluable contribution, particularly the examination of assessment levels by Gijbels et al. (2005). That said, these analyses remain focused on the discipline of medical education and still do not account for a large amount of variance among the findings.

There are several possible explanations for this variance. In one early and oft-cited work, Barrows posits a taxonomy of PBL implementations (1986). The taxonomy moves from lecture-based cases to closed-loop problem based learning and includes a claim that the closed-loop approach is best positioned to enhance at least four different educational objectives. It seems logical to expect that the type of PBL implementation might play a role in learning outcomes. Another potential source of variance is the problem types students with which engage. Jonassen has proposed a typology of problems (ranging from logical problems to dilemmas) that includes features like associated learning activities, inputs, success criteria, context, structuredness, and abstractness (2000). Problem types may prove even more important with the examination of work outside of medical education and the associated diagnosis-solution problems that pervade it.

In the interest of filling these existing and emerging gaps, the purpose of this analysis is twofold: 1) To investigate differences across a broad range of disciplines and assessment levels in PBL outcomes, and 2) To characterize PBL implementations and investigate features, such as Jonassen's typology of problems (2000) and Barrows' taxonomy (1986), that may act as moderators in student achievement.

Literature Review

Definition of PBL. As Barrows (1996) noted PBL has taken on a myriad of definitions pushed, in part, by institutions wanting to refine their particular approach. This becomes true to an even greater extent with PBL expanding to several different disciplines and contexts (Savery, 2006; Savery & Duffy, 1995). Changes to PBL as initiated by institutions to reflect their needs and the needs of their discipline have made it somewhat difficult to construct a clear statement about what is and what is not PBL. With that caveat in mind, and borrowing heavily from Barrows as one of the initial proponents of PBL, the definition for this research includes the following:

- *Ill-structured problems* are presented as unresolved so that students will generate not just multiple thoughts about the cause of the problem, but multiple thoughts on how to solve it (Barrows, 2002). Such problems may not have a single correct answer and should engage students in the exploration of multiple solution paths (Hmelo-Silver & Barrows, 2006).

- A *student centered* approach in which students determine what they need to learn. It is up to the learners to derive the key issues of the problems they face, define their knowledge gaps, and pursue and acquire the missing knowledge (Barrows, 2002; Hmelo-Silver & Barrows, 2006).
- *Teachers act as facilitators* or tutors in the learning process. These tutors, typically faculty, initially prompt students with meta-cognitive questions and in subsequent sessions fade that guidance (Barrows, 2002). Tutors forgo lecturing about content in favor of modeling the kinds of learning processes that lead to success in PBL settings (Hmelo-Silver & Barrows, 2006).
- *Authenticity* forms the basis of problem selection, embodied by alignment to professional or "real world" practice (Barrows, 2002). As such, the problems are inherently cross-disciplinary and require students to investigate multiple subjects (Barrows, 1996) in order to generate a workable solution.

For the purposes of this article, these components constitute the minimum standards of PBL. Barrows and Hmelo-Silver add to these components an addendum that PBL is typically undertaken in a small group setting (Barrows, 2002; Hmelo-Silver & Barrows, 2006). While groups of five to nine students were used in the original McMaster model for PBL (Barrows, 1996), these later definitions allow for the possibility of PBL without small group work. Thus, cases of large group PBL which were investigated by Barrows with favorable results (Barrows, Myers, Williams, & Moticka, 1986), or cases of individual student PBL are encompassed in the definition of PBL used here.

Several meta-analyses have already been conducted on PBL, and a robust meta-synthesis of these efforts is available in this issue (Strobel & Barneveld, 2009). Although much is known as a result of this work, all of the studies maintain a narrow focus in terms of discipline. Both of the most recent meta-analyses (Dochy et al., 2003; Gijbels et al., 2005) include a stated goal of encompassing PBL literature irrespective of discipline but between them only included a single study in high school economics (Son & Van Sickle, 2000). Also, as mentioned above, all of these analyses were unable to account for variance across study findings.

Assessment in PBL. In an attempt at parsing some of the variance, Gijbels et al. (2005) aggregated outcomes by level of assessment. These efforts are based on a theoretical framework for dividing the assessment of problem solving into several distinct parts (Sugrue, 1993, 1995). At the concept level, assessment consists of defining, identifying or even generating examples of fundamental constructs. This is declarative knowledge in its purest sense. Revealing relationships between these concepts is assessment at the principles level. Principles might be rule-based or more emergent but in general they rely on some sort of an underlying probabilistic model to define the associations between concepts. The application level, referred to as application conditions and procedures by Sugrue (1993), assesses the ability of learners to correctly invoke principle and concept knowledge to achieve a goal state. A key component of application-level work is procedures for using principle and concept knowledge in new situations.

Rather than replicate the rationale for using this framework, readers are referred to the justification given by Gijbels et al. (2005). Note they do reiterate a limitation voiced initially by Sugrue (1995). Specifically, this includes the acceptance of one or more probabilistic models to assess at the principle, and by extension, application levels. For some domains, such as mathematics, probabilistic models are more available or perhaps come with less debate than in a domain like history. That said, Sugrue's definitions for principles and concepts are based on Merrill's Component Display Theory (Sugrue, 1993) which carries with it the intention for broad based application to a variety of subject matter (Merrill, 1983). Note that assumptions would still have to be made for the framework to apply. Either the assessment would have to take a firm stance on any uncertainty within the principles of the domain or be robust in allowing for alternative probabilistic models.

Thought process, as first introduced by Albanese and Mitchell (1993), is used to describe several different constructs. Most of these can be included with the previously described levels of assessment. Their last use of process, however, deserves separate consideration. This version of process as a form of assessment differentiates forward-driven reasoning from backward-driven reasoning. Experts commonly use forward- or data-driven reasoning, which relies on having a well-defined cognitive structure or schema from which a diagnosis can be achieved almost simultaneously with recognition of symptoms (Gilhooly, 1990; Patel, Groen, & Norman, 1991). While experts reason quite well using this approach, novices tend to struggle and are prone to commit errors when using it (Claessen & Boshuizen, 1985; Hmelo, Gotterer, & Bransford, 1997; Patel et al., 1991). Backward- or hypothesis-driven reasoning involves testing a series of preliminary diagnoses and reasoning backward through the probabilistic model of principles to determine the concepts or presenting symptoms that should be present if the diagnosis is accurate (Hmelo et al., 1997). This is a reasoning process that is specifically reinforced in PBL, for good or ill, and has yet to be examined in a systematic way during or since its introduction by Albanese and Mitchell (1993).

Problem Types. One cited shortcoming of PBL and related student-centered approaches to learning is a focus on problem solving without a corresponding examination of the underlying problems with which students engage (Jonassen, 2000). In an effort to fill this void, Jonassen derived eleven problem types through collecting and conducting a cognitive task analysis across hundreds of sample problems. At one end of a loose spectrum are highly structured problems with a focus on accurate and efficient paths to an optimal solution where context is a secondary concern. The other extreme includes ill-structured problems where context is crucial, solutions may not even exist, and evaluation is more about the evidence and chain of reasoning employed than the solution itself. Each problem type is described in terms of associated learning activities, inputs, success criteria, context structuredness, and abstractness. Within each type, there is variability in the level of abstraction as well as complexity. In the case of complexity, there is so much variability that no attempt is made to describe it in relation to problem type. Following is a short summary of problem types as defined in the typology and as they relate to PBL. Note that Jonassen's typology was not intended to focus exclusively on the context of PBL, but rather problem solving as a whole. As such, problem types at both ends of the loose continuum are likely to be inappropriate for PBL and will be discussed as such below. That said they were included in this analysis because much of the typology strays from absolutes, and because their appropriateness for PBL remains an empirical question.

It seems likely that the first couple of problem types would not align well with PBL. They are heavily constrained, highly structured and as such it is difficult to imagine a group engaged in a meaningful search for knowledge in advance of working towards solutions. Logical problems are highly constrained and generally abstracted, such as drawing four straight lines that intersect all points in a 3x3 array of dots. They rely on a single approach to reasoning to unlock the solution (in this particular case, a willingness to draw a triangle that goes outside the bounds of the dot array, which is then bisected by the fourth line). Algorithmic problems cover the formulas you might expect for symbol manipulation domains like math or physics but also things like recipes for cooking. They tend to focus on following appropriate steps to arrive at a solution state.

The next series of problems are likely a good fit for PBL. They tend to align much better in terms of authenticity outside of a formal learning environment and are comparatively ill structured. Story problems have underlying algorithms but add two critical components. First, they provide a context wrapper for what amounts to an algorithmic problem, and they require the learner to engage in a process of unpacking this story into relevant components and making decisions about appropriate procedures for solving it. Rule-Using problems represent one of the more diverse categories both in terms of complexity and structuredness. Examples can be as simple as doubling a recipe to as complex as playing chess. Internet searching is also classified as a rule using problem, which can be quite ill structured given the ability to invoke one of several different search strategies. Decision-Making problems are just what they sound like, selecting from a set of alternatives and their associated consequences. They also involve associated activities—such as generating additional alternatives, and assessing the risks and benefits of alternatives. Trouble-Shooting problems are about goal state and current state discrepancies. They require several different kinds of knowledge (systems, procedural, and strategic). A classic example might be a lap-top and projector that are failing to show the computer screen. An expert technician will systematically reduce the problem space to diagnose the fault. Diagnosis-Solution problems are rife within classic medical education PBL but clearly expand beyond it as well. They are the next logical step after trouble-shooting and involve resolving the fault state through weighing alternative options and monitoring progress.

One example is pruning and caring for a fruit tree left fallow for several years. Strategic performance consists of thinking both strategically and tactically. For example, a teacher implements a lesson strategically and employs a combination of tactical teaching approaches in support of that overall strategy. However, the teacher may need to adapt those tactics on the fly when misconceptions are noted among the learners. The classic examples of case analysis problems are those emerging from business and law schools. The intention is to be more reflective of authentic situations, to be quite ill structured by leaving much of the case open ended or ill defined, including such things as the goal state itself. Cases do not lend themselves to a set process for solution beyond high level heuristics. Design problems might include something like designing a robot to explore Mars. They incorporate knowledge that crosses disciplines (such as computer science and mechanical engineering), and tend to be quite complex and ill structured. Less technical examples might include writing a prose poem or designing instruction. Often, design problems require the use of artificial systems to aid in their solution, such as use of the ADDIE model in Instructional Systems Design.

The most ill structured of problems are Dilemmas, in part because they may not have a solution at all. For this reason, dilemmas may be ill suited to problem-based learning since it carries a tacit assumption of there being a correct response. These might be the traditional dichotomy represented in many debate settings (for example abortion or immigration reform) or be far more complex (such as global warming). Possible solutions will generally marginalize one or more stakeholders because of complex interrelationships and multiple concerns. It is unclear how effective PBL can be at handling the kinds of value judgments embedded within dilemmas.

PBL Method. In an effort to classify and differentiate among several different methods of PBL, Barrows created a taxonomy (1986). This work highlights what he characterized as quality differences between the various approaches, specific to the following educational goals, which are adapted here to a context wider than medical education: 1) structuring knowledge of all types in a way that supports problem solving, 2) a reasoning process for problem solving, 3) self-directed learning skills, and 4) increased motivation for learning. Presumably, as these educational goals are achieved to a greater extent, corresponding increases should be seen in learning outcomes—particularly but not limited to those centered on principle and application level assessments. The taxonomy centers on how much these goals are facilitated by each PBL method. Methods are characterized according to 1) the complexity of the problems, 2) the focus on teacher or student centered learning and 3) the order of problem and case and information presentation.

In lecture-based cases, teacher-directed information is presented prior to cases in which all of the relevant information is already provided. While there is still a need to diagnose the problem and generate a solution, there is no need for free inquiry. The idea is to provide some sort of a context for the information provided, but this represents the poorest alignment to the four learning goals above. Case-based lectures are largely the same with the exception of exposing students to cases before corresponding lectures. With the added benefit of having a context in mind, this should facilitate a better structuring of the knowledge acquired from the teacher. The case method approach utilizes the same fully elaborated cases (e.g., all necessary data are provided up front), followed by a discussion that is directed partially by students and partially by the teacher. Modified case-based approaches open up the case to allow for some inquiry, but it is generally cued or constrained. This represents the first method in the taxonomy that is largely student-directed. Problem based as a method is not only student directed, but centers on a problem that allows for free inquiry. Teachers act as facilitators and help students in recalling relevant prior knowledge and potentially lead them towards identifying any of their own misconceptions. Finally, in closed-loop problem based approaches learners are asked to revisit the problem to determine any improvements they could make to their reasoning process. As part of that exercise, they evaluate the information sources used as well as their own prior knowledge. This last method represents the best alignment to the educational goals above.

Although there is a great deal of face validity in Barrows' taxonomy it has yet to be empirically tested. This work is an effort to validate some of these claims as well as determine the extent to which observed variance in PBL findings can be explained by the PBL method employed, the discipline of the intervention, the problem types used, or the assessment level of the outcome.

Methods

Although meta-analysis is not a primary research study it shares common traits in terms of formulating a problem, collecting data (studies in this case), coding the data, analysis, and interpretation (Cooper & Hedges, 1994a). It was used here to answer the following research questions: 1) To what extent does discipline of study moderate PBL outcomes? 2) To what extent is the PBL method, as defined by Barrows' taxonomy (1986), a moderating variable on PBL outcomes? 3) To what extent do problems types, as identified by Jonassen (2000), moderate PBL outcomes? 4) To what extent does level of assessment (Gijbels et al., 2005) moderate PBL outcomes? and 5) Can these factors, in any combination, be used to reliably predict PBL outcomes? The following sections are an attempt to make the process for this analysis as transparent as possible from the search strategies employed (White, 1994) to the inclusion criteria and statistical procedures used.

Inclusion Criteria. Inclusion criteria included the following elements. The first was quantitative outcomes focused on either student learning or their reasoning processes as described above. These outcomes either had to include a statistical significance for inclusion in the vote count analysis, or preferably include enough data to calculate an effect size. The second was a comparison between a PBL (treatment) condition and a control (lecture) condition. The PBL treatment had to include engagement with ill-structured and authentic problems, student-directed learning, and tutors acting as facilitators, as described under the PBL definition above. In keeping with the recent discussions of small group interaction as typical rather than required (Barrows, 2002; Hmelo-Silver & Barrows, 2006), studies without small group interaction were kept.

Literature Search. Searching began with existing meta-analyses and reviews (Albanese & Mitchell, 1993; Berkson, 1993; Dochy et al., 2003; Gijbels et al., 2005; Kalaian et al., 1999; Vernon & Blake, 1993) for primary research previously covered. Once these articles were reviewed and coded, a list of journals and keywords with frequency counts was compiled to determine which databases to search and the best search terms to use. Updates to this list continued throughout the search process. Subject descriptors from the database thesaurus were used when applicable for the best fit of search terms. A thorough search was then conducted of the electronic databases Education Resources Information Center (ERIC), PsychInfo, Education Full Text, Google Scholar, Communication of the ACM, CiteSeer, and Digital Dissertations looking for empirical articles that fit the inclusion criteria. While no age limit was placed on most databases, Digital Dissertations was searched only back to 2002. This was partially an effort to make the set manageable and partially under the assumption that dissertations older than that would have an alternate publication venue (e.g., conferences, technical reports or journal articles). The Medline search was restricted to 2004-2007 since this literature is so well covered by the two most recent reviews (Dochy et al., 2003; Gijbels et al., 2005). Finally, as the articles were reviewed citations of empirical PBL studies were flagged and sourced for possible inclusion.

Coding Scheme. Two researchers independently coded each study along several dimensions and then met to discuss their findings until consensus was achieved. The vast majority of revisions were the result of omission by either coder rather than direct conflict. Specific dimensions reported in this study included assessment type as originally used in Gijbels et al. (2005), with possible values of *concept*, *principle*, and *application* as described above. In addition we also classified an assessment type for reasoning *process* as originally used by Albanese and Mitchell (1993). These are reported separately because of conflicting statements about the utility of data- vs hypothesis-driven reasoning (Albanese & Mitchell, 1993; Hmelo et al., 1997; Kirschner, Sweller, & Clark, 2006). Problem type was coded with possible levels of *logical*, a*lgorithmic, story, rule using, decision making, trouble-shooting, diagnosis-solution*, *strategic performance*, *case analysis*, or *dilemma.* In all cases, inferences were made based on available descriptions of the intervention or, in the case of medical education literature, diagnosis-solution problems were assumed. PBL method was coded as *lecture-based cases, case-based lecture, case method, modified case-based, problem-based* or *closed-loop problem-based* only if explicitly reported as such in the primary research article. Finally the discipline of study was recorded as precisely as possible with the intention of later collapsing into broader categories. In some cases, natural categories existed, such as medical education, but even this was refined. Medical education was reserved for medical doctors, allied health was added for professionals like respiratory therapists, nurse anesthetists, and dentists. The final categories for discipline include *allied health*, *business*, *engineering*, *medical education*, *science*, *social science*, *teacher education*, and *other* which included disciplines like aviation, kinesiology, and textiles.

For this meta-analysis, the common metric is standardized mean difference (*d*). Like prior PBL analyses (Vernon & Blake, 1993) the denominator for calculation of *d* depends on the available data from each study. When possible, the pooled estimate of the population standard deviation was used. Effect sizes were also computed from *p* values using the lowest reported value (e.g., $p < .05$ became $p = .05$). Although this was quite rare it likely resulted in an underestimation of any calculated effect sizes (Shadish & Haddock, 1994). All effect sizes were calculated with the aid of a freely available tool for the purpose, ESFree (http://inst.usu.edu/~aewalker/ESFree/app_about/) with directionality and inputs confirmed by both researchers. Effect sizes are reported as positive when PBL students performed better than lecture students. The one exception was with process-level assessment. Here when PBL students exhibited more hypothesis- or less data-driven reasoning the PBL outcomes are positive but readers should not associate this direction with any kind of value judgment.

Results

The analysis reports effect sizes weighted by sample (Cooper, 1989). This was particularly important in a study that involved sample sizes as small as 8 (Boshuizen, Schmidt, & Wassamer, 1993) and as large as 2,469 (Martenson, Ericksson, & Ingelman-Sundberg, 1985). Specific to the PBL literature, Colliver advocated for effect sizes of 0.8-1.0 (2000) before PBL can be seen as a success. This seems unnecessarily strict as it would fail to accept several behavioral, educational and even pharmaceutical interventions currently in common practice (Albanese, 2000). Instead, these findings will be discussed in terms of Cohen (1988) who reluctantly classified effect sizes of 0.2 as small, 0.5 as medium and 0.8 as large. For unit of analysis, this synthesis makes the assertion that multiple outcomes from the same study should not be condensed when there is a strong rationale for expected differences (for example assessment of principles as opposed to assessment of concepts). Thus, the unit of analysis will be at the outcome level.

Vote count analysis (Bushman, 1994) is reported alongside each finding as a means of more conservatively estimating any observed differences and also to obtain a comparison for the purposes of missing data—put simply, a check to see if the findings of the more inclusive vote count differed substantially from the meta-analysis, which can only include studies with sufficient data to compute effect sizes. Much more controversial is the use of inferential statistics in meta-analysis, which has been both advocated for (Cooper & Hedges, 1994a) and argued against (Glass, 2000, 2006). At their core, these arguments center on meta-analysis as a probabilistic sampling of studies from a defined population of research. The results below report confidence intervals (at the 95% level), tests of heterogeneity, and a regression analysis, all of which imply a nebulous theoretical population of PBL studies. Rather than an attempt to take sides in the debate, these are presented as data to be interpreted or ignored by the reader. Unless specifically mentioned, all of the results and analyses below exclude the reasoning process outcomes.

In total, 201 outcomes across 82 studies had codeable effect sizes that met our inclusion criteria. See Appendix A for the full set of findings including 8 additional process level outcomes. Of the 201 outcomes, findings approached a small effect size in favor of PBL (d_w = 0.13, +/- .025). The vote count analysis with 68 positive outcomes and only 21 negative is statistically significant in favor of PBL ($p < .001$). Of particular interest in the context of this study, the homogeneity test $Q = 954.27$ is not statistically significant at the .01 level, justifying the grouping of outcomes to assess the impact of different variables.

Discipline of Study. In terms of discipline, as can be seen below in Table 1, there are somewhat large discrepancies across the included outcomes. While teacher education studies seemed to do quite well ($d_w = 0.64$), particularly in contrast with engineering ($d_w = 0.05$), and science ($d_w = 0.06$) which show essentially identical outcomes as lecture-based approaches. Many of the vote count analyses violate the assumptions of chi square by not having at least five outcomes in each cell. That said, the vote count evidence seems to back up the effect size computations, with science and engineering showing an even mix, and the others, particularly at $d_w > 0.15$, with uniformly positive findings.

While the vast majority of outcomes (n = 133) are in medical education, those results ($d_w = 0.09$) are not the most promising. This should be encouraging for researchers doing PBL in other areas, especially studies involving teacher education ($d_w = 0.64$), the catch all other category ($d_w = 0.48$), or in the social sciences ($d_w = 0.30$). There is also a clear need for additional quantitative controlled studies in teacher education, social

Table 1. *Discipline area outcomes.*

Discipline	sig. +	sig. -	$N_{outcomes}$	d_w	CI_{Lower}	CI_{Upper}
teacher education	1	0	4	0.635	0.443	0.827
other	5	0	13	0.482	0.307	0.658
social science	3	0	6	0.299	0.100	0.499
allied health	5	0	22	0.258	0.179	0.336
business	3	0	6	0.159	0.026	0.292
medical education	45[a]	16	133	0.085	0.056	0.115
science	4	4	12	0.062	-0.063	0.187
engineering	2	1	5	0.048	-0.197	0.292
all	68[a]	21	201	0.127	0.101	0.152

[a]Significant ($p < .05$) sign test on the vote count analysis.

science, business, and engineering and a less dramatic need for work in the sciences and the other category.

Assessment Level. Assessment level, as can be seen in Table 2, departs from the findings of Gjibels et al. (2005). Whereas their largest favorable findings for PBL also came at the principle and application level they found much larger effects for application outcomes. In this study, both principle (d_w = .21) and application (d_w = .33) level assessments had favorable but modest performance. The principle level vote count was statistically significant in favor of PBL. At the application level an overwhelming number of studies (*n* = 27) had favorable findings. Several factors contributed to these differences in the effect sizes as compared with previous meta-analyses. Specifically, this analysis includes new studies such as Enarson and Cariaga-Lo (2001), which contained negative principle level outcomes. Some studies that were included in previous meta-analyses either had more modest effect size computations (Finch, 1999) based on available data or were coded differently, with some principle level assessments judged as application level instead. Concept level outcomes appear to be almost identical (d_w = -.04) between PBL and lecture, which is backed up by the vote count analysis and quite close to previous meta-analyses.

In a separate analysis a total of eight process level outcomes across five different studies show that PBL students engage in more hypothesis-driven and less data-driven reasoning than their lecture-based counterparts (d_w = 0.49, +/- 0.23). Although the number of outcomes precludes a meaningful sign test the vote count data (2 positive outcomes, 0 negative outcomes) parallel these results. When these results are broken down further it appears that PBL students engage in a great deal of hypothesis-driven reasoning (d_w = 1.04, +/- 0.35) whereas the lack of data-driven reasoning (d_w = 0.28, +/- 1.04) is more modest. This makes sense given that backwards-driven reasoning tends to reflect the intended process that PBL students are asked to undertake (Hmelo et al., 1997). The implications in terms of learning will be discussed below.

Problem Type. As expected, Table 3 reveals the vast majority of the literature includes diagnosis-solution problem types (d_w = 0.11), which come close to the medical education findings above. This should come as no surprise since there are far more medical educa-

Table 2. *Assessment level outcomes.*

Assessment Level	sig. +	sig. -	$N_{outcomes}$	d_w	CI_{Lower}	CI_{Upper}
concept	19	15	73	-0.043	-0.092	0.005
principle	12[a]	4	40	0.205	0.142	0.268
application	28	0	60	0.334	0.287	0.382
mixed (concept & application)	0	0	1	0.168	-0.357	0.692
missing	9[a]	2	27	0.067	0.018	0.115
all	68[a]	21	201	0.127	0.101	0.152

[a]Significant (p < .05) sign test on the vote count analysis.

tion (n = 133) than allied health (n = 22) results contributing to the diagnosis solution outcomes. Of particular note is a rough trend that seems to vary somewhat in conjunction with Jonassen's typology (2000). Specifically, design problems (d_w = 0.74) did quite well, one of the largest single effect sizes found in the review. Strategic-performance problems also did well (d_w = 0.53). While both of these results need to be interpreted with a great deal of caution, since they include only 5 outcomes total, they do suggest a trend in favor of problems that are bit further along the continuum than the PBL roots of diagnosis solution. In addition, there is evidence of a decline at both extremes of the typology. Story problems did not perform much better in PBL than lecture in the one instance it was used (d_w = 0.11) and dilemmas (d_w = -0.18) showed better performance among lecture-based students.

In addition to needing more information about some types of problems (e.g., story, troubleshooting, strategic performance, and design) there is no information at all about

Table 3. *Problem type outcomes.*

Problem Type	sig. +	sig. -	$N_{outcomes}$	d_w	CI_{Lower}	CI_{Upper}
story	0	0	1	0.112	-1.089	1.313
troubleshooting	2	0	6	0.194	0.024	0.365
diagnosis solution	52[a]	16	153	0.107	0.079	0.135
strategic performance	0	0	2	0.528	0.207	0.850
design	1	0	3	0.740	0.521	0.960
dilemmas	3	3	8	-0.179	-0.333	-0.026
missing	10[a]	2	28	0.262	0.179	0.344
all	68[a]	21	201	0.127	0.101	0.152

[a]Significant ($p < .05$) sign test on the vote count analysis.

others (logical, algorithmic, rule using, and decision-making) and many of the outcomes (n = 28) could not be coded with certainty.

PBL Method. Although Barrows taxonomy (1986) is often cited and includes a plea for both researchers and practitioners to describe their PBL interventions in terms of the critical components of the taxonomy, very few have been explicit and transparent in doing so. The amount of information available for PBL method is rather disappointing, as can be seen in Table 4 below. Of the two studies that specified PBL method, both used closed-loop problem based learning. The five total outcomes from this work included assessments at the concept, principle, and application level and indicated some of the largest findings in favor of PBL (d_w = 0.54). These outcomes favor PBL at a moderate level, and seem to agree with the vote count trend of three positive outcomes and none that are negative.

The fact that PBL does so much better when it uses the closed-loop problem based approach provides support for Barrows' claims about potential benefits in terms of education goals. These findings need to be interpreted with caution since they are based on

Table 4. *PBL method outcomes.*

PBL Method	sig. +	sig. -	$N_{outcomes}$	d_w	CI_{Lower}	CI_{Upper}
closed-loop	3	0	5	0.538	0.419	0.658
Missing	65 [a]	21	196	0.107	0.081	0.133
All	68 [a]	21	201	0.127	0.101	0.152

[a]Significant ($p < .05$) sign test on the vote count analysis.

a rather small amount of evidence. In addition this interpretation makes the assumption that closed-loop PBL increased things like self-directed learning and motivation without including measurement, at least in this meta-analysis, of these variables.

Regression. A regression analysis was used to determine the ways in which these variables combine to impact cognitive outcomes. Specifically, this involved backward elimination linear regression with effect size (weighted by sample) as the dependent variable, and assessment level, problem type, PBL method, and discipline as the predictors. Threshold alphas were set at .5 for initial entry and .1 for subsequent removal. The resulting model does explain a statistically significant portion of the variance, $R^2 = .25$, $F\ (6, 194) = 10.75$, $p < .001$. In terms of individual variables, outcomes are more favorable when they diverge from the disciplines of allied health ($t = -2.15$, $p < .033$) or medical education ($t = -4.15$, $p < .001$) and avoid assessment at the concept level ($t = -2.34$, $p < .021$). They also benefit from application level assessment ($t = 3.70$, $p < .001$) and use of the closed-loop PBL method ($t = 3.81$, $p < .001$). In the final regression model, the independent variables are able to explain 25% of the variability in cognitive outcomes. Some caution is warranted as the regression falls 54 outcomes shy of the recommended 15 data points per variable (Stevens, 1999).

Conclusions

The major contribution of this analysis is the inclusion of 47 outcomes outside the fields of medical education and allied health. Since its modern inception in medical education a robust interest in PBL has resulted in use across several disciplines (Savery & Duffy, 1995). This represents the first full synthesis to examine the impact of PBL. Across almost all of the analyses run, PBL students either did as well as or better than their lecture-based counterparts, and they tended to do better when the subject matter was outside of medical education, a result that is bolstered by the multiple regression analysis. More inclusive vote count analyses either found similarity in results, or a greater number of statistically significant findings in favor of PBL.

In comparison with previous meta-analyses there are cases of both agreement and departure. Early findings pointed to concept or content knowledge differences favoring lecture (Albanese & Mitchell, 1993; Dochy et al., 2003; Vernon & Blake, 1993), although the same was not always true for the vote count (Dochy et al., 2003; Vernon & Blake, 1993). Our findings match almost exactly with the more recent analysis by Gijbels et al.(2005), which is particularly interesting given the expansion of disciplines covered. This is important given recent claims about the ineffectiveness of minimally guided instruction as a whole (Kirschner et al., 2006). Even more encouraging are the favorable outcomes at the principle and application level of assessment.

Our process level outcomes agree with previous findings (Albanese & Mitchell, 1993) that PBL students engage in far more backward-driven reasoning. According to some, this is undesirable as it results in more errors during problem solving (Albanese & Mitchell, 1993) and may persist even after the educational intervention is complete (Kirschner et al., 2006). If that does hold true, it is interesting that PBL students managed to do so well despite these errors. There is agreement that backwards-driven reasoning tends to take much more time (Hmelo et al., 1997; Kirschner et al., 2006). Hmelo et al. (1997) also point out that while experts may engage in forward reasoning for a typical case, they will use backward-driven reasoning when presented with a novel situation. It appears that both reasoning processes have merit and value depending on the situation, and a look at the long-term reasoning process of PBL students should be undertaken.

As initially posited by Barrows (1986), problem type does appear to play a role in the effects of PBL. Jonassen's much more in-depth view of problem types (Jonassen, 2000) is at least a step towards a better understanding of some of the variation across study findings. The sharp contrast between design problems ($d_w = .74$) and dilemmas ($d_w = -0.18$) is quite surprising given their immediate proximity within Jonassen's typology. One possible explanation is the nature of each problem type. While both support multiple solution paths, dilemmas incorporate competing interest and may not have a correct solution at all. Still, it remains odd that a consistent upward trend, with effect sizes increasing as problems move towards the extreme of Jonassen's typology, takes such a sudden downward turn.

While much more needs to be known about which PBL methods were employed before confident assertions can be made, closed-loop problem based learning appears to improve student learning outcomes ($d_w = 0.54$). That said, there may well be a relationship between PBL method and problem type that we simply do not yet have enough data to reveal. Barrows did discuss problem selection and presentation as a feature of the PBL method employed (Barrows, 1986). While his discussion is notably short it does involve the level of information about a problem provided to a student ahead of time and implies the degree to which the information is dynamic. These relate to Jonassen's (2000) descriptions for problem structure and problem complexity.

Limitations. Sugrue's assessment framework (Sugrue, 1995) includes constructs that are not reported here but do align with PBL as a whole. Specifically, these include metacognitive elements of planning and monitoring and motivational elements of perceived self-efficacy, task difficulty, and task attraction. Both meta-cognition and motivation relate directly to definitional works of PBL (Barrows, 1986, 1996; Savery, 2006; Savery & Duffy, 1995). While the vast majority of studies reporting learning outcomes do not measure these constructs they hold great potential to explain other differences that might be expected between PBL and lecture-based students in educational settings (Albanese, 2000). This could be particularly true with performance-based outcomes that are appropriately motivating, ask learners to draw on their domain-specific knowledge, and make use of their meta-cognitive skills. Put more simply, this analysis uses just a portion of a much larger theoretical framework for the assessment of problem solving.

Similarly, alignment to Jonassen's work is also a focused view of a much larger theory. His specific interest is promoting overall skill in problem solving, of which problem variation is only one component. The other components include representations such as the social context of the problem solving task, and individual differences which can impact the nature of the problem itself. Additionally, as described above, he discusses a wide range of problem complexity within each problem type. This specifically deals with the number of interrelationships, the components of the problem, the ways in which they are represented, and the degree to which they or the context of the problem change over time. With Jonassen discussing a positive relationship between problem complexity and problem difficulty (2000), and Kirschner et al. (2006) discussing problem complexity as contributing to non-germane load and poor learning outcomes it seems likely that this may be a source of variance not yet explained.

Purists will likely question the inclusion of studies that did not utilize small group interactions. Of the 201 outcomes 160 not only employed small group interaction, but reported enough data to calculate an average group size. Of the remaining 41, a total of 6 outcomes across 3 studies clearly did not use small group interaction (Johnson, Flesher, Jehng, & Ferej, 1993; Robertson, 2005; Yang, 2002). For convenience, these are specifically labeled in the Appendix. Despite the high level of missing data it seems unlikely that this number will fluctuate a great deal.

Caution is warranted when interpreting the vote count. It is common for authors to engage in directional null hypothesis significance testing (NHST), even when it is clearly not warranted. As an example, concept level assessment in PBL should be non-directional since the existing set of findings is so close. If authors have a directional hypothesis they would never report a statistically significant negative finding, they would simply report no significant results. Thus, it is possible that the number of negative PBL findings is under reported here.

A final obvious limitation is the use of meta-analysis as a whole. For the most part, these are described well elsewhere (Cooper & Hedges, 1994b; Glass, 2000, 2006). Broadly speaking, this is just one of many possible perspectives on a much larger body of literature. A great deal of caution should be noted due to the complete exclusion of qualitative work here. That said, past qualitative findings have been quite favorable both in terms of deep understanding (Fyrenius, Silén, & Wirell, 2007) and generating new knowledge (Pearson, 2006).

Future Work. Clearly, more work needs to be done. The lack of homegeneity found across studies as a whole is certainly exciting in a meta-analysis because it warrants additional parsing of the data. However, after parsing the data several ways no homogeneity among findings was discovered. This may be due in part to the limitations described above or the need for additional research studies in underrepresented disciplines. It could also be the result of unrepresented problem types, the need to classify the vast bulk of the outcomes according to PBL method or perhaps a combination of all of these. In part, some of these may be addressed in the next phase of this review. A survey of primary literature authors is underway to ascertain their level of agreement with our characterization of their work, and to request missing information.

Dochy et al. (2003) examined methodological quality of PBL studies in terms of internal threats to validity. After confirming coding with primary source authors it is our intention to replicate that work and perhaps broaden it in scope. The detailed examination of the validity and reliability of instrumentation as reported in this issue (Belland, French, & Ertmer, 2009), in relation to effect size outcomes, may play a crucial role in interpreting results.

Both of the early meta-analytic reviews of PBL reported findings outside the realm of examinations or performance. These include noncognitive outcomes such as the self-reported preparation of graduates (Albanese & Mitchell, 1993), their evaluation or level of satisfaction with their programs (Albanese & Mitchell, 1993; Vernon & Blake, 1993), the level of faculty satisfaction (Albanese & Mitchell, 1993), and their academic activities—such as use of resources (Vernon & Blake, 1993). To our knowledge, none of these noncognitive outcomes have been assessed in a subsequence meta-analysis. Given the conjecture that some of the most favorable outcomes for PBL may not be strictly cognitive in nature (Albanese, 2000), future analyses should investigate some of these dimensions.

This work, however, is an examination of PBL that is focused on cognitive outcomes and includes a wide range of everything from standardized licensure exams to full problem simulations to evaluation and rating in the context of practice. Even when the scope is limited to standardized tests of concepts, PBL is able to hold its own in comparison to lecture-based approaches. Recent criticism leveled against minimally guided instruction as a whole makes the point that controlled experiments favor direct and guided approaches to instruction (Kirschner et al., 2006). To a certain extent, these results agree. As part of their criticism, Kirschner et al. cite past reviews of science and engineering, two of the least favorable disciplines for PBL. Yet these disciplines could not be characterized as favoring either direct instruction or PBL, and when a wider range of disciplines is examined there are several cases in which PBL clearly performs better in controlled experiments.

Exactly when PBL leads to the most favorable findings becomes clearer in the regression analysis. PBL may do best outside of medical education and allied health, when assessment is at the application rather than the concept level and when the intervention uses the full closed-loop approach. The R^2 value of .25, accounting for 25% of the variability in cognitive outcomes may not seem all that compelling at first. To put these numbers in perspective the GRE, when combined with other factors, only accounts for 10-12% of the variability in graduate GPA (Anderson, 2006). As a result of this meta-analysis, we know more about the conditions under which PBL performs similar to or better than lecture-based approaches. A logical next step is to investigate why interventions like a closed-loop approach to PBL, or disciplines outside the field of origin are more efficacious homes for this kind of instruction.

Acknowledgements

This work is supported in part by a new faculty grant from Utah State University and an award from the NSF DRK12 program (DUE 0554440).

References

Aaron, S., Crocket, J., Morrish, D., Basualdo, C., Kovithavongs, T., Mielke, B., et al. (1998). Assessment of exam performance after change to problem-based learning: Differential effects by question type. *Teaching and Learning in Medicine, 10*(2), 86-91.

Akinoglu, O., & Tandogan, R. (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.

Albanese, M. (2000). Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills. *Medical Education, 34*(9), 729-738.

Albanese, M., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine, 68*(1), 52-81.

Alleyne, T., Shirley, A., Bennett, C., Addae, J., Walrond, E., West, S., et al. (2002). Problem-based compared with traditional methods at the faculty of Medical Sciences, University of the West Indies: A model study. *Medical Teacher, 24*(3), 273-279.

Anderson, J. L. (2006). Predicting final GPA of graduate school students: Comparing artificial neural networking and simultaneous multiple regression. *College and University, 81*(4), 19-29.

Antepohl, W., & Hezrig, S. (1997). Problem-based learning supplementing in the course of basic pharmacology-results and perspectives from two medical schools. *Naunyn-Schmiedeberg's Archives of Pharmacology, 355*(18), 43.

Antepohl, W., & Hezrig, S. (1999). Problem-based learning versus lecture based-learning in a course of pharmacology: A controlled, randomized study. *Medical Education, 33*(2), 106-113.

Barrows, H. S. (1986). A taxonomy of problem-based learning methods. *Medical Education, 20*(6), 481-486.

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New directions for teaching and learning*(68), 3-12.

Barrows, H. S. (2002). Is it Truly Possible to Have Such a Thing as dPBL? *Distance Education, 23*(1), 119-122.

Barrows, H. S., Myers, A., Williams, R. G., & Moticka, E. J. (1986). Large group problem-based Learning: A possible solution for the '2 sigma problem'. *Medical Teacher, 8*(4), 325-331.

Barrows, H. S., & Tamblyn, R. M. (1976). An evaluation of problem-based learning in small groups utilizing a simulated patient. *Journal of Medical Education, 51*(1), 52-54.

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education*. New York: Springer Publishing Company.

Beachey, W. D. (2004). *A comparison of problem-based learning and traditional curricula in baccalaureate respiratory therapy education.* Unpublished dissertation, The University of North Dakota.

Belland, B., French, B., & Ertmer, P. A. (2009). Validity and problem-based learning research: A review of the instruments used to assess intended learning outcomes. *Interdisciplinary Journal of Problem Based Learning, 3*(1), 59-89.

Berkson, L. (1993). Problem-based learning: have the expectations been met? *Academic Medicine, 68*(10 Suppl.), S79-88.

Blake, R., & Parkison, L. (1998). Faculty Evaluation of the Clinical Performances of Students in a Problem-Based Learning Curriculum. *Teaching & Learning in Medicine, 10*(2), 69-73.

Block, S. D., & Moore, G. T. (1994). Project Evaluation. In D. C. Tosteston, S. J. Adelstein & S. T. Carver (Eds.), *New pathways to medical education: Learning to learn at Havard Medical School* (pp. 114-122). Cambridge, MA: Harvard University Press.

Boshuizen, H. P. A., Schmidt, H. G., & Wassamer, A. (1993). Curriculum style and the integration of biomedical and clinical knowledge. In P. A. J. Bouhuys, H.G.Schmidt & J. H, M.van Berkel (Eds.), *Problem-based learning as an educational strategy* (pp. 33-41). Maastricht: Network Publications.

Bouchard, G. J. (2004). *The effect of a problem-based learning curriculum on performance on the Physician Assistant National Certifying Examination.* Unpublished doctoral dissertation, Seton Hall University, College of Education and Human Services.

Bovee, M., & Gran, D. (2000). Comparison of two teaching methods in a chiropractic clinical science course. *Journal Of Allied Health, 29*(3), 157-160.

Bushman, B. J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 193-213). New York: Russell Sage Foundation.

Ceconi, A. (2006). *Influence of problem-based learning instruction on decision-making skills in respiratory therapy students.* Unpublished Dissertation, Seton Hall University.

Chan, D. H., Leclair, K., & Kaczorowski, J. (1999). Problem-based small-group learning via the Internet among community family physicians: a randomized controlled trial. *Computers In Medical Practice, 16*(3), 54-58.

Chang, C. (2001). Comparing the impacts 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.

Cheaney, J. (2005). *Problem-based learning in an on-line biotechnology course.* Unpublished dissertation, Iowa State University.

Cheaney, J., & Ingebritsen, T. (2005). Problem-based learning in an online course: A case study. *International Review of Research in Open and Distance Learning, 6*(3), 1-18.

Claessen, H. F., & Boshuizen, H. P. (1985). Recall of medical information by students and doctors. *Medical education, 19*(1), 61-67.

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

Colliver, J. A. (2000). Effectiveness of problem-based learning curricula: research and theory. *Academic Medicine, 75*(3), 259-266.

Cooper, H. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.

Cooper, H., & Hedges, L. (1994a). Research synthesis as a scientific enterprise. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 3-14). New York: Russell Sage Foundation.

Cooper, H., & Hedges, L. (1994b). *The Handbook of research synthesis*. New York: Russell Sage Foundation.

Coulson, R. L. (1983). Problem-based student-centred leaming of the cardiovascular system using the problem-based learning module (P.B.L.M.). *The Physiologist, 26*(4), 220-224.

Derry, S., Hmelo-Silver, C., Nagarajan, A., Chernobilsky, E., & Beitzel, B. (2006). Cognitive transfer revisited: Can we exploit new media to solve old problems on a large scale? *Journal of Educational Computing Research, 35*(2), 145-162.

Distlehorst, L. H., & Robbs, R. S. (1998). A comparison of problem-based learning and standard curriculum students: Three years of retrospective data. *Teaching and Learning in Medicine, 10*(3), 131-137.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and instruction : the journal of the European Association for Research on Learning and Instruction, 13*(5), 533-568.

Dods, R. F. (1997). An action research study of the effectiveness of problem-based learning in promoting the acquisition and retention of knowledge. *Journal for the Education of the Gifted, 20*(4), 423-437.

Doig, K., & Werner, E. (2000). The marriage of lecture-based curriculum and problem-based learning: Are the offspring vigorous? *Medical Teacher, 22*(2), 173-178.

Doucet, M. D., Purdy, R. A., Kaufman, D. M., & Langille, D. B. (1998). Comparison of problem-based learning and lecture format in continuing medical education on headache diagnosis and management. *Medical education, 32*(6), 590-596.

Dyke, P., Jamrozik, K., & Plant, A. (2001). A randomized trial of a problem-based learning approach for teaching epidemiology. *Academic Medicine, 76*(4), 373-379.

Eisenstaedt, R. S., Barry, W. E., & Glanz, K. (1990). Problem-based learning: Cognitive retention and cohort traits of randomly selected participants and decliners. *Academic Medicine, 65*(9 (supplement)), S11-S12.

Enarson, C., & Cariaga-Lo, L. (2001). Influence of curriculum type on student performance in the United States Medical Licensing Examination Step 1 and Step 2 exams: problem-based learning vs. lecture-based curriculum. *Medical Education, 35*(11), 1050-1055.

Farquhar, J., Haf, J., & Kotabe, K. (1986). Effect of two preclinical curricular on NMBE part 1 examination performance. *Journal of Medical Education, 61*(5), 368-373.

Farr, C., Ownbey, S., Branson, D., Cao, H., & Starr, C. (2005). Multimedia and Problem-Based Instruction in the Textiles Laboratory. *Clothing & Textiles Research Journal, 23*(4), 360-367.

Finch, P. M. (1999). The effect of problem-based learning on the academic performance of students studying podiatric medicine in Ontario. *Medical Education, 33*(6), 411-417.

Fyrenius, A., Silén, C., & Wirell, S. (2007). Students' conceptions of underlying principles in medical physiology: an interview study of medical students' understanding in a PBL curriculum. *Advances in Physiology Education, 31*(4), 364-369.

Gallagher, S., & Stepien, W. (1996). Content acquisition in problem-based learning: Depth versus breadth in American Studies. *Journal for the Education of the Gifted, 19*(3), 257-275.

Geertsma, R. H., Meyerowitz, S., Salzman, L. F., & Donovan, J. C. (1977). An independent study program within a medical curriculum. *Journal of Medical Education, 52*(2), 123-132.

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.

Gilhooly, K. J. (1990). Cognitive psychology and medical diagnosis. *Applied Cognitive Psychology, 4*(4), 261-272.

Glass, G. V. (1999, July). *Meta-analysis at 25*. Paper presented to Office of Special Education Programs Research Project Directors' Conference, U.S. Department of Education. Washington D.C.

Glass, G. V. (2006). Meta-Analysis: The quantitative synthesis of research findings. In J. Green, G. Camilli & P. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 427-438). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Goodman, L. J., Brueschke, E. E., Bone, R. C., Rose, W. H., Williams, E. J., & Paul, H. A. (1991). An experiment in medical education: A critical analysis using traditional criteria. *JAMA, 265*(18), 2373-2375.

Gordon, P. R., Rogers, A. M., Comfort, M., Gavula, N., & McGee, B. P. (2001). A taste of problem-based learning increases achievement of urban minority middle-school students. *Educational Horizons, 79*(4), 171-175.

Gulseçen, S., & Kubat, A. (2006). Teaching ICT to teacher candidates using PBL: A qualitative and quantitative evaluation. *Educational Technology & Society, 9*(2), 96-106.

Heale, J., Davis, D., Norman, G., Woodward, C., Neufeld, V., & Dodd, P. (1988, November). *A randomized controlled trial assessing the impact of problem-based versus didactic teaching methods in CME.* Paper presented at the Annual Conference on Research in Medical Education Conference on Research in Medical Education.

Herring, S. M., & Evans, R. (2005, December). *The effects of problem-based learning on students understanding of animal behavior. Studies in Teaching: 2005 Research Digest.* Paper presented at the Presented at Annual Research Forum, Winston-Salem, North Carolina.

Hesterberg, L. J. (2005). *Evaluation of a problem-based learning practice course: Do self-efficacy, critical thinking, and assessment skills improve?* Unpublished dissertation, University of Kentucky.

Hmelo, C. E. (1998). Problem-based learning: Effects on the early acquisition of cognitive skill in medicine. *Journal of the Learning Sciences, 7*(2), 173-236.

Hmelo, C. E., Gotterer, G. S., & Bransford, J. D. (1997). A theory-driven approach to assessing the cognitive effects of PBL. *Instructional Science, 25*(6), 387-408.

Hmelo-Silver, C. E., & Barrows, H. S. (2006). Goals and strategies of a problem-based learning facilitator. *The interdisciplinary Journal of Problem-based Learning, 1*(1), 21-39.

Hoffman, K., Hosokawa, M., Blake Jr., R. L., Headrick, L., & Johnson, G. (2006). Problem-based learning outcomes: Ten years of experience at the University of Missouri-Columbia School of Medicine. *Academic Medicine, 81*(7), 617-625.

Johnson, S. D., Flesher, J. W., Jehng, J.-C. J., & Ferej, A. (1993). Enhancing electrical troubleshooting skills in a computer-coached practice environment. *Interactive Learning Environments, 3*(3), 199-214.

Jonassen, D. H. (2000). Toward a Design Theory of Problem Solving. *Educational Technology Research and Development, 48*(4), 63-85.

Jones, J. W., Beiber, L. L., Echt, R., Scheifley, V., & Ways, P. O. (1984). A Problem-based curriculum, ten years of experience. In H. G. Schmidt & M. L. d. Volder (Eds.), *Tutorials in problem-based learning. New direction in training for health professions.* Assen, Netherlands: Van Gorcum.

Kalaian, H. A., Mullan, P. B., & Kasim, R. M. (1999). What can studies of problem-based learning tell us? Synthesizing and modeling PBL effects on National Board of medical Examination Performance: Hierarchical linear modeling meta-analytic approach. *Advances in Health Sciences Education, 4*(3), 209-221.

Kaufman, A., Mennin, S., Waterman, R., Duban, S., Hansbarger, C., Silverblatt, H., et al. (1989). The New Mexico experiment: Educational innovation and institutional change. *Academic Medicine, 64*(6), 285-294.

Kaufman, D. M., & Mann, K. V. (1998). Comparing achievement on the Medical Council of Canada Qualifying Examination part I of students in conventional and problem-based learning curricula. *Academic Medicine, 73*(11), 1211-1213.

Kennedy, S. J. (2007). *Learning and transfer compared in two teaching methods: Online problem-based learning and the traditional lecture method.* Unpublished doctoral dissertation, Capella University.

Kirschner, P., Sweller, J., & Clark, R. (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.

LeJeune, N. (2002). *Problem-based learning instruction versus traditional instruction on self-directed learning, motivation, and grades of undergraduate computer science students.* Unpublished doctoral dissertation, University of Colorado at Denver.

Lewis, K. E., & Tamblyn, R. M. (1987). The problem-based learning approach in Baccalaureate nursing education: How effective is it? *Nursing Papers, 19*(2), 19-26.

Lieux, E. M. (1996). A comparative study of learning in lecture versus problem-based format. *About Teaching, 50*, 25-27.

Login, G., Ransil, B., Meyer, M., Truong, N., Donoff, R., & McArdle, P. (1997). Assessment of preclinical problem-based learning versus lecture-based learning. *Journal of Dentistry Education, 61*(6), 473-479.

Lyons, E. B. (2006). *Examining the effects of problem-based learning on the critical thinking skills of associate degree nursing students in a Southeastern community college.* Unpublished doctoral dissertation, Mississippi State University.

Martenson, D., Ericksson, H., & Ingelman-Sundberg, M. (1985). Evaluation of active and problem oriented teaching methods. *Medical education, 19*, 34-42.

Matthews, B. (2004). *The effects of direct and problem-based learning instruction in an undergraduate introductory engineering graphics course.* Unpublished doctoral dissertation, North Carolina State University.

Maxwell, N., Mergendoller, J., & Bellisimo, Y. (2005). Problem-based learning and high school macroeconomics: A comparative study of instructional methods. *Journal of Economic Education, 36*(4), 315-331.

McGee, M. R. (2003). *A comparison of traditional learning and problem-based learning in pharmacology education for athletic training students.* Unpublished doctoral dissertation, The University of North Carolina at Greensboro.

Mennin, S. P., Friedman, M., Skipper, B., Kalishman, S., & Snyder, J. (1993). Performances on NMBE I, II, III by medical students in the problem-based learning and conventional tracks as the University of New Mexico. *Academic Medicine, 68*(8), 616-624.

Mergendoller, J., Maxwell, N., & Bellisimo, Y. (2000). Comparing problem-based learning and traditional instruction in high school economics. *Journal of Educational Research, 93*(6), 374-382.

Mergendoller, J., Maxwell, N., & Bellisimo, Y. (2006). The effectiveness of problem-based instruction: A comparative study of instructional methods and student characteristics. *The Interdisciplinary Journal of Problem-based Learning, 1*(2), 49-69.

Merrill, M. D. (1983). Component display theory. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status* (pp. 279-333). Hillsdale, NJ: Lawrence Erlbaum Associates.

Moore, G. T., Block, S. D., Briggs-Style, C., & Mitchell, R. (1994). The influence of the new pathway curriculum on Havard medical students. *Academic Medicine, 69*(12), 983-989.

Moore-West, M., & O'Donnell, M. J. (1985). Program evaluation. In A. Kaufman (Ed.), *Implementing problem-based medical education: Lessons from successful innovations* (pp. 180-206).

Murray-Harvey, R., & Slee, P. (2000, December 4-6). *Problem based learning in teacher education: Just the beginning!* Paper presented at the Australian Association for Research in Education, Sydney.

Neufeld, V. R., & Barrows, H. S. (1974). The "McMaster Philosophy": an approach to medical education. *Journal of medical education, 49*(11), 1040-1050.

Patel, V. L., Groen, G. J., & Norman, G. R. (1991). Effects of conventional and problem-based medical curricula on problem solving. *Academic Medicine, 66*, 380-389.

Pearson, J. (2006). Investigating ICT using problem-based learning in face-to-face and online learning environments. *Computers & Education 47*(1), 56-73.

Phelan, S. T., Jackson, J. R., & Berner, E. S. (1993). Comparison of problem-based and traditional education on student performance in the obstetrics and gynecology clerkship. *Obstetrics and Gynecology, 82*(1), 159-161.

Polanco, R., Calderon, P., & Delgado, F. (2001). *Effects of a problem-based learning program on engineering students' academic achievements, skill development and attitudes in a Mexican university*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Polglase, R. F., Parish, D. C., Buckley, R. L., Smith, R. W., & Joiner, T. A. (1989). Problem-based ACLS instruction: A model approach for undergraduate emergency medical education. *Annals of Emergency Medicine, 18*(9), 997-1000.

Prince, K., van Mameren, H., Hylkema, N., Drukker, J., Scherpbier, A., & Van Der Vleuten, C. P. M. (2003). Does problem-based learning lead to deficiencies in basic science knowledge? An empirical case on anatomy. *Medical Education, 37*(1), 15-21.

Rich, K. R., Keim, R. G., & Shuler, C. F. (2005). Problem-based learning versus traditional educational methodology: a comparison of preclinical and clinical periodontics performance. *Journal of Dental Education, 69*(9), 649-662.

Richards, B. F., Ober, K. P., Cariaga-Lo, L., Camp, M. G., Philip, J., McFarlane, M., et al. (1996). Rating of students' performances in a third-year internal medicine clerkship: A comparison between problem-based and lecture-based curricula. *Academic Medicine, 71*(2), 187-189.

Robertson, C. L. (2005). *Development and transfer of higher order thinking skills in pilots*. Unpublished dissertation, Capella University.

Santos-Gomez, L., Kalishman, S., Rezler, A., Skipper, B., & Mennin, S. P. (1990). Residency performance of graduates from a problem-based and conventional curriculum. *Medical Education, 24*(4), 366-375.

Saunders, N. A., Mcintosh, J., Mcpherson, J., & Engel, C. E. (1990). A comparison between University of New Castle and University of Sydney final-year students: Knowledge of competence. In Z. H. Nooman, H. G. Schmidt & E. S. Ezzat (Eds.), *Innovation in medical education: An evaluation of its present status* (pp. 50-54). New York: Springer.

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

Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology, 35*(5), 31-38.

Saye, J. W., & Brush, T. (1999). Student engagement with social issues in a multimedia-supported learning environment. *Theory and Research in Social Education, 27*(4), 472-504.

Schmidt, H. G., Machiels-Bongaerts, M., Hermens, H., ten Cate, T. J., Venekamp, R., & Boshuizen, H. P. A. (1996). The development of diagnostic competence: Comparison of a problem-based, an integrated, and a conventional medical curriculum. *Academic Medicine, 71*(6), 658-664.

Schwartz, R. W., Donnelly, M. B., Nash, P. P., Johnson, S. B., Young, B., & Griffen, W. O., Jr. (1992). Problem-based learning: An effective educational method for a surgery clerkship. *The Journal of Surgical Research, 53*(4), 326-330.

Schwartz, R. W., Donnelly, M. B., Nash, P. P., & Young, B. (1992). Developing students' cognitive skills in a problem-based surgery clerkship. *Academic Medicine, 67*(10), 694-696.

Schwartz, R. W., Donnelly, M. B., Sloan, D. A., & Young, B. (1994). Knowledge gain in a problem-based surgery clerkship. *Academic Medicine, 69*(2), 148-151.

Sevening, D., & Baron, M. (2002). A comparison of traditional teaching methods and problem-based learning in an addiction studies class. *Journal of Teaching in the Addictions, 1*(2), 27-42.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261-281). New York: Russell Sage Foundation.

Shelton, J. B., & Smith, R. F. (1998). Problem-based learning in analytical science undergraduate teaching. *Research in Science and Technological Education, 16*(1), 19-30.

Shin, J., Haynes, B., & Johnston, M. (1993). Effect of problem-based, self-directed undergraduate education on life-long learning. *Journal of the Canadian Medical Association, 148*(6), 969-976.

Shoffner, M. B., & Dalton, D. W. (1998). *Effects of Problem-Based, Networked Hypermedia, and Cooperative Strategies on Visual Literacy Instruction*. Paper presented at the National Convention of the Association for Educational Communications and Technology (AECT), St. Louis, MO, February 18-22.

Shuler, C. F., & Fincham, A. G. (1998). Comparative achievement on National Dental Board Examination Part I between dental students in problem-based learning and traditional educational tracks. *Journal of Dental Education, 62*(9), 666-670.

Smits, P., de Buisonje, C. D., Verbeek, J. H., van Dijk, F. J., Metz, J. C., Cate, O. J. (2003). Problem-based learning versus lecture-based learning in postgraduate medical education. *Scandinavian Journal of Work, Environment & Health, 29*(4), 280-287.

Son, B., & Van Sickle, R. (2000). Problem-solving instruction and students' acquisition, retention and structuring of economic knowledge. *Journal of Research and Development in Education, 33*(2), 95-105.

Spaulding, W. (1969). The undergraduate medical curriculum (1969 model): McMaster University. *Canadian Medical Association Journal, 100*(14), 659-664.

Stevens, J. (1999). *Intermediate statistics: A modern approach* (2nd ed.). Mahway, NJ: Lawrence Erlbaum.

Strobel, J. & van Barneveld, A. (2009). Is PBL effective? A meta-synthesis of meta-analyses comparing problem-based learning to conventional classroom learning. *Interdisciplinary Journal of Problem Based Learning, 3*(1), 44-58.

Sugrue, B. (1993). *Specifictions for the design of problem-solving assessments in science* (No. 84.117G). Los Angeles, CA: National Center for Educational Research and Improvement.

Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem solving ability. *Educational Measurement: Issues and Practice, 14*(3), 29-36.

Tomczak, R. L. (1991). *The effects of problem-based learning on National Board scores, clinical evaluations and residency selection of medical students.* Unpublished doctoral dissertation, Drake University.

Usoh, I. I. (2003). *An investigation into the effectiveness of problem-based learning in an engineering technology program at Nashville State Technical Community College.* Unpublished doctoral dissertation, Tennessee State University.

Van Duijn, A. J. (2004). *Clinical performance of physical therapy students in problem-based, mixed-model, and traditional curricula.* Unpublished doctoral dissertation, University of Central Florida.

Verhoeven, B. H., Verwijnen, G. M., Scherpbier, A. J. J. A., Holdrinet, R. S. G., Oesburg, B., Bulte, J. A., et al. (1998). An analysis of progress test results of PBL and non-PBL students. *Medical Teacher, 20*(4), 310-316.

Vernon, D. T., & Blake, R. L. (1993). Does Problem-based Learning Work? A Meta-analysis of Evaluative Research. *Academic Medicine, 68*(7), 550-563.

Visser, Y. (2002, May). *Effects of problem-based and lecture-based instructional strategies on problem solving performance and learner attitudes in a high school genetics class.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Walton, J., Clark, D., & Glick, N. (1997). An outcomes assessment of a hybrid-PBL course in treatment planning. *Journal of Dental Education, 61*(4), 361-367.

Ward, J., & Lee, C. (2004). Teaching Strategies for FCS: Student Achievement in Problem-Based Learning Versus Lecture-Based Instruction. *Journal of Family and Consumer Sciences, 96*(1), 73-76.

Washington, E., Tysinger, J., Snell, L., & Palmer, L. (1998). Implementing Problem-based Learning in a Family Medicine Clerkship. *Family Medicine Journal, 30*(10), 720-726.

White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 41-55). New York: Russell Sage Foundation.

Whitfield, C., Mauger, E., Zwicker, J., & Lehman, E. (2002). Differences between students in problem-based and lecture-based curricula measured by clerkship performance ratings at the beginning of the third year. *Teaching and Learning in Medicine, 14*(4), 211-217.

Williams, D. C., Hemstreet, S., Liu, M., & Smith, V. D. (1998, June 10-25). *Examining how middle school students use problem-based learning software*. Paper presented at the 10th Annual World Conference on Educational Multimedia and Hypermedia & World Conference on Educational Telecommunications, Freiburg, Germany.

Willis, S. (2002). Problem-based learning in a general psychology course. *Journal of General Education, 51*(4), 282-291.

Woodward, C. A., McAuley, R. G., & Ridge, H. (1981). Unravelling the meaning of global comparative ratings of interns. *Annual Conference on Research in Medical Education, 20*, 149-154.

Yang, S. (2002). *Problem-based learning on the World Wide Web in an undergraduate kinesiology class: An integrative approach to education*. Unpublished doctoral dissertation, The University of New Brunswick

Andrew Walker is Assistant Professor in the Instructional Technology & Learning Sciences department at Utah State University. His research interests include collaborative information filtering for educational resources, technology professional development for in-service teachers, meta-analysis, and problem-based learning.

Heather Leary is a doctoral student in the Instructional Technology & Learning Sciences department at Utah State University. Her research interests include problem-based learning, educational digital libraries, professional development for in-service teachers, and open education.

Appendix A

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	*d*	*p* value
(Aaron et al., 1998), *medical education*						
Research Exam MCQ	12/17	concept	diagnosis-solution	/	=	*p* > .05
Research Exam Written	12/17	application	diagnosis-solution	/	=	*p* > .05
Year-End	121/113	concept	diagnosis-solution	/	-0.25	*p* < .05
(Akinoglu & Tandogan, 2007), *science*						
academic achievement (post-test)	25/25	/	/	/	0.64	p < .05
open-ended questions	25/25	/	/	/	=	ns
(Alleyne et al., 2002), *medical education*						
Exam 1995	20/5	/	diagnosis-solution	/	0.06	*p* = .55
Exam 1996	12/5	/	diagnosis-solution	/	-0.26	*p* = .32
Exam 1997	9/14	/	diagnosis-solution	/	0.06	*p* = .55
Exam 1998	14/10	/	diagnosis-solution	/	-0.04	*p* = .46
Exam 1999	17/23	/	diagnosis-solution	/	-0.15	*p* = .32
(Antepohl & Hezrig, 1997), *medical education*						
short essay	110/110	concept	diagnosis-solution	/	0.44	*p* = .0013
(Antepohl & Hezrig, 1999), *medical education*						
multiple-choice questions	57/55	concept	diagnosis-solution	/	-0.14	*p* = 0.4
short essay	57/55	principles	diagnosis-solution	/	0.37	*p* = 0.07
(Barrows & Tamblyn, 1976), *medical education*						
patient problem multiple choice questions (time 1)	10/10	application	diagnosis-solution	/	0.78	*p* < .05
(Beachey, 2004), *allied health*						
ELE	143/66	concept	/	/	0.21	*p* = .168
Employer Rating	36/29	application	/	/	0.11	*p* = .806
WRRT	95/69	concept	/	/	0.20	*p* = .203
(Blake & Parkison, 1998), *medical education*						
clerkship grades	82/82	/	diagnosis-solution	/	1.05	*p* < .001
(Block & Moore, 1994), *medical education*						
Clerkship Evaluations	63/62	application	diagnosis-solution	/	=	ns
NBME II	63/62	principles	diagnosis-solution	/	=	ns

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	*d*	*p* value
(Boshuizen et al., 1993), *medical education*						
Acceptable Answer Components	4/4	principles	diagnosis-solution	/	1.91	*p* < .05
Answer Relevant Terms	4/4	concept	diagnosis-solution	/	1.26	ns
(Bouchard, 2004), *science*						
critical thinking	8/4	process	story	/	-1.16	ns
problem solving	8/4	application	story	/	0.11	ns
(Bovee & Gran, 2000), *allied health*						
exam	50/54	principles	diagnosis-solution	/	-0.18	*p* = .35
(Ceconi, 2006), *allied health*						
Actual Exam DM Score	44/14	application	diagnosis-solution	/	2.43	*p* < .01
(Chan, Leclair, & Kaczorowski, 1999), *medical education*						
MCQ	11/8	concept	diagnosis-solution	/	-0.46	*p* = .33
(Chang, 2001), *science*						
application level	75/84	principles	troubleshooting	/	-0.04	*p* = .787
knowledge and comprehension level	75/84	concept	troubleshooting	/	0.41	*p* < .05
(Cheaney & Ingebritsen, 2005), *science*						
Post-Unit Exam	53/227	concept	dilemmas	/	-0.33	*p* = .016
(Cheaney, 2005), *science*						
Exam 2004	72/36	concept	dilemmas	/	-1.00	*p* < .001
Exam 2005	58/29	concept	dilemmas	/	-1.03	*p* < .001
(Coulson, 1983), *medical education*						
Exam	72/72	concept	diagnosis-solution	/	-0.56	ns
Retention Exam	72/72	concept	diagnosis-solution	/	=	ns
(Derry, Hmelo-Silver, Nagarajan, Chernobilsky, & Beitzel, 2006), *teacher education*						
post-test	126/101	application	design	/	1.21	*p* < .001
(Distlehorst & Robbs, 1998), *medical education*						
clerkship ratings	154/47	application	diagnosis-solution	closed loop	0.50	*p* = .0028
post-clerkship exam	154/47	application	diagnosis-solution	closed loop	0.30	*p* = .0703
USMLE step 1	154/47	concept	diagnosis-solution	closed loop	0.18	*p* = .271
USMLE step 2	154/47	principles	diagnosis-solution	closed loop	0.39	*p* = .0197
(Dods, 1997), *science*						
Depth of Understanding	30/30	concept	/	/	1.94	*p* < .0005
Depth of Understanding, Hybrid	30/30	concept	/	/	1.07	/

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	d	p value
(Doig & Werner, 2000), *medical education*						
USMLE step 1	/	application	/	/	+	$p < .05$
(Doucet, Purdy, Kaufman, & Langille, 1998), *medical education*						
Key Features Problems post-test	29/34	application	diagnosis-solution	/	1.21	$p = .001$
Knowledge post-test	26/21	concept	diagnosis-solution	/	0.48	$p = .05$
(Dyke, Jamrozik, & Plant, 2001), *medical education*						
Final exam	40/40	/	diagnosis-solution	/	-0.05	ns
Quiz total	40/40	/	diagnosis-solution	/	0.18	ns
Take-home test	40/40	/	diagnosis-solution	/	0.28	ns
(Eisenstaedt, Barry, & Glanz, 1990), *medical education*						
exam 1	58/32	concept	diagnosis-solution	/	-0.69	$p < .001$
exam 2	58/32	concept	diagnosis-solution	/	-0.18	ns
(Enarson & Cariaga-Lo, 2001), *medical education*						
USMLE step 1 1992	77/22	concept	diagnosis-solution	/	-0.35	ns
USMLE step 1 1993	77/22	concept	diagnosis-solution	/		ns
USMLE step 1 1994	77/22	concept	diagnosis-solution	/	-1.16	$p < .001$
USMLE step 1 1995	77/22	concept	diagnosis-solution	/	-0.67	$p < .01$
USMLE step 1 1996	77/22	concept	diagnosis-solution	/	-0.14	ns
USMLE step 1 1997	77/22	concept	diagnosis-solution	/	0.24	ns
USMLE step 1 1998	77/22	concept	diagnosis-solution	/	-0.32	ns
USMLE step 2 1992	60/17	principles	diagnosis-solution	/	-0.46	$p < .05$
USMLE step 2 1993	60/17	principles	diagnosis-solution	/		ns
USMLE step 2 1994	60/17	principles	diagnosis-solution	/	0.19	ns
USMLE step 2 1995	60/17	principles	diagnosis-solution	/	-0.35	ns
(Enarson & Cariaga-Lo, 2001), *continued*						
USMLE step 2 1996	60/17	principles	diagnosis-solution	/	-0.40	$p < .05$
USMLE step 2 1997	60/17	principles	diagnosis-solution	/	-0.14	ns
USMLE step 2 1998	60/17	principles	diagnosis-solution	/	0.38	ns
(Farquhar, Haf, & Kotabe, 1986), *medical education*						
NBME I	40/40	concept	diagnosis-solution	/	0.14	ns
(Farr, Ownbey, Branson, Cao, & Starr, 2005), *other*						
Short Term Gains, 2001	19/20	/	/	/	0.60	$p = .034$
Short Term Gains, 2002	18/20	/	/	/	0.68	$p = .020$

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	d	p value
(Finch, 1999), *medical education*						
(Provintial Registration Exam) Essay scores	26/21	application	diagnosis-solution	/	1.07	$p < .0005$
(Provintial Registration Exam) Multiple-choice scores	26/21	concept	diagnosis-solution	/	0.00	$p > .5$
(Gulseçen & Kubat, 2006), *teacher education*						
achievement	30/49	/	design	/	-0.04	/
(Gallagher & Stepien, 1996), *social science*						
Gain Scores	130/37	concept	dilemmas	/	0.53	$p = .000$
(Geertsma, Meyerowitz, Salzman, & Donovan, 1977), *medical education*						
clerkship grades 1974	16/15	application	diagnosis-solution	/	=	ns
clerkship grades 1975	24/16	application	diagnosis-solution	/	=	ns
NBME I 1973	16/15	concept	diagnosis-solution	/	=	ns
NBME I 1974	24/16	concept	diagnosis-solution	/	=	ns
(Goodman et al., 1991), *medical education*						
NBME I	501/72	concept	diagnosis-solution	/	-0.02	$p = .4$
NBME II	297/36	principles	diagnosis-solution	/	-0.07	$p = .73$
oral 1985 - factual science	12/12	concept	diagnosis-solution	/		/
oral 1985 - obtain history, summarize history, problem solving	12/12	principles	diagnosis-solution	/	0.19	/
oral 1987 - factual science	15/13	concept	diagnosis-solution	/	0.67	/
oral 1987 - obtain history, summarize history, problem solving	15/13	principles	diagnosis-solution	/	0.78	/
(Gordon, Rogers, Comfort, Gavula, & McGee, 2001), *science*						
science grades (6)	/	/	/	/	=	ns
science grades (7-8)	/	/	/	/	+	$p < .05$
(Heale et al., 1988), *medical education*						
MCQ - knowledge	/	/	diagnosis-solution	/	=	Ns
MCQ - retention	/	/	diagnosis-solution	/	=	Ns
SP exam	/	application	diagnosis-solution	/	=	Ns
(Herring & Evans, 2005), *science*						
10 minute quiz	/	/	/	/	=	Ns
(Hesterberg, 2005), *social science*						
WGCTA change scores	39/39	principles	/	/	-0.32	$p = .107$
(Hmelo, 1998), *medical education*						
accuracy (Rush)	14/19	application	diagnosis-solution	/	1.07	/
accuracy (VMS)	20/19	application	diagnosis-solution	/	0.21	/
data-driven reasoning (Rush)	14/19	process	diagnosis-solution	/	0.65	/

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	*d*	*p* value
(Hmelo, 1998), *continued*						
data-driven reasoning (VMS)	20/19	process	diagnosis-solution	/	0.32	/
findings accounted for (Rush)	14/19	principles	diagnosis-solution	/	1.07	/
findings accounted for (VMS)	20/19	principles	diagnosis-solution	/	0.15	/
hypothesis-driven reasoning (Rush)	14/19	process	diagnosis-solution	/	0.56	/
hypothesis-driven reasoning (VMS)	20/19	process	diagnosis-solution	/	0.28	/
reasoning chains (Rush)	14/19	principles	diagnosis-solution	/	1.82	/
reasoning chains (VMS)	20/19	principles	diagnosis-solution	/	0.46	/
science concepts (Rush)	14/19	concept	diagnosis-solution	/	1.29	/
science concepts (VMS)	20/19	concept	diagnosis-solution	/	0.31	/
(Hmelo et al., 1997), *medical education*						
coherence of explanation (reasoning chains)	20/20	principles	diagnosis-solution	/	0.85	$p < .01$
hypothesis-driven reasoning	20/20	process	diagnosis-solution	/	1.25	$p < .001$
use of basic science concepts	20/20	concept	diagnosis-solution	/	0.56	$p < .10$
(Hoffman, Hosokawa, Blake Jr., Headrick, & Johnson, 2006), *medical education*						
USMLE step 1 – 1997	94 total	concept	diagnosis-solution	/	=	ns
USMLE step 1 – 1998	98 total	concept	diagnosis-solution	/	=	ns
USMLE step 1 – 1999	87 total	concept	diagnosis-solution	/	0.51	$p < .01$
USMLE step 1 – 2000	87 total	concept	diagnosis-solution	/	0.51	$p < .01$
USMLE step 1 – 2001	99 total	concept	diagnosis-solution	/	0.51	ns
USMLE step 1 – 2002	94 total	concept	diagnosis-solution	/	0.49	$p < .01$
USMLE step 1 – 2003	89 total	concept	diagnosis-solution	/	0.50	$p < .01$
USMLE step 1 – 2004	90 total	concept	diagnosis-solution	/	0.50	$p < .01$
USMLE step 1 – 2005	95 total	concept	diagnosis-solution	/	0.51	ns
USMLE step 1 – 2006	82 total	concept	diagnosis-solution	/	0.52	p < .01
USMLE step 2 – 1997	96 total	principles	diagnosis-solution	/	=	ns
USMLE step 2 – 1998	96 total	principles	diagnosis-solution	/	0.48	$p < .01$
USMLE step 2 – 1999	86 total	principles	diagnosis-solution	/	0.51	$p < .01$
USMLE step 2 – 2000	87 total	principles	diagnosis-solution	/	=	ns
USMLE step 2 – 2001	99 total	principles	diagnosis-solution	/	0.47	$p < .01$
USMLE step 2 – 2002	93 total	principles	diagnosis-solution	/	0.49	$p < .01$
USMLE step 2 – 2003	89 total	principles	diagnosis-solution	/	0.50	$p < .01$
USMLE step 2 – 2004	91 total	principles	diagnosis-solution	/	0.49	$p < .01$
USMLE step 2 – 2005	89 total	principles	diagnosis-solution	/	=	ns

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	*d*	*p* value
(Jones, Beiber, Echt, Scheifley, & Ways, 1984), *continued*						
FLEX	60/142	/	diagnosis-solution	/	=	ns
NBME I	63/138	concept	diagnosis-solution	/	=	ns
NBME Medicine Shelf Test	331/170	principles	diagnosis-solution	/	=	ns
Residency Director Ratings	/	application	diagnosis-solution	/	=	ns
(D. M. Kaufman & Mann, 1998), *medical education*						
Problem solving/clinical reasoning 1996	81/84	application	diagnosis-solution	/	0.21	ns
Problem solving/clinical reasoning 1997	81/78	application	diagnosis-solution	/	0.45	ns
Total MC 1996	81/84	/	diagnosis-solution	/	0.12	ns
Total MC 1997	81/78	/	diagnosis-solution	/	0.18	ns
(A. Kaufman et al., 1989), *medical education*						
Clerkship Overall	/	application	diagnosis-solution	/	=	ns
Clinical Subscales	/	/	diagnosis-solution	/	=	ns
NBME I	/	concept	diagnosis-solution	/	-	$p < .0001$
NBME II	/	principles	diagnosis-solution	/	+	$p < .01$
(Kennedy, 2007), *allied health*						
post lecture test	16/14	concept	diagnosis-solution	/	0.94	$p < .005$
transfer	16/14	application	diagnosis-solution	/	0.54	ns
(LeJeune, 2002), *engineering*						
grade	8/8	principles	/	/	-1.33	$p = .013$
(Lewis & Tamblyn, 1987), *allied health*						
clinical	20/22	application	diagnosis-solution	/	0.23	$p = .2256$
exam	20/22	concept	diagnosis-solution	/	=	$p = .479$
(Lieux, 1996), *business*						
Pilot	45/36	principles	dilemmas	/	=	ns
Study	55/35	principles	dilemmas	/	-0.22	ns
(Login et al., 1997), *allied health*						
Diagnosis & Treatment 1992	20/18	application	diagnosis-solution	/	-0.11	/
Diagnosis & Treatment 1993	20/21	concept	diagnosis-solution	/	0.23	/
Diagnosis & Treatment 1994	20/21	concept	diagnosis-solution	/	0.25	/
Knowledge 1992	20/18	concept	diagnosis-solution	/	0.10	/
Knowledge 1993	20/21	concept	diagnosis-solution	/	0.29	/
Knowledge 1994	20/21	concept	diagnosis-solution	/	0.39	/

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	*d*	*p* value
(Lyons, 2006), *allied health*						
critical thinking assessment	27/27	principles	/	/	0.12	$p = .413$
NCLEX-RN	27/27	application	/	/	0.43	$p = .386$
(Martenson et al., 1985), *medical education*						
essay	818/1651	/	diagnosis-solution	/	0.16	$p < .00003$
short answer	818/1651	/	diagnosis-solution	/	-0.16	$p < .00003$
short answer - retention	132/94	concept	diagnosis-solution	/	0.41	$p < .001$
(Matthews, 2004), *engineering*						
knowledge gain	24/24	principles	/	/	0.58	$p = .24$
skill performance	24/24	application	/	/	-0.23	$p = .85$
(Maxwell, Mergendoller, & Bellisimo, 2005), *business*						
macroeconomics test	154/191	concept	/	/	0.17	ns
(McGee, 2003), *other*						
content	15/15	concept	/	/	0.15	ns
essay	15/15	application	diagnosis-solution	/	1.37	$p < .05$
(Mennin, Friedman, Skipper, Kalishman, & Snyder, 1993), *medical education*						
NBME I	508/167	concept	diagnosis-solution	/	-0.62	$p = .0001$
NBME II	447/144	principles	diagnosis-solution	/	0.11	$p = .29$
NBME III	313/103	application	diagnosis-solution	/	0.33	$p = .001$
(Mergendoller, Maxwell, & Bellisimo, 2000), *business*						
unit-specific content knowledge	38/75	concept	/	/	=	ns
(Mergendoller, Maxwell, & Bellisimo, 2006), *business*						
macroeconomics test	107/139	concept	/	/	0.25	$p < .05$
(Moore, Block, Briggs-Style, & Mitchell, 1994), *medical education*						
clinical reasoning	63/62	/	diagnosis-solution	/	=	ns
NBME I	61/60	concept	diagnosis-solution	/	-0.01	$p = .96$
(Moore-West & O'Donnell, 1985), *medical education*						
pediatric clerkship - applied basic science	18/18	application	diagnosis-solution	/	1.17	$p = .001$
pediatric clerkship - basic science	18/18	concept	diagnosis-solution	/	0.83	$p = .01$
pediatric clerkship - clinical skills	18/18	application	diagnosis-solution	/	1.17	$p = .001$
(Murray-Harvey & Slee, 2000), *teacher education*						
SOLO grade	73/36	application	/	/	0.28	ns

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	*d*	*p* value
(Patel et al., 1991), *medical education*						
clinical problem alone - backward reasoning	36/36	process	diagnosis-solution	/	0.88	$p < .05$
clinical problem alone - forward reasoning	36/36	process	diagnosis-solution	/	=	ns
(Phelan, Jackson, & Berner, 1993), *medical education*						
clinical gyn	228/78	application	diagnosis-solution	/	0.04	$p = .75$
clinical ob	228/78	application	diagnosis-solution	/	-0.05	$p = .69$
final exam	228/78	/	diagnosis-solution	/	0.07	$p = .48$
(Polanco, Calderon, & Delgado, 2001), *engineering*						
force concept inventory	/	/	/	/	=	$p > .43$
global critical thinking	/	/	/	/	=	$p = .20$
GPA	/	/	/	/	+	$p < .0001$
mechanics baseline test	/	/	/	/	+	$p < .002$
(Polglase, Parish, Buckley, Smith, & Joiner, 1989), *medical education*						
pass rate	18/11	/	diagnosis-solution	/	1.07	/
(Prince et al., 2003), *medical education*						
subtest with context	/	application	diagnosis-solution	/	=	ns
subtest without context	/	application	diagnosis-solution	/	=	ns
(Rich, Keim, & Shuler, 2005), *allied health*						
Clinical exam	233/134	application	diagnosis-solution	/	0.00	$p = .966$
Final	234/274	application	diagnosis-solution	/	0.22	$p = .015$
Midterm	234/274	application	diagnosis-solution	/	0.42	$p = .0001$
(Richards et al., 1996), *medical education*						
clinical ratings - differential diagnosis	364/88	application	diagnosis-solution	/	0.31	$p = .0005$
clinical ratings - factual knowledge	364/88	application	diagnosis-solution	/	0.35	$p = .0001$
clinical ratings - take history	364/88	application	diagnosis-solution	/	0.27	$p = .002$
NBME medicine shelf test	364/88	principles	diagnosis-solution	/	0.08	$p = .80$
(Robertson, 2005), *other*-no small group						
combined subscales	13/13	application	/	/	0.53	/
(Santos-Gomez, Kalishman, Rezler, Skipper, & Mennin, 1990), *medical education*						
nurse ratings	76/40	application	diagnosis-solution	/	-0.13	$p = .50$
supervisor ratings	79/41	application	diagnosis-solution	/	0.33	$p = .09$
(Saunders, Mcintosh, Mcpherson, & Engel, 1990), *medical education*						
MEQ1	243/44	principles	diagnosis-solution	/	-0.59	$p < .001$
MEQ2	243/44	principles	diagnosis-solution	/	0.32	$p < .05$

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	*d*	*p* value
(Saunders et al., 1990), *continued*						
NSW MCQs	240/45	concept	diagnosis-solution	/	-0.42	$p < .01$
Sydney MCQs	242/45	concept	diagnosis-solution	/	-0.77	$p < .001$
(Saye & Brush, 1999), *social science*						
essay - factual statements	24/20	application	dilemmas	/	0.65	$p < .05$
essay - higher order reasoning	24/20	application	dilemmas	/	0.70	/
post-test	24/20	concept	dilemmas	/	=	$p = .99$
(Schmidt et al., 1996), *medical education*						
diagnostic performance - PBL	612 total	application	diagnosis-solution	/	0.33	$p < .0001$
(Schwartz, Donnelly, Nash, & Young, 1992), *medical education*						
MC Exam	22/35	concept	diagnosis-solution	/	=	ns
MC Quiz #1	22/35	concept	diagnosis-solution	/	-0.44	$p < .05$
MC Quiz #2	22/35	concept	diagnosis-solution	/	=	ns
MC Quiz #3	22/35	concept	diagnosis-solution	/	-0.44	$p < .05$
MEE	22/35	application	diagnosis-solution	/	0.44	$p < .05$
SP Exam	22/35	application	diagnosis-solution	/	0.42	$p = .056$
(Schwartz, Donnelly, Nash, Johnson et al., 1992), *medical education*						
MEE - differential diagnosis	42/36	application	diagnosis-solution	/	+	$p < .01$
MEE - history and physical examination	42/36	application	diagnosis-solution	/	=	ns
MEE - interpreting data	42/36	application	diagnosis-solution	/	+	$p < .03$
MEE - order writing	42/36	application	diagnosis-solution	/	=	ns
MEE - ordering labs	42/36	application	diagnosis-solution	/	=	$p = .057$
MEE - patient management	42/36	application	diagnosis-solution	/	=	ns
NBME II-S	42/36	principles	diagnosis-solution	/	0.58	$p = .059$
(Schwartz, Donnelly, Sloan, & Young, 1994), *medical education*						
NBME II-Surgery Shelf Exam	66 total	principles	diagnosis-solution	/	-0.26	ns
(Sevening & Baron, 2002), *social science*						
pretest/posttest scores	/	/	/	/	=	$p = .086$
(Shelton & Smith, 1998), *other*						
year 1 pass rate	33/46	concept	strategic performance	/	0.66	/
year 2 pass rate	33/46	concept	strategic performance	/	0.40	/
(Shin, Haynes, & Johnston, 1993), *medical education*						
overall	41/41	concept	diagnosis-solution	/	0.52	$p < .01$

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	*d*	*p* value
(Shoffner & Dalton, 1998), *teacher education*						
combined raw achievement	28/28	concept & app.	design	/	0.17	/
(Shuler & Fincham, 1998), *allied health*						
NDB I	130/12	/	diagnosis-solution	/	1.33	$p < .001$
(Smits et al., 2003), *medical education*						
knowledge retention	59/59	concept	diagnosis-solution	/	-0.25	$p = 0.126$
(Son & Van Sickle, 2000), *business*						
knowledge acquisition	68/72	concept	/	/	0.45	$p = .05$
knowledge retention	65/64	concept	/	/	0.40	$p = .05$
knowledge structure - PRO	68/72	/	/	/	0.10	$p = .56$
knowledge structure - similarity	68/72	/	/	/	-0.21	$p = .29$
(Tomczak, 1991), *medical education*						
clinical knowledge	57/62	application	diagnosis-solution	/	0.05	$p = .724$
clinical skills	57/62	application	diagnosis-solution	/	0.19	$p = .193$
NBME II	57/62	principles	diagnosis-solution	/	-0.08	ns
(Usoh, 2003), *engineering*						
Knowledge	38/38	concept	troubleshooting	/	-0.18	$p = .429$
Knowledge Application	38/38	principles	troubleshooting	/	0.37	$p = .109$
(Van Duijn, 2004), *allied health*						
CPI Mixed PBL	50/40	application	/	/	0.14	ns
CPI PBL	50/20	application	/	/	0.12	ns
(Verhoeven et al., 1998), *medical education*						
MPT year 4 basic sciences	54/169	concept	diagnosis-solution	/	-0.57	$p < .01$
MPT year 4 clinical sciences	54/169	principles	diagnosis-solution	/	-0.10	ns
MPT year 4 social sciences	54/169	/	diagnosis-solution	/	-0.13	ns
MPT year 6 basic sciences	74/140	concept	diagnosis-solution	/	-0.47	$p < .01$
MPT year 6 clinical sciences	74/140	principles	diagnosis-solution	/	-0.20	ns
MPT year 6 social sciences	74/140	/	diagnosis-solution	/	0.01	ns
(Visser, 2002), *science*						
near-transfer post-test	30/30	/	/	/	-0.01	$p < .05$
(Walton, Clark, & Glick, 1997), *allied health*						
faculty ratings	20/20	application	diagnosis-solution	/	=	ns

study/discipline/outcome(s)	lecture / pbl N	assessment level	problem type	PBL method	*d*	*p* value
(Ward & Lee, 2004), *science*						
VoCATS gain	37/42	concept	/	/	0.04	ns
(Washington, Tysinger, Snell, & Palmer, 1998), *medical education*						
NBME - family medicine clerkship exam	200/400	process	diagnosis-solution	closed loop	0.76	$p < .0001$
(Whitfield, Mauger, Zwicker, & Lehman, 2002), *medical education*						
CPSS, 1 year PBL (1995)	74/25	application	diagnosis-solution	/	0.31	$p = .19$
CPSS, 1 year PBL (1996)	74/35	application	diagnosis-solution	/	0.24	$p = .24$
CPSS, 1 year PBL (1997)	90/14	application	diagnosis-solution	/	0.60	$p = .04$
CPSS, 2 year PBL (1998)	88/22	application	diagnosis-solution	/	0.06	$p = .78$
CPSS, 2 year PBL (1999)	82/22	application	diagnosis-solution	/	0.72	$p < .01$
CPSS, 2 year PBL (2000)	74/17	application	diagnosis-solution	/	0.37	$p = .16$
FK, 1 year PBL (1995)	74/25	application	diagnosis-solution	/	0.15	$p = .52$
FK, 1 year PBL (1996)	74/35	application	diagnosis-solution	/	0.18	$p = .29$
FK, 1 year PBL (1997)	90/14	application	diagnosis-solution	/	0.53	$p = .08$
FK, 2 year PBL (1998)	88/22	application	diagnosis-solution	/	0.08	$p = .70$
FK, 2 year PBL (1999)	82/22	application	diagnosis-solution	/	0.66	$p = .01$
FK, 2 year PBL (2000)	74/17	application	diagnosis-solution	/	0.67	$p = .01$
USMLE step 1	477/135	concept	diagnosis-solution	/	-0.08	$p = .40$
(Williams, Hemstreet, Liu, & Smith, 1998), *science*						
achievement	17/82	/	/	/	0.88	/
(Willis, 2002), *social science*						
presentations	44/30	/	/	/	0.35	/
(Woodward, McAuley, & Ridge, 1981), *medical education*						
1978 supervisor ratings	15/58	application	diagnosis-solution	/	0.14	$p = .272$
1979 supervisor ratings	32/57	application	diagnosis-solution	/	0.59	$p = .0035$
(Yang, 2002), *other*-no small group						
L-PBL long answer	18/19	application	diagnosis-solution	/	0.15	$p = .653$
L-PBL MCQ	18/19	concept	diagnosis-solution	/	0.80	$p = .02$
W-PBL long answer	18/15	application	diagnosis-solution	/	0.38	$p = .29$
W-PBL MCQ	18/15	concept	diagnosis-solution	/	-0.37	$p = .299$

Note. / indicates missing data.