2010; 32: 28–35 MEDICAL TEACHER

BEME RAPID REVIEW

Problem-based learning in pre-clinical medical education: 22 years of outcome research

LISA HARTLING, CAROL SPOONER, LISA TJOSVOLD & ANNA OSWALD
University of Alberta, Canada

Abstract

Purpose: To conduct a systematic review of problem-based learning (PBL) in undergraduate, pre-clinical medical education.

Methods: A research librarian developed comprehensive search strategies for MEDLINE, PSYCINFO, and ERIC (1985–2007). Two reviewers independently screened search results and applied inclusion criteria. Studies were included if they had a comparison group and reported primary data for evaluative outcomes. One reviewer extracted data and a second reviewer checked data for accuracy. Two reviewers independently assessed methodological quality. Quantitative synthesis was not performed due to heterogeneity. A qualitative review with detailed evidence tables is provided.

Results: Thirty unique studies were included. Knowledge acquisition measured by exam scores was the most frequent outcome reported; 12 of 15 studies found no significant differences. Individual studies demonstrated either improved clerkship ($N = 3$) or residency ($N = 1$) performance, or benefits on some clinical competencies during internships for PBL ($N = 1$). Three of four studies found some benefits for PBL when evaluating diagnostic accuracy. Three studies found few differences of clinical (or practical) importance on the impact of PBL on practicing physicians.

Conclusions: Twenty-two years of research shows that PBL does not impact knowledge acquisition; evidence for other outcomes does not provide unequivocal support for enhanced learning. Work is needed to determine the most appropriate outcome measures to capture and quantify the effects of PBL. General conclusions are limited by methodological weaknesses and heterogeneity across studies. The critical appraisal of previous studies, conducted as part of this review, provides direction for future research in this area.

Introduction

Following the introduction of problem-based learning (PBL) in medical curricula in the 1960s (Berkson 1993), numerous medical schools worldwide began to adopt more active learning strategies over what are considered the traditional passive methods (Norman & Schmidt 1992). Active learning is promoted in an effort to improve the quality of education and the quality of graduating physicians and the care they provide. The premise is that PBL-based strategies result in enhanced learning and performance by engaging the students through self-direction and problem-solving (Berkson 1993; Barrows 1996). In theory, PBL learning goes beyond rote memorization and simple acquisition of knowledge characteristic of passive learning strategies, to nurture clinical reasoning, team work and problem-solving which is presumed to result in enhanced deep learning, and better preparation for students' future careers (Schmidt et al. 1987; Norman & Schmidt 1992; Berkson 1993; Vernon & Blake 1993; Sivan et al. 2000).

Since the introduction of PBL, a number of narrative and systematic reviews have been published that compare it to more traditional passive approaches to medical education. Schmidt et al. (1987) published an overview of 15 select studies that compared PBL to conventional programs. The review found that PBL encouraged a student-centred and 'inquisitive style of learning.' They noted that students from conventional programs tended to perform better on 'traditional measures of academic achievement,' although the differences were small (Schmidt et al. 1987). The effect of PBL on clinical competence was weak and inconclusive. The authors discussed the many threats to internal and external validity in this

Practice points

- Despite conflicting evidence, there is a continued movement to integrate problem-based learning (PBL) into undergraduate medical education.
- Results from 30 comparative studies do not provide unequivocal support for enhanced learning through PBL.
- Sources of discrepant findings include heterogeneity in the educational curricula, study methods, and outcomes examined.
- Measures are required that reflect the outcomes that PBL seeks to effect, specifically factors that result in improved practice and quality or outcomes of patient care.
- Future research needs to address the methodological shortcomings of previous studies and follow recommendations for improved reporting.

Correspondence: Lisa Hartling, University of Alberta, Pediatrics, Aberhart Centre One, Room 9424, 11402 University Avenus, Edmonton, Alberta, t6H 5w9, Canada. Tel: 780-492-6124; fax: 780-407-6435; email: hartling@ualberta.ca

ISSN 0142–159X print/ISSN 1466–187X online/10/010028–8 © 2010 Informa Healthcare Ltd.
DOI: 10.3109/01421590903200789

body of literature that make it difficult to reach general conclusions.

Berkson (1993) published an overview of 12 studies that had been published prior to September 1992. She found no evidence of enhanced problem-solving skills, knowledge acquisition, or motivation for learning among PBL students. Moreover, she reported that PBL was costly and stressful to students and faculty. She concluded that 'the graduate of a PBL curriculum is, to date, difficult to distinguish from his or her traditional counterpart' and that any differences in postgraduate clinical performance may be related to a variety of factors other than the specific medical school curriculum (Berkson 1993).

The first two systematic reviews appeared in 1993 (Albanese & Mitchell 1993; Vernon & Blake 1993). Vernon and Blake (1993) identified 35 studies from 19 institutions, published between 1970 and 1992, and included 22 studies in the meta-analysis. While the authors found no difference in factual or clinical knowledge, and better performance on national licensing exams among the traditional students, they claimed PBL was superior to traditional methods. They based this positive conclusion on student and faculty attitudes, better student mood and attendance, better clinical performance, benefits in terms of academic process variables, and enhanced 'humanism' among PBL students. While the authors conducted a comprehensive search and clearly described their statistical methods, this review had major methodological flaws according to established guidelines to assess the quality of systematic reviews (Table 1, available at www.medicalteacher.org) (Oxman & Guyatt 1991).

The second systematic review to appear in the published literature also had major methodological flaws (Albanese & Mitchell 1993). The findings were similar to those of Vernon and Blake; however, their conclusions towards PBL were less favourable. Despite findings that PBL was 'more nurturing and enjoyable' for students, there were several factors that mitigated their conclusions. These included concerns over gaps in knowledge, reasoning processes, and higher costs and service utilization of practicing physicians from a PBL program. They suggested caution when implementing PBL and recommended a hybrid program that capitalizes on the benefits offered by both PBL and traditional curricula with a gradual increase in self-directed learning as students progress through their medical education.

Kalaian et al. (1999) published a systematic review of the effects of PBL on the National Board of Medical Examination's (NBME) performance. A meta-analysis of six studies with 22 samples/classes showed that PBL students performed better on the clinical science exam (NBME II) but poorer on the basic science exam (NBME I); however, the differences were not statistically significant. The authors identified factors associated with the included studies that were predictive of positive outcomes with PBL: randomized designs, amount of the school's experience with PBL, and earlier publication date. This review was limited by the small number of studies and outcomes assessed, as well as having major methodological flaws. Despite these limitations, the article does draw attention to the need to consider differences across PBL curricula in future evaluations and in developing conclusions and recommendations.

In 2000, Colliver (2000) conducted a review of eight studies, including three randomized trials, published between 1992 and 1998. The review methods were poorly described and the review itself had major methodological flaws. The three randomized trials showed no beneficial effect of PBL on NBME performance, diagnostic reasoning, or clinical problem solving. One trial showed some positive effect of PBL on interpersonal skills, although Colliver argued that the effect was at best moderate and the results were highly confounded. He also noted that any beneficial effects observed in the non-randomized studies were likely due to selection bias and 'use of outcomes that directly reflect the activities and experiences of the curriculum tracks' (Colliver 2000). Colliver (2000) concluded that there is 'no convincing evidence that PBL improves knowledge base and clinical performance, at least not of the magnitude that would be expected given the resources required for a PBL curriculum'.

Newman (2003) conducted a pilot systematic review and meta-analysis of the effectiveness of PBL based on a select sample of 15 studies identified from previous reviews. PBL resulted in more positive attitudes to clinical practice while non-PBL resulted in better consultation skills. For approaches to learning, results from two studies favoured PBL. One study showed that 'satisfaction with the learning environment' favoured PBL. There was no significant difference in knowledge acquisition and several factors were identified that influenced the outcomes, including 'study design, randomization, level of education, and assessment format.' Newman (2003) concluded that the 'limited high quality evidence available from existing reviews does not provide robust evidence about the effectiveness of different kinds of PBL in different contexts'. Aside from the limited search strategy, this pilot study followed accepted methods for systematic reviews.

The same year, Dochy et al. (2003) published a meta-analysis of 43 studies to evaluate PBL in terms of knowledge and skills, as well as to identify factors that may modify the effects of PBL. This review was not restricted to medical education but included evaluations of PBL in all forms of tertiary education. The analysis showed moderate significant effects on practice skills favouring PBL. There was a trend for scores on knowledge tests to be lower in the non-PBL group, although the authors deemed the effect to be small and not of practical significance. While the appropriateness of combining these data in a meta-analysis is questionable due to substantial heterogeneity across studies, the analysis did provide some insight into potential effect modifiers that warrant attention in future research. These exploratory analyses, which were based on small numbers of studies, suggested that study design, scope of PBL (i.e., course- vs. curriculum-based), students' level of expertise, retention period, and assessment methods may explain variability in effect estimates. The authors cite as their main limitation the compromised internal validity of the primary research studies.

Most recently, Koh et al. (2008) conducted a systematic review that evaluated PBL on 37 outcomes of physician competency (identified by the authors) post-graduation. This review was one of the most methodologically rigorous to date in that it comprised a comprehensive and/or systematic approach to searching, study selection, data extraction, and quality assessment. The authors identified 13 unique relevant studies although four only provided self-reported data which the authors acknowledge as being prone to inaccuracy. The analysis yielded significant results supporting PBL for 7 of the 37 competencies. The competencies for which there were strong levels of evidence fell within the social and cognitive dimensions. Those with little or weak evidence fell in the technical, research, teaching, managerial, and knowledge dimensions. Koh et al. pointed out a number of limitations of their review some of which stem from the nature of the literature, in particular, the challenge of disentangling the effects of PBL from other curricular changes.

Notwithstanding these conflicting reviews and some scepticism around the effectiveness of PBL, there is continued support for and movement towards curriculum reform and integration of PBL and other active learning strategies in undergraduate medical education. The objective of this systematic review was to contribute to the evidence base through a methodologically rigorous synthesis of the literature that evaluated PBL in undergraduate medical education prior to clinical rotations (i.e., pre-clinical). It is at this stage of medical education that the application of PBL may be the most contentious (Shanley 2007).

Methods

Research question

In undergraduate pre-clinical medical education, what is the evidence from comparative studies of the effectiveness of PBL versus passive learning strategies in terms of evaluative outcomes (Vernon & Blake 1993)?

Search strategy

A research librarian developed comprehensive search strategies in consultation with experts in medical education to identify relevant studies in MEDLINE, PSYCINFO, and ERIC from 1985 to December 2007. The complete search strategy for MEDLINE appears in Appendix A (available at www.medicalteacher.org); the search strategies for the other databases are available from the authors on request. To identify additional relevant studies, we checked the references lists of related reviews and all included studies. Further, we hand-searched conference proceedings from the Association of Medical Education of Europe (AMEE) and the American Association for Medical Colleges (AAMC) for 2007 and 2008.

Screening and selection of studies

Two reviewers (LH, KR, DD, LT, or CS) independently screened each title and abstract generated from the searches. The full manuscripts for all titles deemed potentially relevant were retrieved and examined in detail by two reviewers (LH, CS, DD, or LB) to determine relevance to the review. Disagreements or uncertainty with respect to inclusion were resolved through discussion or in consultation with a third party with expertise in medical education (AO or JE).

Primary research studies were included if they compared some form of PBL (either individual courses or partial/entire curricula) to passive learning strategies during pre-clerkship undergraduate medical education; and, reported primary data that measured effectiveness. Acceptable outcomes were 'evaluative measures' as defined by Vernon and Blake (1993). These include measures with unambiguous good or bad (positive or negative) dimensions (e.g., test scores); modes of reasoning, processes of learning or acquiring knowledge, and attitudes or opinions were not considered (Vernon & Blake 1993). For practical reasons studies were restricted to those conducted in Europe, North America, or Australia.

Assessment of methodological quality

Two reviewers independently assessed the methodological quality of included studies using the following instruments: Cochrane Risk of Bias tool (randomized and non-randomized trials); Newcastle-Ottawa Quality Assessment Scale (cohort and cross-sectional studies); and Cochrane Effective Practice and Organization of Care criteria (interrupted-time series). Discrepancies were resolved through consensus.

Data extraction

One reviewer extracted data (LH) and a second reviewer (CS) checked data for accuracy and completeness. Data were extracted using an electronic form that captured: study design and primary objective; location and setting of the study; population; description of the intervention and comparison; results; authors' conclusions; and stated limitations.

Analysis

We did not perform a quantitative analysis because of substantial heterogeneity across studies in the interventions and comparisons, study populations, study designs, and outcomes assessed. A qualitative analysis of the studies was performed by two reviewers (LH, CS). For the analysis, the studies were grouped by design (trials vs. other) based on hierarchies of evidence that place trials among the more robust study designs with less susceptibility for bias. The results reported within each study were then summarized by outcome. Evidence tables were developed that detail the quantitative findings, author conclusions, and stated limitations for each study. The results are presented according to curriculum-wide versus course-based PBL based on evidence that the scope of PBL may create variability in results (Dochy et al. 2003).

Results

Overview

Figure 1 presents a flow diagram of the selection process. From over 6000 citations, 38 were identified as relevant to the

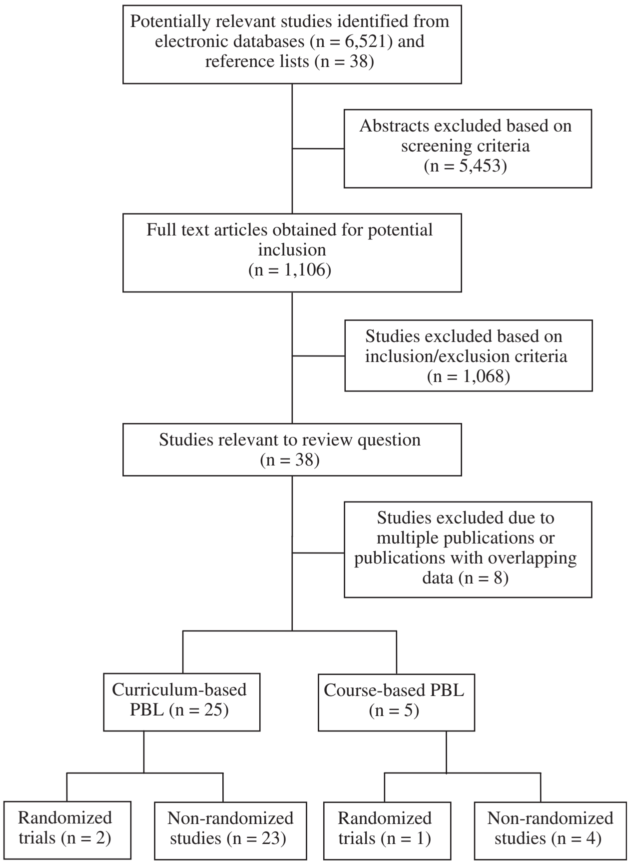

Figure 1. Flow of studies through the selection process.

review. Five of these represented multiple publications (Sachs et al. 1985; Woodward et al. 1988; Moore et al. 1990; Block & Moore 1994; Schmidt et al. 1995), two publications were subsequently updated with data from additional years (Baca et al. 1990; Distlehorst & Robbs 1998), and one study included duplicate data for outcomes of interest (Richards & Cariaga 1993). We identified four potentially relevant abstracts from hand-searching conference proceedings; however, none has been published in full and the abstracts contain insufficient data either to assess relevance or to include in the review. Therefore, a total of 30 unique studies were included: 25 curriculum- and 5 course-based comparisons. The studies and their results are detailed in Evidence Tables 1–4, while a summary of findings is presented in Table 2 (available at www.medicalteacher.org). A list of excluded studies and reasons for exclusion are available from the authors; the majority of studies were excluded because they were not comparative (72%).

Curriculum-wide PBL

Of the 25 curriculum-based PBL evaluations there were: two randomized trials; 11 cohort studies; one interrupted time series; and, 11 cross-sectional studies. The methodological quality of the studies is presented by components in Appendix B (available at www.medicalteacher.org). The randomized trials were at high risk of bias primarily due to lack of blinding. The interrupted time series was generally well conducted but did not report a formal test for trend over time. General limitations with the cohort studies included self-selection of the PBL cohort ($N=7$), no control for confounding in the design or analysis ($N=6$), and inadequate reporting of losses to follow-up ($N=8$). General limitations with the cross-sectional studies were self-selection of the PBL cohort ($N=6$), selection of the PBL and non-PBL cohorts from different sources ($N=10$), inadequate reporting of how outcomes were assessed ($N=6$), and inadequate or unclear response rates ($N=7$).

The majority of studies were conducted in North America (11 US, 5 Canada, 5 Netherlands, 1 Netherlands/Belgium, 1 Netherlands/Italy/Germany, 1 Norway, 1 Australia). The studies were published between 1986 and 2006, with 1996 being the median year of publication. The curriculum-wide PBL programs varied across studies: while the focus of the PBL programs involved small group learning with problem-based discussions, the programs differed with respect to the level of student independence, extent of supplementary lectures, use of other active learning strategies (e.g., simulated patients), and timing and extent of community-based exposure (Evidence Table 1, available at www.medicalteacher.org).

Most studies evaluated the short-term performance of undergraduate medical students by comparing scores on standardized national examinations. Three studies assessed performance during clerkships (Richards et al. 1996; Whitfield et al. 2002; Distlehorst et al. 2005), two during internship (Woodward 1990; Rolfe et al. 1995), two during residency (Santos-Gomez et al. 1990; Hoffman et al. 2006), and three of post-graduation clinical practice (Woodward et al. 1990; Shin et al. 1993; Tamblyn et al. 2005). Four studies evaluated diagnostic accuracy (Boshuizen et al. 1993; Schmidt et al. 1996; Hmelo 1998; Patel et al. 2001).

The two randomized trials evaluated standardized exam performance and clinical reasoning skills: one trial found no significant differences between PBL and non-PBL students on these items (Moore et al. 1994), while the second trial found a significant difference for the NBME Part I examination favouring the non-PBL group but no differences for the NBME Parts II or III (Mennin et al. 1993). Mennin et al. (1993) concluded that a hybrid program that provides more structured learning early in the curriculum may be most appropriate. Moore et al. (1994) concluded that PBL fosters self-directed learning and may develop humanism in the pre-clinical years.

Among the 23 non-randomized studies, the PBL intervention, the comparator group, and presentation of data differed on so many characteristics that it was not possible to combine the results; however, 13 examined knowledge acquisition. Of these, 11 found no significant difference in exam scores (Farquhar et al. 1986; Van Hessen & Verwijnen 1990; Verwijnen et al. 1990; Albano et al. 1996; Richards et al. 1996; Verhoeven et al. 1998; Way et al. 1999; Enarson & Cariaga-Lo 2001; Whitfield et al. 2002; Distlehorst et al. 2005; Lycke et al. 2006), while two found a significant benefit for PBL students (Remmen et al. 2001; Hoffman et al. 2006). Three studies showed improved clerkship performance (Richards et al. 1996; Whitfield et al. 2002; Distlehorst et al. 2005); effect sizes were small (Whitfield et al. 2002; Distlehorst et al. 2005) to moderate (Richards et al. 1996; Whitfield et al. 2002). One study of first-year interns showed benefits for PBL on 4 of 13 pre-defined competencies, while graduates from traditional schools were rated better on the other competencies. The competencies at which the PBL graduates excelled were a focus of the PBL curriculum (i.e., self-directed learning, interpersonal relationships) (Rolfe et al. 1995). While Woodward (1990) found favourable trends for interns from a PBL school with respect to supervisor ratings, she commented on the 'lack of distinctiveness' between PBL and non-PBL interns in their profile of competencies. Hoffman et al. (2006) found gains in residency performance, such as improved communication and collaboration skills, maturity, and initiative. In contrast, Santos-Gomez et al. (1990) found no significant differences overall in residency performance as rated by either nurses or doctor-supervisors.

Three studies examined the influence of PBL versus traditional curricula on the knowledge and practices of physicians (Woodward et al. 1990; Shin et al. 1993; Tamblyn et al. 2005). Shin et al. (1993) examined knowledge with respect to management of hypertension among primary care physicians who had followed PBL or traditional curricula. While they found a significant difference in test scores favouring those from a PBL curriculum (68% vs. 62%; $p < 0.01$), the authors questioned whether the small difference they detected was clinically significant. Tamblyn et al. (2005) studied practicing family physicians to determine differences in rates of mammographic screening, continuity of care, and prescribing patterns. They found that physicians from a community-oriented PBL curriculum showed significant practice improvements; however, the changes were not significantly greater than simultaneous changes in the three comparison schools that followed traditional curricula (Tamblyn et al. 2005). Woodward et al. (1990) compared patterns of billing to a provincial health plan for PBL graduates versus graduates from the province's non-PBL medical schools. In general, they observed that PBL graduates 'provided fewer services, saw fewer patients, and earned less'; however, their 'cost per patient seen was higher'. The authors attribute these findings to different types of services provided by PBL graduates, notably fewer minor assessments and more psychotherapy services. The authors were unable to assess the effect of other confounding factors (e.g., personal values and characteristics, selection criteria, type of curriculum, post-graduate education) because of the study design and data source. Both Shin and Woodward point out that there is no information regarding the impact of the observed differences on the quality or outcome of patient care.

Three of the four studies evaluating diagnostic accuracy reported some benefits for PBL students. Hmelo et al. (1998) found greater increase in diagnostic accuracy among first year PBL students ($p < 0.05$). Schmidt et al. (1996) found that students from PBL and integrated curricula made more accurate diagnoses compared to a conventional curriculum; there were no differences in the final 2 years between the PBL and integrated curricula. Based on responses to one specific clinical problem, Boshuizen found that PBL students demonstrated significantly higher quality and accuracy of answers; however, no data were reported (Boshuizen et al. 1993). In contrast, Patel et al. (2001) found that family medicine and internal medicine residents from a conventional curriculum showed significantly better diagnostic accuracy over PBL interns in two different case studies (Family Medicine: case 1: 83% vs. 60%; case 2: 33% vs. 5%; Internal Medicine: case 1: 92% vs. 82%; case 2: 42% vs. 21% [all significant at $p < 0.05$]).

Course-based PBL

Five studies evaluated the effectiveness of PBL within the context of a specific course or topic that varied by study (Evidence Tables 3 and 4, available at www.medicalteacheer.org): two trials (1 randomized and 1 non-randomized) and three cohort studies. The trials were at high risk of bias due to non-randomization in one and lack of blinding in both. General methodological limitations with the cohort studies were inadequate or lack of reporting regarding the representativeness of the PBL cohort ($N = 2$), lack of control for confounding in the analysis ($N = 2$), and inadequate reporting of losses to follow-up ($N = 3$).

The studies were conducted in the US ($N = 2$), France ($N = 1$), Australia ($N = 1$), and England ($N = 1$) and were published between 1990 and 2005 (median year 1999). The most common evaluative outcome was knowledge acquisition measured by exam scores.

Only one study was a randomized trial: the study found no significant difference in knowledge in terms of overall academic assessment, although the authors concluded that PBL students had a 'richer learning experience' (Dyke et al. 2001). In a non-randomized trial, Eisenstaedt et al. (1990) found a significant difference in short-term recall favouring the traditional lecture-based format (80% [SE 1.7] vs. 67% [SE 2.7]); however, long-term retention of information was no different between groups on follow-up assessment 2 years after the course. The three observational studies showed inconsistent findings. Hinduja et al. (2005) found significantly better exam scores for traditional students (37/50 [SD 3.9] vs. 32.35/50 [SD 4.9], $p < 0.001$). Casassus et al. (1999) found no difference in knowledge acquisition between PBL and traditional students; however, PBL students showed better problem-solving skills. Finally, Sivam et al. (1995) found that students from the traditional curriculum were at or below the national average on a standardized exam while PBL students performed at or better than the national average.

Discussion

The existing literature provides inconsistent findings with respect to the effectiveness of PBL relative to more traditional methods in undergraduate medical education. Discrepant results likely stem from the heterogeneity of interventions, varied study designs, relatively short follow-up periods (i.e., during or at the end of the course/curriculum), and the difficulty in objectively evaluating some important professional characteristics such as team work and life-long learning. The two randomized trials of PBL produced conflicting results. Authors of both studies commented on the relative merits of PBL: PBL does not appear to hamper knowledge acquisition and may develop characteristics that result in 'better' clinicians post-graduation (Mennin et al. 1993; Moore et al. 1994). One study suggested that a hybrid program that provides more structured learning early in the curriculum with a gradual increase in active and self-directed learning may be optimal (Mennin et al. 1993). Few studies assessed the impact of PBL curricula during medical school among practicing physicians. One study showed no significant differences compared to three comparison schools in quality of primary care (Tamblyn et al. 2005), while another suggested a small benefit in terms of knowledge acquisition post-graduation (Shin et al. 1993). A further study identified differences in billing patterns for PBL graduates but could not associate these differences with quality or outcomes of patient care (Woodward et al. 1990).

While this review was conducted according to methodologically rigorous guidelines (Higgins & Green 2006), it is limited by the weaknesses of the included studies. Our detailed assessment of the methodological quality of the included studies highlights some of these weaknesses and areas for improvement in future research. There were few randomized trials which represent the highest level of evidence for the effectiveness of an intervention. Those trials included were at high risk of bias primarily due to the lack of blinding which can lead to overestimates of an intervention's effects. Blinding can be challenging in these studies due the nature of the intervention; blinding of outcomes assessors and objectively measured outcomes should be considered. The Newcastle–Ottawa Quality Assessment Scale provides a framework for consideration when designing a study in order to limit biases due to sample selection, confounding, outcome assessment, and inadequate follow-up.

Numerous limitations of this literature have been identified and discussed previously (Schmidt et al. 1987; Berkson 1993; Vernon & Blake 1993; Leung & Johnston 2006). Key limitations previously discussed include: bias introduced by self-selection which may result in cohorts with abilities and preferences that are commensurate with the curricular track they choose (Eisenstaedt et al. 1990; Mennin et al. 1993; Vernon & Blake 1993; Antepohl & Herzig 1999); difficulty controlling for many extraneous variables that may affect outcomes, particularly in studies that extend over a period of time (Schmidt et al. 1987); difficulty identifying and isolating the relative contributions of different curricular components that may affect outcomes (Schmidt et al. 1987; Schmidt et al. 1996; Tamblyn et al. 2005); and the limitation of existing outcomes and measurement tools to directly capture important areas of physician competence (Berkson 1993; Vernon & Blake 1993; Distlehorst et al. 2005). The heterogeneity across studies in study populations, research designs, PBL interventions, control groups, and outcome reporting limits generalizability. Publication bias has been raised as an issue in this body of literature (Vernon & Blake 1993) and is difficult to assess outside of a meta-analysis. Vernon and Blake (1993) suggested that researchers may be reluctant to publish negative results on this topic.

A recently published study supports the need for improved quality of reporting in the field of medical education research (Cook et al. 2007). Specifically, future reports must provide adequate and precise descriptions of the interventions and controls in order to permit comparisons and interpretation (Kalaian et al. 1999; Newman 2003; Cook et al. 2007). Measures need to be developed and carefully chosen to reflect the outcomes that different learning strategies seek to influence (Vernon & Blake 1993; Kalaian et al. 1999). Our review focused on evaluative outcomes, therefore the conclusions may be inconsistent with studies that report attitudes or opinions. For instance, previous studies have demonstrated that PBL is more enjoyable for students, and this may be the basis for claims of effectiveness and impetus for its implementation. Rigorous research methods need to be employed and where necessary, there should be careful consideration of potential-effect modifiers such as student and program characteristics, and the extent of a school's experience with the experimental intervention (Kalaian et al. 1999). We found that many of the non-randomized studies did not control for confounding in the design or analysis, or the analysis controlled for a limited number of potential confounders. The value of randomized designs is that all known or unknown variables are balanced across groups, hence the findings are attributable to the intervention under study. Finally, research is needed to evaluate the effect of PBL on the quality of care and resource utilization of post-graduate clinicians (Albanese & Mitchell 1993).

In summary, 22 years of evidence does not provide unequivocal support for enhanced learning through PBL. What many conclude is that there are no deficits in knowledge; however, no investigations were designed to assess equivalence (i.e., designed and powered to determine whether the difference in effect between groups lies within an upper and lower equivalence level of clinically/practically acceptable differences) (Higgins & Green 2006). The advantages of PBL are difficult to quantify and there are questions around the most appropriate outcome measures for evaluating learning strategies. The majority of studies examined short-term knowledge acquisition measured by standardized exams. The extent to which this outcome correlates with what makes a good physician is unclear. The studies in this area are generally methodologically weak. The threat to internal validity (i.e., potential risk of bias) puts into question any positive findings, particularly where effect sizes are small. Furthermore, the heterogeneity of interventions limits generalizability.

Acknowledgments

We would like to thank the following individuals for their assistance: Kelly Russell (screening articles); Dr. Donna Dryden (screening articles and methodological advice); Liza Bialy (inclusion); Marie-Claire St-Jacques (article retrieval and administrative tasks); Robin Leicht (article retrieval); Dr. Janet Ellsworth (content expertise); Kenneth Bond (quality assessment); Michele Hamm (quality assessment); and, Krystal Harvey (quality assessment). Funding for this work was provided by the Faculty of Medicine and Dentistry at the University of Alberta, Edmonton, Alberta, Canada.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

LISA HARTLING coordinated the project, screened articles and applied inclusion criteria, extracted data, and drafted the manuscript.

CAROL SPOONER screened articles and applied inclusion criteria, checked data extraction, and reviewed and edited the manuscript.

LISA TJOSVOLD conducted the literature searches, screened articles for inclusion, and assisted with manuscript preparation.

ANNA OSWALD assisted with inclusion and manuscript preparation, and provided content expertise.

References

Albanese MA, Mitchell S. 1993. Problem-based learning: A review of literature on its outcomes and implementation issues. Acad Med 68(1):52–81.

Albano MG, Cavallo F, Hoogenboom R, Magni F, Majoor G, Manenti F, Schuwirth L, Stiegler I, van der Vleuten C. 1996. An international comparison of knowledge levels of medical students: The Maastricht Progress Test. Med Educ 30(4):239–245.

Antepohl W, Herzig S. 1999. Problem-based learning versus lecture-based learning in a course of basic pharmacology: A controlled, randomized study. Med Educ 33(2):106–113.

Baca E, Mennin SP, Kaufman A, Moore-West M. 1990. Comparison between a problem-based, community-oriented track and a traditional track within one medical school. In: Nooman ZM, Schmidt HG, Ezzat ES, editors. Innovation in medical education: An evaluation of its present status. New York: Springer. pp 9–26.

Barrows HS. 1996. Problem-based learning in medicine and beyond: A brief overview. New Dir Teach Learn 68:3–12.

Berkson L. 1993. Problem-based learning: Have the expectations been met? Acad Med 68(10 Suppl):S79–S88.

Block SD, Moore GT. 1994. Project evaluation. In: Tosteson DC, Adelstein SJ, Carver ST, editors. New pathways to medical education: Learning to learn at Harvard Medical School. Cambridge, MA: Harvard University Press. pp 114–122.

Boshuizen HPA, Schmidt HG, Wassmer L. 1993. Curriculum style and the integration of biomedical and clinical knowledge. In: Bouhuys PAJ, Schmidt HG, van Berket HJM, editors. Problem-based learning as an educational strategy. Maastricht: Network Publications. pp 33–41.

Casassus P, Hivon R, Gagnayre R, d'Ivernois JF. 1999. An initial experiment in haematology instruction using the problem-based learning method in third-year medical training in France. Hematol Cell Ther 41(4):137–144.

Colliver JA. 2000. Effectiveness of problem-based learning curricula: Research and theory. Acad Med 75(3):259–266.

Cook DA, Beckman TJ, Bordage G. 2007. Quality of reporting of experimental studies in medical education: A systematic review. Med Educ 41:737–745.

Distlehorst LH, Dawson E, Robbs RS, Barrows HS. 2005. Problem-based learning outcomes: The glass half-full. Acad Med 80(3):294–299.

Distlehorst LH, Robbs RS. 1998. A comparison of problem-based learning and standard curriculum students: Three years of retrospective data. Teach Learn Med 10(3):131–137.

Dochy F, Segers M, Van den Bossche P, Gijbels D. 2003. Effects of problem-based learning: A meta-analysis. Learn Instr 13:533–568.

Dyke P, Jamrozik K, Plant AJ. 2001. A randomized trial of a problem-based learning approach for teaching epidemiology. Acad Med 76(4): 373–379.

Eisenstaedt RS, Barry WE, Glanz K. 1990. Problem-based learning: Cognitive retention and cohort traits of randomly selected participants and decliners. Acad Med 65(9 Suppl):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. Med Educ 35(11):1050–1055.

Farquhar LJ, Haf J, Kotabe K. 1986. Effect of two preclinical curricula on NBME Part I examination performance. J Med Educ 61(5):368–373.

Higgins JPT, Green S. 2006. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6. Chichester, UK: John Wiley & Sons, Ltd, Cochrane Collaboration.

Hinduja K, Samuel R, Mitchell S. 2005. Problem-based learning: Is anatomy a casualty? Surgeon J Royal Coll Surg Edinburgh & Ireland 3(2):84–87.

Hmelo CE. 1998. Problem-based learning: Effects on the early acquisition of cognitive skill in medicine. J Learn Sci 7(2):173–208.

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

Kalaian HA, Mullan PB, Kasim RM. 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. Adv Health Sci Educ 4(3):209–221.

Koh GC, Khoo HE, Wong ML, Koh D. 2008. The effects of problem-based learning during medical school on physician competency: A systematic review. CMAJ Can Med Assoc J 178:34–41.

Leung GM, Johnston JM. 2006. Evidence-based medical education – quo vadis? J Eval Clin Pract 12(3):353–364.

Lycke KH, Grottum P, Stromso HI. 2006. Student learning strategies, mental models and learning outcomes in problem-based and traditional curricula in medicine. Med Teach 28(8):717–722.

Mennin SP, Friedman M, Skipper B, Kalishman S, Snyder J. 1993. Performances on the NBME I, II, and III by medical students in the problem-based learning and conventional tracks at the University of New Mexico. Acad Med 68(8):616–624.

Moore G, Block S, Mitchell R. 1990. A randomized controlled trial evaluating the impact of the new pathway curriculum at Harvard medical school. Cambridge, MA: Harvard University Medical School. Final report.

Moore GT, Block SD, Style CB, Mitchell R. 1994. The influence of the New Pathway curriculum on Harvard medical students. Acad Med 69(12):983–989.

Newman M. 2003. A pilot systematic review and meta-analysis on the effectiveness of problem-based learning. Special Report 2. Newcastle, UK: Learning & Teaching Subject Network.

Norman GR, Schmidt HG. 1992. The psychological basis of problem-based learning: A review of the evidence. Acad Med 67:557–565.

Oxman AD, Guyatt GH. 1991. Validation of an index of the quality of review articles. J Clin Epidemiol 44:1271–1278.

Patel VL, Arocha JF, Leccisi MS. 2001. Impact of undergraduate medical training on house staff problem-solving performance: Implications for problem-based curricula. J Dental Educ 65(11):1199–1218.

Remmen R, Scherpbier A, van der Vleuten C, Denekens J, Derese A, Hermann I, Hoogenboom R, Kramer A, Van Rossum H, Van Royen P, et al. 2001. Effectiveness of basic clinical skills training programmes: A cross-sectional comparison of four medical schools. Med Educ 35(2):121–128.

Richards BF, Ober KP, Cariaga-Lo L, Camp MG, Philp J, McFarlane M, Rupp R, Zaccaro DJ. 1996. Ratings of students' performances in a third-year internal medicine clerkship: A comparison between problem-based and lecture-based curricula. Acad Med 71:187–189.

Richards BF, Cariaga LD. 1993. A comparison between students in problem-based and traditional curricula at the same medical school: Preparing for the NBME Part I. Med Educ 27(2):130–136.

Rolfe IE, Andren JM, Pearson S, Hensley MJ, Gordon JJ. 1995. Clinical competence of interns. Programme Evaluation Committee (PEC). Med Educ 29:225–230.

Sachs LA, Altman M, Trzebiatowski GL, Williams JH, Verny R, Bellchambers M. 1985. A comparison of medical students' performances in independent study and traditional programs. J Med Educ 60(8):602–609.

Santos-Gomez L, Kalishman S, Rezler A, Skipper B, Mennin SP. 1990. Residency performance of graduates from a problem-based and a conventional curriculum. Med Educ 24(4):366–375.

Schmidt HG, Machiels-Bougaerts M, Hermans H, Tencafe O, Venekamp R, Boshvizen HPA. 1995. The development of diagnostic competence: A comparison between a problem-based, an integrated, and a conventional medical curriculum. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 18–22.

Schmidt HG, Dauphinee WD, Patel VL. 1987. Comparing the effects of problem-based and conventional curricula in an international sample. J Med Educ 62(4):305–315.

Schmidt HG, Machiels-Bongaerts M, Hermans H, ten Cate TJ, Venekamp R, Boshuizen HP. 1996. The development of diagnostic competence: Comparison of a problem-based, an integrated, and a conventional medical curriculum. Acad Med 71(6):658–664.

Shanley PF. 2007. Viewpoint: Leaving the 'empty glass' of problem-based learning behind: New assumptions and a revised model for case study in preclinical medical education. Acad Med 82(5):479–485.

Shin JH, Haynes RB, Johnston ME. 1993. Effect of problem-based, self-directed undergraduate education on life-long learning. CMAJ Can Med Assoc J 148:969–976.

Sivam SP, Latridis PG, Vaughn S. 1995. Integration of pharmacology into a problem-based learning curriculum for medical students. Med Educ 29(4):289–296.

Sivan A, Wong Leung R, Woon C. 2000. An implementation of active learning and its effect on the quality of student learning. Innov Educ Train Int 37(4):381–389.

Tamblyn R, Abrahamowicz M, Dauphinee D, Girard N, Bartlett G, Grand'Maison P, Brailovsky C. 2005. Effect of a community oriented problem based learning curriculum on quality of primary care delivered by graduates: Historical cohort comparison study. Brit Med J 331(7523):1002.

Van Hessen PAW, Verwijnen GM. 1990. Does problem-based learning provide other knowledge?. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, editors. Teaching and assessing clinical competence. Groningen, The Netherlands: Boekwerk Publications. pp 446–451.

Verhoeven BH, Verwijnen GM, Scherpbier AJJA, Holdrinet RSG, Oeseburg B, Bulte JA, van der Vleuten CPM. 1998. An analysis of progress test results of PBL and non-PBL students. Med Teach 20(4): 310–316.

Vernon DT, Blake RL. 1993. Does problem-based learning work? A meta-analysis of evaluative research. Acad Med 68(7):550–563.

Verwijnen M, Van der Vleuten C, Imbos T. 1990. A comparison of an innovative medical school with traditional schools: An analysis in the cognitive domain. In: Nooman ZH, Schmidt HG, Ezzat ES, editors. Innovation in medical education: An evaluation of its present status. New York: Springer. pp 41–49.

Way DP, Biagi B, Clausen K, Hudson A. 1999. The effects of basic science pathway on USMLE Step 1 scores. Acad Med 74 (10 Suppl):S7–S9.

Whitfield CF, Mauger EA, Zwicker J, Lehman EB. 2002. Differences between students in problem-based and lecture-based curricula measured by clerkship performance ratings at the beginning of the third year. Teach Learn Med 14(4):211–217.

Woodward C. 1990. Monitoring an innovation in medical education: The McMaster experience. In: Nooman ZM, Schmidt HG, Ezzat ES, editors. Innovation in medical education: An evaluation of its present status. New York: Springer. pp 27–39.

Woodward CA, Ferrier BM, Goldsmith C, Cohen M. 1988. Billing patterns of general practitioners and family physicians in Ontario: A comparison of graduates of McMaster Medical School with graduates of other Ontario medical schools. Res Med Educ 27:276–281.

Woodward CA, Ferrier BM, Cohen M, Goldsmith A. 1990. A comparison of the practice patterns of general practitioners and family physicians graduating from McMaster and other Ontario medical schools. Teach Learn Med 2(2):79–88.