4-2009

Expert Versus Novice Tutors: Impacts on Student Outcomes in Problem-Based Learning

Heather M. Leary
Utah State University

Andrew E. Walker
Utah State University

M. Harrison Fitt
Utah State University

Brett E. Shelton
Utah State University

Expert Versus Novice Tutors: Impacts on Student Outcomes in Problem-Based Learning

Heather M. Leary, Andrew E. Walker, M. Harrison Fitt, Brett E. Shelton
Utah State University

Please address all correspondence to:
Andrew Walker
Utah State University
Department of Instructional Technology and Learning Sciences
Logan, UT 84322
Phone: 435-797-2614
Fax: 435-797-2693
Email: andy.walker@usu.edu

Paper presented at the
Annual Meeting of the American Educational Research Association
San Diego, April 13-17, 2009

Expert versus Novice Tutors: Impacts on Student Outcomes in Problem Based Learning

Heather Leary, Andrew E. Walker, M. Harrison Fitt, Brett E. Shelton
Utah State University

The tutor is an essential part of problem based learning (PBL). However, tutor characteristics and role are inconsistent. Meta-analysis was used to investigate both the role and training of PBL tutors as moderators of student learning. Weighted effect sizes were calculated on student outcomes with a modest favorable overall effect size for PBL; a vote count shows favorable results as well. Results indicate a mixture of peers and instructors do best when compared to peers and instructors alone. Tutor training appears to make a difference by itself, but when considered with tutor background, tutor training does not appear to moderate student learning. A framework for study factors and recommendations for future work are provided.

Introduction

As an approach to learning, Problem Based Learning (PBL) begins by presenting authentic, complex, and ill-structured problems to students. Rather than lecturing, instructors act as tutors, encouraging students to acquire the information they need to solve each problem. Students typically work in small groups and are given increased control over their learning (Barrows, 1986, 1996, 2002). Experimental studies of PBL go back more than three decades (Neufeld & Barrows, 1974). Several meta-analyses have been conducted (Albanese & Mitchell, 1993; Dochy, Segers, Van den Bossche, & Gijbels, 2003; Gijbels, Dochy, Van den Bossche, & Segers, 2005; Kalaian, Mullan, & Kasim, 1999; Vernon & Blake, 1993) the most recent of which (Walker & Leary, in press) incorporates findings across diverse subject areas and educational levels (Savery, 2006; Savery & Duffy, 1995) in which PBL has been used. From this prior work we know that PBL in the aggregate has effect sizes that favor PBL to a small degree, that these differences depend in part on the kind of assessment used, and that unexplained variance in findings remains. One potential source for that variance is the instructor, also known as a facilitator or tutor, referred to throughout this manuscript as the tutor. While there are empirical efforts to examine tutor behavior, tutor training, and the role of the tutor in PBL interventions (Hmelo-Silver & Barrows, 2006; Moust, de Grave, & Gijselaers, 1990), there have been no efforts to systematically examine tutors in the context of the existing literature base.

There is general consensus that training tutors is important (Baroffio, Nendaz, Perrier, & Vu, 2007; Barrows, 1996; Bochner, Badovinac, Howell, & Karimbux, 2002; Chan, 2008; Dolmans et al., 2002; Eagle, 1992; Hmelo-Silver & Barrows, 2006) but it is unclear how effective that training may be. For example, tutors may turn small group sessions into lectures even after training (Moust et al., 1990). More research is required regarding the links between tutor training and corresponding student learning outcomes. In contrast to recommendations for training, there is a fair amount of debate about the optimal background for tutors. Some call for content experts (Barrows, 1996; Hmelo-Silver & Barrows, 2006; Schmidt, Van Der Arend, Moust, Kokx, & Boon, 1993), some argue content experts are not necessary (Barrows, 1986, 1998; Swanson, Stalenhoef-Halling, & van der Vleuten, 1990), while others claim that content experts and content novices should be used at different stages of PBL instruction (Schmidt, Van Der Arend, Kokx, & Boon, 1994). To complicate this more, some studies determine if you use a content expert you must conduct facilitation training (de Volder, 1982; Silver & Wilkerson, 1991). To address the existing gaps in the literature and lack of empirical knowledge to inform recommendations, this meta-analysis investigates the impacts of tutor background and tutor training on student learning outcomes.

Literature Review

PBL originated in the late 1960s as a response to low enrollments and general dissatisfaction with medical education at McMaster University (Barrows, 1996). From these modern origins, PBL became utilized as a pedagogical approach in several subject areas and educational contexts (Savery, 2006; Savery & Duffy, 1995). Over the years PBL has taken on varied definitions as a result of institutions altering the approach to meet their own particular needs (Barrows, 1996). Multiple definitions make characterizing precisely what PBL means difficult. Since Barrows is one of the originators of PBL in its current forms, the definition adopted for this work follows one from his more recent writing (Barrows, 2002):

- Unresolved and ill-structured problems are presented to students who generate not just multiple thoughts about the cause of the problems, but multiple thoughts on how to solve them.
- A student centered approach in which students determine what they need to learn. It is up to the learners to generate a list of the key issues for a particular problem, identify the key issues they need to know more about, and then pursue and acquire the missing knowledge.
- Tutors, typically instructors, act as facilitators and guides for learning. They initially ask students meta-cognitive questions about their problem-solving process and in subsequent sessions fade that guidance in favor of students taking on the tutor role.
- Authenticity forms the basis of problem selection, embodied by alignment to professional or "real world" practice. As such, problems are cross-disciplinary and unconstrained.
- PBL is typically undertaken in a small group setting (Barrows, 2002).

In terms of meta-analyses, there has been a great deal of prior work on the PBL literature, enough to warrant a synthesis of existing meta-analytic reviews (Barneveld & Strobel, in press). One of the major findings of this work is that effect size differences are impacted by the nature of the assessment. There are several approaches to categorizing assessments, including a subset of the Sugrue (1995) framework employed by Gijbels et al (2005) and then Walker & Leary (in press). Both of these analyses found large differences in effect size comparisons based on the underlying nature of the assessment, broken down into the concept, principles and application level. The concept level is centered on constructs, including their definition, identification, or the generation of examples. This is best described as declarative knowledge. Both of the meta-analyses mentioned above found that PBL students performed at about the same level as control students. This was not the case with principles level assessment, which covers the relationship between concepts. These principles may be rule-based or more heuristic but are generally built on an underlying probabilistic model that defines relationships between concepts. Both meta-analyses found that PBL students performed better with principles level assessments, but Gjibels et al's (2005) findings showed much greater differences. The application level assessments for both meta-analyses were in almost complete agreement, showing modest support for PBL. Application level assessments examine the learner's ability to employ concept and principle level knowledge in order to achieve a goal. A critical factor at this level is applying this knowledge in new situations. Based on these results and the work of other meta-analysists in the area of PBL (Kalaian et al., 1999) it seems clear that findings should take into account the assessment level of the measures used.

Like many components of PBL, different opinions exist about the role, function and ideal traits of tutors (Blumberg, Michael, & Zeitz, 1990). To a certain extent, these definitions evolved over time, which may explain the range of tutor characteristics that various authors advocate for and their subsequent study results (Bochner, Badovinac, Howell, & Karimbux, 2002; Kwizera, Dambisya, & Aguirre, 2001).

According to Barrows (1998), the role and function of the PBL tutor is to raise student awareness in higher cognitive thinking and question development. Hmelo-Silver and Barrows later added that tutors facilitate the collaborative construction of knowledge by students (2006, p. 21). Their role is not just to facilitate knowledge construction, but to work themselves out of a job by encouraging students to take on increasing responsibility for their own learning. This is accomplished through modeling desired behaviors, monitoring discussions, and focusing student efforts on deep and critical thinking (Hmelo-Silver & Barrows, 2008).

Throughout the literature, tutors are characterized along continua that cover content experts to content novices, and faculty members to student peers. Determining which characteristics promote student learning the best has been debated in the literature with several strong opinions but little consensus. Outside of the PBL literature, the distinction between expert and novice tutors is not clearly defined. Frequently, expert tutors take the form of subject-matter experts not necessarily trained in tutoring skills while novice tutors are characterized as peer tutors (Annis, 1983; Pata, Lehtinen, & Sarapuu, 2006; Roscoe & Chi, 2007; van Rosmalen et al., 2008). However, in the study of pedagogical learning agents and artificial intelligence, there is closer agreement with some PBL definitions of an expert tutor in that intelligent tutors are a representation of a domain expert also in possession of advanced tutoring skills (Baylor, 2000; Biswas, Leelawong, Schwartz, Vye, & Teachable Agents Grp, 2005; Cho, Kim, & Yun, 2005; de Antonio, Ramirez, Imbert, & Mendez, 2005; Kim, Baylor, & Grp, 2006; Nussbaum, Rosas, Peirano, & Cardenas, 2001).

At the inception of PBL, McMaster University promoted the idea of using a content novice tutor to keep the faculty members from reverting to old teaching habits, such as lecture (Barrows, 1996). As PBL spread from McMaster University, the definition was refined, declaring content expertise to be less important than facilitation expertise (Barrows & Tamblyn, 1980; de Volder, 1982; Dolmans et al., 2002; Eagle, 1992). Subsequent recommendations can be confusing, with some close overlaps and some direct contradictions. For example, some recommend that tutors be content experts with facilitation training (Barrows, 1996; Gilkison, 2003; H.G. Schmidt & Moust, 1995; H.G. Schmidt et al., 1993) while others similarly advocate for faculty or instructor tutors with facilitation skills (Bochner et al., 2002; H.G. Schmidt et al., 1994). In contrast, others purport that using content novices positively impacts student outcomes

(Silver & Wilkerson, 1991), especially in non cognitive areas such as self-directed learning. Still more view content novices as equally effective to content experts (de Volder, de Grave, & Gijselaers, 1985; Hendry, Phan, Lyon, & Gordon, 2002; Kwizera et al., 2001; Moust, de Volder, & Nuy, 1989; Moust & Schmidt, 1994; Park, Susarla, Cox, Silva, & Howell, 2007; Regehr et al., 1995; Steele, Medder, & Turner, 2000; Swanson et al., 1990), claiming that content expertise is at odds with good facilitation since an expert will constantly inject their content knowledge (Des Marchais, Bureau, Dumais, & Pigeons, 1992; Moust et al., 1990; Silver & Wilkerson, 1991) and suppress the student–directed design. Hmelo-Silver & Barrows (2006) confirm that tutor training is the most critical factor, and they elevate content expertise to being a bonus but not a critical element for success. Effective tutors are represented as expert learners who can model their own learning strategies. Further complicating these definitions, PBL often draws on several content areas in which no one tutor can be expertly versed. For example, a forensic accounting problem may incorporate elements of auditing, accounting, management, and criminal justice.

In the absence of clear consensus within the literature, we defined tutor background as follows. Content novices are faculty with no relevant expertise to the PBL curricula or a student or advanced peer (eg. a doctoral student teaching in a graduate course). Content experts are faculty members with expertise covering at least one of the relevant disciplines or a lecturer (e.g. a doctoral student teaching undergraduate students).

It is generally agreed that tutors should be trained in the process of PBL (Baroffio, Nendaz, Perrier, & Vu, 2007; A. M. C. Daniel, 2004; de Volder, 1982; Hmelo-Silver & Barrows, 2006; Wikerson & Hundert, 1991). This idea is supported by the literature outside of PBL showing the most effective tutors have been trained in facilitation skills (Annis, 1983; P. A. Cohen, Kulik, & Kulik, 1982; Merrill, Reiser, Merrill, & Landes, 1995; Topping, 1996). Expanding on facilitation skills, tutors should have a good grasp of the PBL process, which they impact to a great extent (Barrows, 1998; Dolmans & Ginns, 2005; Hmelo-Silver & Barrows, 2006), and understand how their role as a tutor changes during the course of a particular problem. In addition, they should have a great deal of familiarity with the problem either as a result of closely collaborating with the case designer or through co-authoring it (L. C. Chan, 2008; Davis, Nairn, Paine, Anderson, & Oh, 1992; Johansen & Bircher, 1992). There are several recommendations, some contradictory and some with more agreement with respect to tutors and their role in PBL. Despite a large volume of empirical findings and several meta-analyses (Albanese & Mitchell, 1993; Dochy et al., 2003; Gijbels et al., 2005; Kalaian et al., 1999; Vernon & Blake, 1993; Walker & Leary, in press), no systematic review exists of the literature across disciplines regarding the impact of tutor background and tutor training on PBL student outcomes. This work extends our previously reported findings (Walker & Leary, in press) by examining the impacts of the tutor on student learning outcomes.

Methods

Meta-analysis is considered by some to be a form of primary research, in which the subjects are studies rather than people (Cooper & Hedges, 1994a). The advantage of this approach is the ability to look at differences between factors that individual studies may not otherwise include (such as an exploration of instructor and peer tutors in PBL). This is accomplished by placing study outcomes on a common scale (Cohen's *d* in this case), then examining differences between the sets of studies using peer or instructor tutors. This particular analysis set out to address the following research questions about existing studies of PBL, while taking into account the types of assessment employed:

1) To what extent does the background of the tutor moderate the student learning?
2) To what extent does the training of the tutor moderate student learning?
3) To what extent is there an interaction between tutor training and tutor background that moderates student learning?

Inclusion Criteria

To be considered, studies had to report quantitative outcomes comparing a PBL treatment that included ill-structured and authentic problems, a student centered approach, and tutors acting as facilitators rather than lecturers, with a comparison or control condition. As suggested by the research questions, the outcomes had to be focused on student learning as opposed to non-cognitive assessments like student motivation, indications of self-directed learning, or reasoning process. While these are critical outcomes associated with PBL (Barrows, 1986), determining impacts on student learning is of interest to a wider audience, including skeptics.

Literature Search

Literature was obtained as reported in our previous findings (Walker & Leary, in press) with updates to account for new work. We began with primary research reported in existing meta-analyses, then used keywords obtained from each study to search Education Resources Information Center (ERIC), PsychInfo, Education Full Text, Google Scholar, Communications of the ACM, CiteSeer, and Digital Dissertations. Date restrictions of 2008-2009 were placed on searches since this literature has been well canvassed in our prior analysis. In addition to our updated search we obtained additional author referrals to other primary research articles as part of an author survey described below.

Coding Scheme

Each study was separately coded by two researchers. Discrepancies, largely due to omission rather than differences of opinion, were resolved until consensus was achieved (Stemler, 2004). More specifically, tutor background was coded as *unknown*, *peer*, *instructor*, *mixed* peer and instructor, or *automated* when computers took on the tutor role. Here, instructors were defined as tutors who held at least one degree level higher than their student population. The most frequent instructor case being medical education faculty and medical students, other more dramatic educational differences included K-12 teachers acting as tutors for elementary students. Peers consisted of tutors who were at the same educational level (for instance senior undergraduate students tutoring freshman and sophomores). Mixed and unknown are self-explanatory. This coding scheme parallels the call for faculty tutors (Bochner et al., 2002; H.G. Schmidt, 1994). The primary reason for our coding choice was that designation of instructors or faculty was the most common mechanism used in studies to characterize tutors. Thus instructors may take on a tutor role, but the terms not synonymous in this manuscript.

Tutor training was coded as *yes*, *no*, *automatic* or *unknown*. Training was initially coded as *yes* only when some form of training took place. Partway through coding primary studies, we modified our yes criteria to include prior experience with PBL facilitation since several primary studies reported experience either alongside or in place of references to training. No and unknown are self-explanatory. Automatic was coded separately since it is not meaningful to compare the training of human and computer-based tutors.

Assessment level was coded at the *concept*, *principles*, *application*, or *unknown* level based on the following. Many of the instruments in this field are well established, standardized measures (for example the USMLE) with clear ties to Sugrue's framework as established by prior meta-analysts (Gijbels et al., 2005). For other measures, such as teacher created tests information was drawn from the measure or sample questions when provided, or inferred based on descriptions within the article. When no information was available or the description was ambiguous, the assessment level was coded as unknown.

After the initial coding of studies was performed, a survey was sent to study authors. Out of 183 studies we received responses back from authors of 30 different studies. Authors from another 5 studies began the survey but then withdrew their participation, citing a lack of time, or an inability or lack of desire to find data that in some cases was 30 years old. The purpose of the survey was to ask authors to respond to our characterization of their work, supply missing information, and provide referrals to other studies. For the most part, our coding had high levels of agreement with study authors and their responses consisted largely of supplying missing information, although there were some departures. One participant who withdrew noted a different recollection of their work, and one study (with a single outcome) was dropped when the author pointed out the coded findings were a confirmatory analysis rather than a feature of the reported work.

Study outcomes were placed on the common scale of standardized mean difference (*d*). The most common formula used was the pooled estimate of the population standard deviation. On rare occasions the only available data was a non-specific p value (e.g. $p < 0.05$). These were assumed to be $p = 0.05$, an obvious underestimate of effect size (Shadish & Haddock, 1994). A positive effect size indicates findings in favor of PBL, whereas negative effect sizes favor the control students who typically experienced a more traditional lecture.

Results

Due to the wide variation in the number of participants across studies, this analysis uses effect size outcomes weighted by sample size (Cooper, 1989). Benchmarks for what constitutes a meaningful effect size difference are a bit more problematic. A great deal of debate has arisen specific to the context of problem-based learning effect sizes (Albanese, 2000; Colliver, 2000). Cohen, with the caveat that social sciences are quite complex, describes a rough framework for effect sizes with 0.2 as small or characteristic of new inquiry or poorly controlled studies, 0.5 as medium or noticeable differences, and 0.8 as large or grossly differentiated (1988). Our position is that any improvement is worth noting as long as the cost of the intervention is reasonable.

Each set of findings includes a vote count analysis alongside weighted effect sizes (Bushman, 1994). Since statistical significance testing (either positive in favor of PBL or negative in favor of a control group) are all that is required, a vote count is a more inclusive approach to confirm meta-analysis results. The vote count consists of a simple chi-square analysis of the significantly positive and significantly negative findings. While many meta-analysts recommend a vote count, some debate continues about whether or not inferential statistics are appropriate (Cooper & Hedges, 1994a; Glass, 2000, 2006). The analyses below include confidence intervals (reported at the 95% level) and a factorial ANOVA, both of which assume the encoded studies derive from a meaningfully defined population of PBL research. As discussed above, this includes quantitative research using a wide range of assessment tools, where the featured treatment uses PBL and the comparison is characterized as traditional or lecture based. The confidence intervals can be used as an indication of variance irrespective of stance on inferential statistics with meta-analysis, and the factorial ANOVA can be examined or ignored as desired. An alpha level of 0.05 was set for statistical significance for analyses original to this paper, and as a threshold for the vote count.

A total of 210 outcomes with codeable effect sizes from 84 studies were utilized in the meta-analysis. Overall the weighted effect size (d_w = 0.16, ± 0.04) was somewhat modest, approaching a small difference in favor of PBL. In the aggregate, these findings were not homogenous Q = 1170.11, justifying a further parsing of the data—although none of the break out findings were homogenous either.

To aid in all three research questions a 4x2x4 factorial ANOVA between tutor training (peer, instructor, mixed, unknown), tutor background (yes, unknown), and assessment level (concept, principles, application, unknown) was run. Automated tutors (N=3) were removed from the ANOVA since it was a characterization for both tutor background and tutor training, but were left in the descriptives (see Table 1 and Table 2). Even though assessment level was included in the ANOVA, it is not a central focus of this analysis, rather a necessary context for the tutor factors. Within this analysis, there was a main effect for assessment level $F(3, 204)$ = 7.93, p = 0.00, which matches our prior work (Walker & Leary, in press)

Before conducting the ANOVA, several high correlations were found. For example, instructors have a positive correlation with trained or experienced tutors (r = 0.48, p < 0.01). In combination with unequal cell sizes, this necessitates some decision making about the shared variance across factors. Our analysis takes the position that an ordering of effects is meaningful since tutor background, which may include established ideas about the role of instructors (Silver & Wilkerson, 1991) is in place before any PBL training occurs. Thus background was left intact, but training and then assessment level are reported only to the extent that they explain unique variation in student learning (Stevens, 1999).

Tutor Background

Research question one deals directly with the impact of tutor background on student learning. As can be seen below in Table 1, most of the effect size results are fairly close together. With the exception of cases where the tutor background is unknown, (d_w = .04) the effect sizes range between .23 and .34. This relative agreement is even more pronounced in the vote count, where findings favored PBL over control conditions with the exception of peer, mixed, and automated tutors, which failed to meet the minimum chi-square criteria of at least five observations per cell. The vote count however is not sensitive to magnitude of differences.

TABLE 1. *Tutor Background Outcomes*

Background	sig. +	sig. -	$N_{outcomes}$	d_w	CI_{Lower}	CI_{Upper}
Mixed	2	0	14	**0.33**	0.07	0.59
Automated	1	0	3	0.34	-0.50	1.18
Peer	5	0	13	**0.26**	0.03	0.49
Instructor	43 [a]	12	123	0.23	0.15	0.31
Unknown	15 [a]	9	57	**0.04**	-0.06	0.13
Total	69 [a]	20	210	0.16	0.12	0.20

[a]Significant ($p < 0.05$) sign test on the vote count analysis.

The use of peer and a mixture of peer and instructor tutors are notable in that they both did about the same as instructors alone—despite disagreement about their use in the literature. The factorial ANOVA showed no main effect for tutor background $F(3, 204) = 1.30$, $p = 0.28$. The unknown data may drastically change these results, particularly if they favored use of backgrounds with a small number of outcomes. That said use of instructors as tutors is clearly the norm among this literature. Thus, assignment of the unknowns is more likely to lower the instructor scores than any others. The close proximity of the peer and instructor findings should be interpreted with caution given the disparate number of outcomes. The end result for the first research question is mixed. The vote count and inferential statistics indicate that there is no impact based on tutor background, but the weighted effect sizes and small amount of variability as shown in the confidence intervals suggest that there is an impact, and when the background is unknown the outcomes are lower.

Tutor Training

Research question one examines the impacts of tutor training on student learning. When tutor training is considered independently, trained or experienced tutors perform close (d_w = 0.25) and slightly above the overall mean (d_w = 0.16), and automated tutors once again performed best (d_w = 0.34). In contrast, when tutor training information is unknown, students do slightly worse (d_w = 0.08). Once again, these weighted effect sizes parallel the vote count, which shows uniform agreement that PBL students tend to do better, in every case but automated which lacks the minimum number of findings.

TABLE 2. *Tutor Training Outcomes*

Training	sig. +	Sig. -	$N_{outcomes}$	d_w	CI_{Lower}	CI_{Upper}
yes	30 [a]	6	99	0.25	0.17	0.34
automated	1	0	3	0.34	-0.50	1.18
unknown	35 [a]	15	108	**0.08**	0.00	0.16
Total	69 [a]	20	210	0.16	0.12	0.20

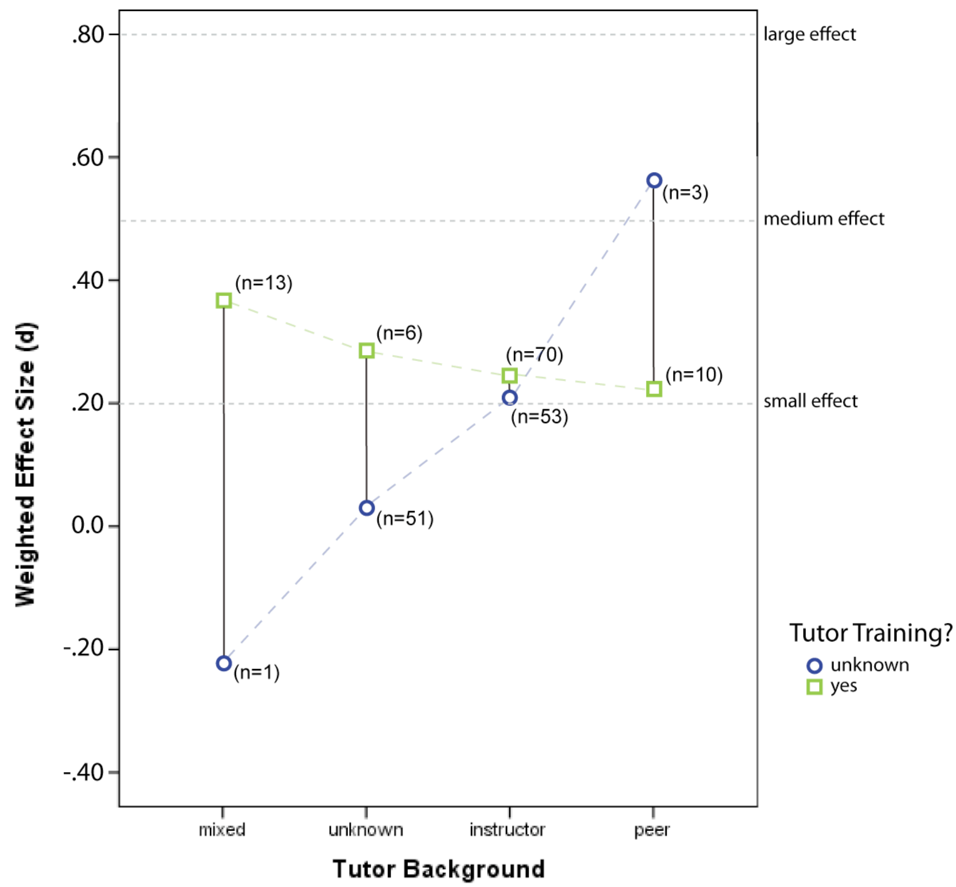
[a]Significant ($p < 0.05$) sign test on the vote count analysis.

None of the coded studies openly admitted to neglecting training for their tutors, or using tutors with prior experience. However, there are many more studies (N = 108) failing to explicitly describe training or experience tutors as compared to the number with unknown tutor background ($N = 57$). The volume of missing information is noteworthy, as is the consistently poor performance of the outcomes with unknown training or unknown background. Much like tutor background, there was no main effect for tutor training, $F(1, 206) = 0.33$, $p = 0.86$. Much like the first research question, evidence in regards to question two is mixed. The inferential statistics and vote count suggest that tutor training has no impact. The weighted effect sizes suggest that training may not improve learning outcomes substantially, but when training is unknown these outcomes appear to drop.

Interaction Effects

The third and final research question is an examination of whether or not student learning resulting from training or unknown tutor training changes with different tutor backgrounds. Figure 1 shows the weighted effect size for training and unknown tutors across the four categories of tutor background. Take note of the number of outcomes associated with each data point and keep in mind that clusters of many outcomes should be interpreted with more confidence than a handful or single outcomes. Cohen's thresholds for small, medium and large effect sizes are shown, alongside a plot of weighted effect size and the number of outcomes at each intersection of tutor training and tutor background. Visually, the results look to be a classic case of interaction with differences between unknown and trained tutors dependant upon their background. However, there is no statistically significant $F(3, 204) = 1.40$, $p = 0.25$ interaction effect. This also holds true when adding in the context of assessment level, with the three way interaction (background x training x assessment level) also failing to achieve statistical significance $F(9, 198) = 0.03$, $p = 0.992$. This is most likely due to the fact that the greatest differences, such as mixed tutor background, are derived from a small collection of outcomes.

FIGURE 1. *Student Learning as a Function of Tutor Background and Tutor Training.*

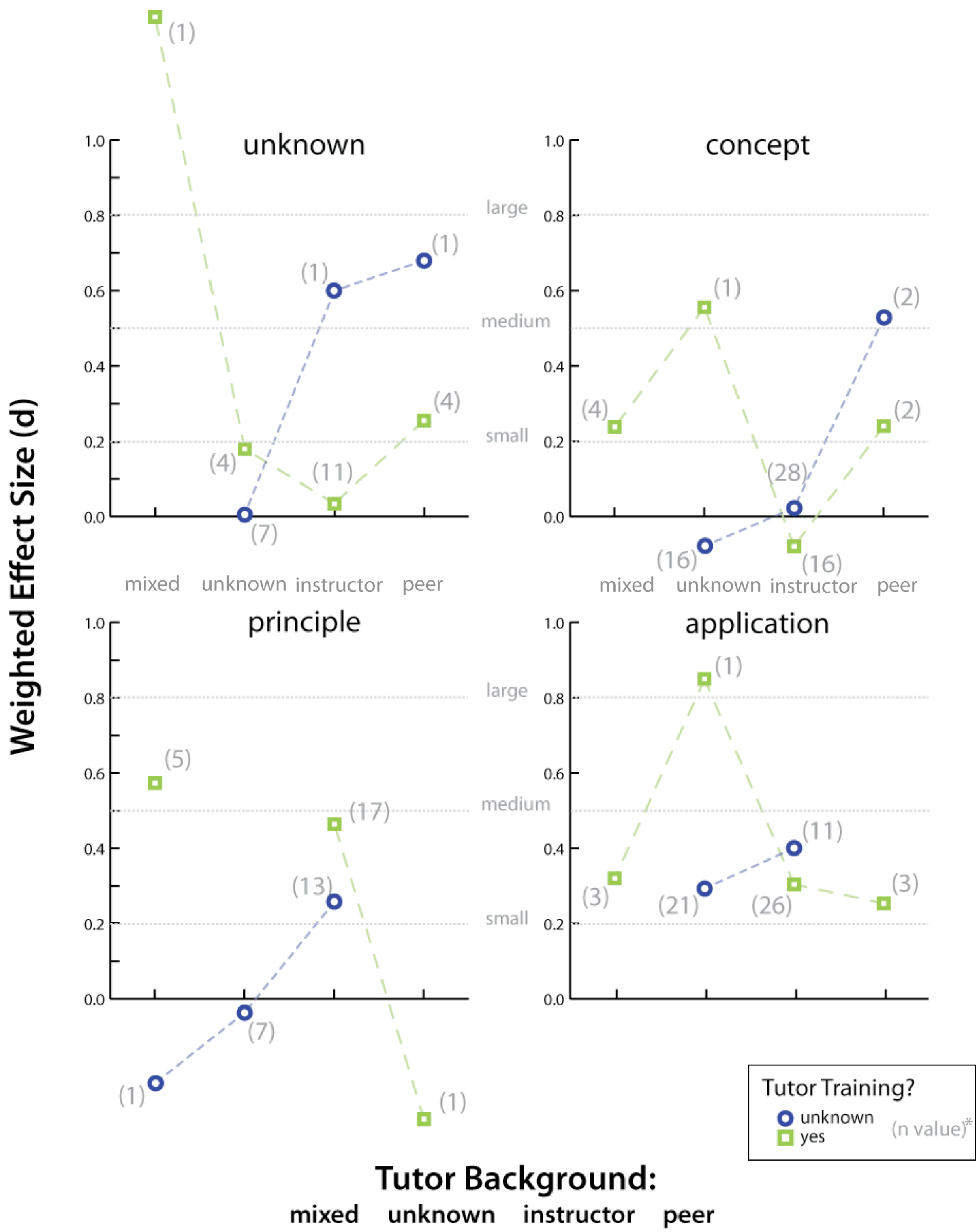


* (n value) indicates the number of study outcomes for each data point.

Still, there are obvious trends that bear discussion. When tutor training is known, the weighted effect size decrease slightly through mixed ($d_w = 0.37$), unknown ($d_w = 0.28$), instructor ($d_w = 0.24$), and peer ($d_w = 0.22$) tutor backgrounds in a nearly linear manner. Conversely, when tutor training is unknown, the weighted effect size increases dramatically from tutor backgrounds that are mixed ($d_w = -0.22$), unknown ($d_w = 0.03$), instructor ($d_w = 0.21$), then finally to peer ($d_w = 0.65$). These trends are emphasized with connecting slopes. In addition, the distance between scores within each tutor background category is shown with vertical lines. Relatively large differences exist, possibly due to the small number of studies, between weighted effect sizes for categories of *mixed* and *peer* tutor training. Less difference exists in the *unknown* category, and very little exists in the *instructor* category.

However, what is not taken into consideration in Figure 1 is the assessment level within each category of tutor background. Figure 2 offers additional insight by charting the assessment tutor background and training broken out by assessment levels of concept, principle, application, and unknown. As with Figure 1 horizontal lines act as markers within each smaller graph denoting thresholds for small, medium and large effect sizes. Once again, care should be taken to examine not only the data points, but also the number of outcomes associated with them.

FIGURE 2. *Student Learning as a Function of Tutor Background and Tutor Training Charted Against Four Categories of Assessment Strategy.*



* (n value) indicates the number of study outcomes for each data point.

The change in unknown tutor training across tutor background categories generally maintains a similar upward trend across assessment levels, but the relative consistency of outcomes with trained tutors is no longer maintained. Focusing on data points with several outcomes, some anticipated trends emerge. The concept level assessment contains several effect sizes that actually favor control students. In particular, unknown training and background, and instructors who received training. The application level assessment has consistently positive effects clustered almost exclusively in the small to medium range. One surprise is the number of principle level outcomes that are at or below an effect size of zero, the largest collection of outcomes again being unknown training and unknown background. The aggregate of these principle level outcomes is just beneath the application level findings, but the outcomes are much more distributed at the principle level.

A conservative interpretation of Figure 2 is that application level assessments appear to be robust with almost all configurations of tutor background and training. With the exception of that application level unknown training and unknown background performs poor by comparison to the overall weighted effect size ($d_w = 0.16$) although it is not at all clear why that is the case. Finally, training tutors appears to have quite diverse effects, particularly in regard to instructors where performance is equivalent to or below control conditions at the unknown and concept assessment levels and findings in the small to medium range at the principle and application assessment level. Support for research question three is mixed, but this may be due largely to the lack of outcomes in particular areas.

Conclusions and Discussion

As noted in the literature review there are advocates for using content experts (Schmidt et al., 1993), those who believe it is not necessary (Swanson et al., 1990), and those who claim content expertise is important, but secondary to other factors—specifically training in the tutor role (Hmelo-Silver & Barrows, 2006). Based on the available data, content expertise may indeed not be a factor. Both mixed and peer tutors did about as well as or perhaps better, albeit very slightly, than instructor tutors alone (see Table 1). In looking at Figure 2, these trends are somewhat stable even when broken down by assessment level. For the most part, mixed and peer tutors continue to do about as well or slightly better than instructor tutors. Notable departures come at the principle level, where peer tutors did poorly. In this case and in some of the cases where peers or mixed tutors performed markedly better, there are dramatically fewer outcomes than those involving instructors. This of course relies on a small number of outcomes (N=14 for mixed, N=13 for peers) and additional investigations are needed to confirm these results, particularly those that make a direct comparison with instructor tutors. If these findings are confirmed, scaling PBL as an intervention may be more cost effective since peers are less costly.

Additional investigations are also needed to explain why mixed and peer tutors performed similarly to instructors. Hmelo-Silver & Barrow's initial research of the tutor role was recently expanded to a rich exploration of the knowledge building process (2008) within PBL groups. The tutor in their study is both an instructor possessing robust content expertise and an expert facilitator. Some prompts from the tutor did include direct questions to focus student efforts, questions perhaps best asked by experts. However, the bulk of the tutor's efforts centered on meta-cognitive style questions that assisted self-directed learning and self-monitoring by the students. The fact that an expert and experienced tutor is focused more on facilitation skills than content may partially explain why peer tutors, who lack content knowledge when compared to instructors, are associated with such similar student learning outcomes. This also aligns with prescriptions for students to gradually take on the role of the tutor themselves (Hmelo-Silver & Barrows, 2006). Clearly, this would not be realistic or beneficial if content expertise were a focal point of the tutor role.

Based on the weighted effect sizes, tutor facilitator training or experience with facilitating slightly improves student learning as compared with the overall mean (Table 2). This is encouraging, as it is highly recommended throughout the literature (Baroffio et al., 2007; Daniel, Tosteson, Adelstein, & Carver, 1994; de Volder, 1982; Hmelo-Silver & Barrows, 2006; Wilkerson & Hundert, 1991). Also encouraging is the fact that none of the 84 studies included here, or even the 183 studies considered, explicitly stated that they used inexperienced and untrained tutors. This provides some indication that at least for tutor training, a fairly clear consensus in favor of training is being followed. That said, it is disappointing that these improvements are not at all dramatic. In fact, for instructor tutors (Figure 1) for which there is an overwhelming amount of data—there are no real differences in the weighted effect size. The main effect and vote count for training show no effect at all.

In looking at Figure 2, training differences with respect to outcome type appear much more stable when training is unknown then when training occurred. When compared to Figure 1, the trend line for unknown training follows a similar upward trend from mixed to unknown to instructor to peer background tutors at all four assessment levels. This is decidedly not the case with trained tutors, which vary widely depending up on the assessment level, perhaps due to the relatively small number of outcomes for a few extreme data points. Similarly erratic patters can be found when focusing on the breakdown of instructor tutors by assessment level. At the concept and application level, training appears to slightly detract from student learning. At the principles level, however, the students of trained instructors do markedly better. This is reversed for unknown assessment level outcomes, but with only a single outcome of unknown instructor tutor training the discrepancy should be regarded with caution. Reasons for these variations between unknown and trained instructor tutors by assessment level are unclear.

One potential reason is the wide variation in what was coded as training for this study. Recall that either formal training or experience was coded as a yes. Experience in doing PBL poorly, however, is far different from interventions that engage in thorough and sustained training of tutors. Even within the realm of training, large differences are possible. A two hour short course on PBL facilitation was coded the same as year long iterative training. These coding decisions were based primarily on the available data. Although almost half of the studies with codeable effect sizes mentioned training, they typically did not elaborate on the type or extent of that training. This is not necessarily a reflection on the quality of the articles in this area. Space is at a premium, particularly in journal articles, and it is not realistic to expect an exhaustive discussion of how tutors were prepared.

It is quite possible that with additional data, a main effect for tutor background, tutor training, or an interaction effect (background x training) might be found. In order for that to occur many more studies are needed. Despite the lack main effects, the weighted effect sizes for both training and tutor background clearly show that unknown tutor training and tutors with unknown backgrounds are not effective. Or to be more precise, studies that failed to report this information did not show student learning outcomes that were as favorable as those that did. Again, it is unclear as to why and perhaps more importantly it is unclear exactly what did occur.

Automated tutors did well, but it is unclear why this is the case. Perhaps an in-depth investigation of the tutoring and knowledge building process like those previously conducted with humans (Hmelo-Silver & Barrows, 2006, 2008) would help illuminate this gap. As a point of speculation, machine based tutors would be unable to usurp the tutor role to engage in lecture, a phenomenon that has been documented by others (Moust et al., 1990). Much like peer and mixed tutors, the use of automated tutors is based on a handful of outcomes (*N*=3) from just two studies. Assuming these results remain stable, automated tutors could be another potential source of lowering PBL costs.

Overall we need to know more about the use of mixed tutors and peers to better understand their impact on outcomes and to make meaningful comparisons with instructors. It is important for studies to report more information about the training and background of the tutors, as definitions and characteristics vary across implementations and there is variation in all of the outcomes reported here both in the aggregate and when broken down by background, training, and assessment level. The tutor is a central role in the PBL process, the training and background of these tutors are sited throughout the literature, and one clear finding from this analysis is that studies not reporting them appear to do worse.

Limitations

There are several limitations to point out. The first is a limitation of the methodology itself (Cooper & Hedges, 1994b; Glass, 2000, 2006). Meta-analysis by definition is an examination of the quantitative findings from a field of inquiry. Further, this particular analysis, in an effort not to compare disparate findings included only studies with both PBL and a control condition. Both of these decisions limit coverage of the available literature, and qualitative work in particular. Others might legitimately criticize the work as being too broad, because it includes widely disparate forms of assessment. We calculated effect sizes for everything from measures uniquely crafted for a particular study (Akinoglu & Tandogan, 2007) to standardized licensure exams (Hoffman, Hosokawa, Blake Jr., Headrick, & Johnson, 2006) to open ended essays (Saye & Brush, 1999) and performance evaluations (Woodward, McAuley, & Ridge, 1981). These are important considerations. Our position is that assuming factors like the background of the tutor and their training are important, then they should be robust and observable across existing studies.

The classification of tutor background comes with no small amount of controversy. Although it centers on the role of the tutor (as an instructor, peer, or automated tutor) and assumes that instructors have a high level of content knowledge, many of the authors responding to our survey were quick to point out that this classification is generally far from the true situation. PBL problems are intended to be cross-disciplinary and as such require a wide range of content knowledge that any one tutor will rarely posses. Our coding decision in this case emerged from the primary research studies examined, and the vast majority of them discussed faculty, teachers, or researchers as taking on the tutor role.

Only a small portion of the authors responded to our survey providing feedback and additional information on only 30 of the 86 studies. This response rate (35%) is quite low in comparison to the response rate (64%) of other author surveys (Asch, Jedrziewski, & Christakis, 1997). However, Asch et al collected data from authors of work published just a few years prior, whereas some of the work in this study was published more than two decades ago. It is possible that the willingness to respond to the survey may have been based on some form of systematic bias. An analysis of the weighted effect sizes ($d_w = .21$) from just these 30 studies is fairly similar when compared to the effect size from studies without an author response ($d_w = .15$), making a systematic bias on learning outcomes less likely.

Future Work

We currently have a detailed taxonomy of problem-based learning which differentiates between various approaches (Barrows, 1986). In addition to this mechanism for categorizing the intervention, we have a typology of problems that students can engage with (Jonassen, 2000). Still missing, despite the importance of the tutor role (Barrows, 1998) is a mechanism to succinctly describe the role of the tutor within a PBL intervention. Past work has been able to use these categories to find some differences among PBL results (Walker & Leary, in press), and it is reasonable to expect that a similar categorization of tutors might accomplish the same. A tutor classification might clearly differentiate between subject matter expertise, the tutor background, whether or not the tutor is scaffolding with the intention that students will eventually be taking on the tutor role, the kinds of prompts and questions the tutor gives, the fidelity of the PBL intervention as evidenced but the tutors' actions, and much more information about the training including the duration and whether or not it is sustained. Future PBL studies might also do well to characterize the nature of their students and their interactions with tutors. An interesting factor that has been noted in the literature (Schmidt, Van Der Arend, Moust, Kokx, & Boon, 1993; Schmidt & Moust, 1998) but has not been the primary focus of research is the level of knowledge the students bring to the discussion. Student knowledge is a determining factor in how the student uses the tutor for learning.

With this in place the researcher as well as practitioner community might be able to determine why there is so much variation in student learning outcomes. That variation may also be due to the obvious need for replication work. In particular, more needs to be known about the use of peer tutors in the context of PBL—authors have understandably shown some reluctance to this approach because it directly contradicts recommendations in the field, but the results reported here show a tendency for peers to perform about as well as instructors. An investigation into the reasons for the strong performance of automated tutors is also needed. It is perhaps less useful to investigate untrained tutors, but the ability to better characterize the training, and more importantly the actions of tutors within a PBL intervention, is vital in any effort to explain why training makes little difference, especially for instructors, when considered alongside tutor background. Finally, a great deal of work remains with respect to the kinds of outcomes examined. Our focus was on student learning, but non-cognitive outcomes like self-directed learning and increased motivation for learning were not examined. It may be that tutor training and tutor background look very different outside the context of student learning.

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.
* al-Haddad, M. K., & Jayawickramarajah, P. T. (1991). Problem-based curriculum: outcome evaluation. *Medical Teacher, 13*(4), 273-279.

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.

Annis, L. F. (1983). The processes and effects of peer tutor. *Human Learning, 2*(1), 39-47.

* 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.

Asch, D., Jedrziewski, M., & Christakis, N. (1997). Response rates to mail surveys published in medical journals. *Journal Of Clinical Epidemiology, 50*(10), 1129-1136.

Barneveld, A. V., & Strobel, J. (in press). 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).

Baroffio, A., Nendaz, M. R., Perrier, A., & Vu, N. V. (2007). Tutor training, evaluation criteria and teaching environment influence students' ratings of tutor feedback in problem-based learning. *Advances In Health Sciences Education, 12*(4), 427-439.

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. (1998). The essentials of problem-based learning. *Journal of Dental Education, 62*(9), 630-633.

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

* 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.

Baylor, A. L. (2000). Beyond butlers: Intelligent agents as mentors. *Journal of Educational Computing Research, 22*(4), 373-382.

* 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, United States -- North Dakota.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & Teachable Agents Grp, V. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*(3-4), 363-392.

* Blake, R., Hosokawa, M., & Riley, S. (2000). Student performances on Step 1 and Step 2 of the United States Medical Licensing Examination following implementation of a problem-based learning curriculum. *Academic Medicine, 75*(1), 66-70.

Blumberg, P., Michael, J. A., & Zeitz, H. (1990). Roles of Student-Generated Learning Issues in Problem-Based Learning. *Teaching and Learning in Medicine, 2*(3), 149-154.

Bochner, D., Badovinac, R. L., Howell, T. H., & Karimbux, N. Y. (2002). Tutoring in a Problem-Based Curriculum: Expert versus Nonexpert. *Journal of Dental Education, 66*(11), 1246-1251.

* 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, United States -- New Jersey.

* Bouchard, J. (2006). *Physics students' approaches to learning and cognitive processes in solving physics problems.* Unpublished doctoral dissertation, McGill University.

* 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.

Chan, L. C. (2008). The Role of a PBL Tutor: A personal perspective. *Kaohsiung Journal of Medical Sciences, 24*(3), S34-38.

* 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, United States -- Iowa.

* 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.

Cho, K., Kim, S. I., & Yun, S. H. (2005). *The design and implementation of an active peer agent providing personalized user interface* Paper presented at the Knowledge-Based Intelligent Information and Engineering Systems.

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

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational Outcomes of Tutoring: A Meta-analysis of Findings. *American Educational Research Journal, 19*(2), 237-248.

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 leaming module (P.B.L.M.). *The Physiologist, 26*(4), 220-224.

Daniel, A. M. C. (2004). Problem Based Learning Curriculum for Clinical Laboratory Science. Unpublished Dissertation. University of Houston.

Daniel, C., Tosteson, S., Adelstein, J., & Carver, S. T. (1994). *New pathways to medical education : learning to learn at Harvard Medical School*. Cambridge, Mass: Harvard University Press.

Davis, W. K., Nairn, R., Paine, M. E., Anderson, R. M., & Oh, M. S. (1992). Effects of Expert and Non-Expert Facilitators on the Small-group Process and on Student Performance. *Academic Medicine, 67*(7), 470-474.

de Antonio, A., Ramirez, J., Imbert, R., & Mendez, G. (2005). *Intelligent virtual environments for training: An agent-based approach* Paper presented at the Multi-Agent Systems and Applications Iv.

de Volder, M. L. (1982). Discussion groups and their tutors: Relationships between tutor characteristics and tutor functioning. *Higher Education, 11*(3), 269-271.

de Volder, M. L., de Grave, W. S., & Gijselaers, W. (1985). Peer teaching: Academic achievement of teacher-led versus students-led discussion groups. *Higher Education, 14*(6), 643-650.

* Derry, S., Hmelo-Silver, C. E., 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.

Des Marchais, J. E., Bureau, M. A., Dumais, B., & Pigeons, G. (1992). From traditional to problem-based learning: a case report of complete curriculum reform. *Medical Education, 26*(3), 190-199.

* 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.

Dolmans, D. H. J. M., Gijselaers, W. H., Moust, J. H. C., De Grave, W. S., Wolfhagen, I. H. A. P., & Van Der Vleuten, C. P. M. (2002). Trends in research on the tutor in problem-based learning: conclusions and implications for educational practice and research. *Medical Teacher, 24*(2), 173-180.

Dolmans, D. H. J. M., & Ginns, P. (2005). A short questionnaire to evaluate the effectiveness of tutors in PBL: validity and reliability. *Medical Teacher, 27*(6), 534-538.

* 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.

Eagle, C. J. (1992). Effects of tutors with case expertise on problem-based learning issues. *Academic Medicine, 67*(7), 465-469.

* 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.

* 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.

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.

Gilkison, A. (2003). Techniques used by 'expert' and 'non-expert' tutors to facilitate problem-based learning tutorials in an undergraduate medical curriculum. *Medical Education, 37*, 6-14.

Glass, G. V. (2000). Meta-analysis at 25. Paper presented at the UC-Berkeley-Stanford Colloquium on Meta-Analysis. Retrieved March 17, 2008, from http://glass.ed.asu.edu/gene/papers/meta25.html

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.

Hendry, G. D., Phan, H., Lyon, P. M., & Gordon, J. (2002). Student evaluation of expert and non-expert problem-based learning tutors. *Medical Teacher, 24*(5), 544-549.

* 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, United States -- 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.

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building *Cognition and Instruction, 26*(1), 48-94.

* 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.

Johansen, M. L., & Bircher, D. F. M. (1992). Students as tutors in problem-based learning: does it work? *Medical Education, 26*(2), 163-165.

* 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.

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, 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.

Kim, Y., Baylor, A. L., & Grp, P. (2006). Pedagogical agents as learning companions: The role of agent competency and type of interaction. *Educational Technology Research and Development, 54*(3), 223-243.

Kwizera, E. N., Dambisya, Y. M., & Aguirre, J. H. (2001). Does tutor subject-matter expertise influence student achievement in the problem-based learning curriculum at UNITRA Medical School? *South African Medical Journal, 91*(6), 514-516.

* LeJeune, N. (2002). *Problem-based learning instruction versus traditional instruction on self-directed learning, motivation, and grades of undergraduate computer science students.* University of Colorado at Denver, Denver, CO.
* 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 Denistry 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, United States -- North Carolina.
* 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, United States -- North Carolina.
* 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. (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, D. C., Reiser, B. J., Merrill, S. K., & Landes, S. (1995). Tutoring: Guided Learning by Doing. *Cognition and Instruction, 13*(3), 315-372.

* 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).

Moust, J. H. C., de Grave, W. S., & Gijselaers, W. H. (1990). The tutor role: A neglected variable in the implementation of problem-based learning. In Z. H. Nooman, H. G. Schmidt & E. S. Ezzat (Eds.), *Innovation in medical education: An evaluation of its present status* (pp. 135-151). New York: Springer.

Moust, J. H. C., de Volder, M. L., & Nuy, H. J. P. (1989). Peer teaching and higher level cognitive learning outcomes in problem-based learning. *Higher Education, 18*(6), 737-742.

Moust, J. H. C., & Schmidt, H. G. (1994). Facilitating small-group learning: a comparison of student and staff tutors' behavior. *Instructional Science, 22*(4), 287-301.

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

Nussbaum, M., Rosas, R., Peirano, I., & Cardenas, F. (2001). Development of intelligent tutoring systems using knowledge structures. *Computers & Education, 36*(1), 15-32.

Park, S. E., Susarla, S. M., Cox, C. K., Silva, J. D., & Howell, T. H. (2007). Do tutor expertise and experience influence students performance in a problem-based curriculum? . *Journal of Dental Education, 71*(6), 819-824.

Pata, K., Lehtinen, E., & Sarapuu, T. (2006). Inter-relations of tutor's and peers' scaffolding and decision-making discourse acts. *Instructional Science, 34*(4), 313-341.

* 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.

* 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.

Regehr, G., Martin, J., Hutchinson, C., Murnaghan, J., Cusimano, M., & Reznick, R. (1995). The Effect of Tutor’s Content Expertise on Student Learning, Group Process, and Participants Satisfaction in a Problem-Based Learning Curriculum. *Teaching and Learning in Medicine, 7*(4), 225-232.

* Rich, K. R., Keim, R. G., & Shuler, C. F. (2005). Problem-based learning versus traditional educational methodology: a comparison of preclinical and clinical periodonitcs 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, United States -- Minnesota.

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*(4), 534-574.

* Ross, P. N. (2003). *The effect of technology-enhanced, problem-based instruction on student attitudes, academic engagement, and acquisition and retention of knowledge.* Unpublished doctoral dissertation, George Mason University, United States -- Virginia.

* 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. (1994). Resolving inconsistencies in tutor expertise research: Does lack of structure cause students to seek tutor guidance? . *Academic Medicine, 69*(8), 656- 662.

Schmidt, H. G., & Moust, J. H. C. (1995). *What makes a tutor effective? A structural equations modelling approach to learning in a problem-based curricula*. Paper presented at the Annual Meeting of the American Educational Research Association.

Schmidt, H. G., Van Der Arend, A., Kokx, I., & Boon, L. (1994). Peer versus staff tutoring in problem-based learning. *Instructional Science, 22*(4), 279-285.

Schmidt, H. G., van der Arend, A., Moust, J. H. C., Kokx, I., & Boon, L. (1993). Influence of tutors' subject-matter expertise on student effort and achievement in problem-based learning. *Academic Medicine, 68*(10), 784-791.

* 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.

Silver, M., & Wilkerson, L. (1991). Effects of tutors with subject expertise on the problem-based tutorial process. *Academic Medicine, 66*(5), 298-300.

* Smits, P., de Buisonje, C. D., Verbeek, J. H., van Dijk, F. J., Metz, J. C., & ten 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.

Steele, D. J., Medder, J. D., & Turner, P. (2000). A comparison of learning outcomes and attitudes in student- versus faculty-led problem-based learning: an experimental study. *Medical Education, 34*(1), 23-29.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Practical Assessment, Research & Evaluation, 9(4).

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

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

Swanson, D. B., Stalenhoef-Halling, B. F., & van der Vleuten, C. P. M. (1990). Effect of tutor characteristics on test performance of students in a problem-based curriculum. In W. Bender, R. J. Hiemstra, A. J. J. A. Scherpbier & R. P. Zwierstra (Eds.), *Teaching and assessing clinical competence* (pp. 129–134). Groningen, The Netherlands: Boek Werk Publications.

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

Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: a typology and review of the literature. *Higher Education, 32*(3), 321-345.

* 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, United States -- Tennessee.

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

van Rosmalen, P., Sloep, P., Kester, L., Brouns, F., de Croock, M., Pannekeet, K., et al. (2008). A learner support model based on peer tutor selection. *Journal of Computer Assisted Learning, 24*(1), 74-86.

* 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.

Walker, A., & Leary, H. (in press). A Problem Based Learning Meta Analysis: Differences Across Problem Types, Implementation Types, Disciplines, and Assessment Levels. *Interdisciplinary Journal of Problem Based Learning*.

* 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.
* 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.
Wikerson, L., & Hundert, E. (1991). Becoming a problem-based tutor: Increasing self-awareness through faculty development. In B. Boud & G. Feletti (Eds.), *The Challenge of Problem Based Learning*. London: Kogan Page.
Wilkerson, L., & Hundert, E. (1991). Becoming a problem-based tutor: Increasing self-awareness through faculty development. In B. Boud & G. Feletti (Eds.), *The Challenge of Problem Based Learning*. London: Kogan Page.
* 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.* The University of New Brunswick.

Studies marked with an asterisk (*) were included in the analysis.