Effects of problem-based learning: A meta-analysis.

Pergamon

Learning and Instruction 13 (2003) 533–568

Learning and Instruction

www.elsevier.com/locate/learninstruc

Effects of problem-based learning: a meta-analysis

Filip Dochy [a,b,*], Mien Segers [b], Piet Van den Bossche [b], David Gijbels [b]

[a] *University of Leuven, Afdeling Didactiek, Vesaliusstraat 2, 3000 Leuven, Belgium*
[b] *University of Maastricht, The Netherlands*

Abstract

This meta-analysis has two aims: (a) to address the main effects of problem based learning on two categories of outcomes: knowledge and skills; and (b) to address potential moderators of the effect of problem based learning. We selected 43 articles that met the criteria for inclusion: empirical studies on problem based learning in tertiary education conducted in real-life classrooms. The review reveals that there is a robust positive effect from PBL on the skills of students. This is shown by the vote count, as well as by the combined effect size. Also no single study reported negative effects. A tendency to negative results is discerned when considering the effect of PBL on the knowledge of students. The combined effect size is significantly negative. However, this result is strongly influenced by two studies and the vote count does not reach a significant level. It is concluded that the combined effect size for the effect on knowledge is non-robust. As possible moderators of PBL effects, methodological factors, expertise-level of students, retention period and type of assessment method were investigated. This moderator analysis shows that both for knowledge- and skills-related outcomes the expertise-level of the student is associated with the variation in effect sizes. Nevertheless, the results for skills give a consistent positive picture. For knowledge-related outcomes the results suggest that the differences encountered in the first and the second year disappear later on. A last remarkable finding related to the retention period is that students in PBL gained slightly less knowledge, but remember more of the acquired knowledge.

Keywords: Learning environments; Problem-based learning; Constructivism; Meta-analysis

* Corresponding author. Tel.: 32 16 325914.
E-mail address: filip.dochy@ped.kuleuven.ac.be (F. Dochy).

0959-4752/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0959-4752(02)00025-7

1. Introduction

The complexity of today's society is characterized by an infinite, dynamic and changing mass of information, the massive use of the internet, multimedia and educational technology, a rapid changing labor market demanding a more flexible labor force that is directed towards a growing proportion of knowledge-intensive work in teams and lifelong learning (Nonaka & Takeuchi, 1995; Quinn, 1992; Tynjälä, 1999). As a consequence, today's information community expects graduates not only to have a specific knowledge base but also to be able to apply this knowledge to solve complex problems in an efficient way (Engel, 1997; Poikela & Poikela, 1997; Segers, 1996). Educational research has shown that successful problem solvers possess an organized and flexible knowledge base and master the skills to apply this knowledge for problem solving (Chi, Glaser, & Rees, 1982).

Educational practices have been criticized for not developing these prerequisites of professional expertise (Mandl, Gruber, & Renkl, 1996). An important challenge for today's higher education is the development and implementation of instructional practices that will foster in students the skill to apply knowledge efficiently. For this purpose references are made to the design of "powerful learning environments" (De Corte, 1990a, 1990b; Honebein, Duffy & Fishman, 1993; Tynjälä, 1999). Such powerful learning environments should support the constructive cumulative, goal-oriented acquisition processes in all students, they should allow for the flexible adaptation of the instructional support, especially the balance between self-discovery and direct instruction (De Corte, 1995). Further, such environments should use as much as possible representative authentic, real life contexts that have personal meaning for the learners, and offer opportunities for distributed and co-operative learning through social interaction. Finally, powerful learning environments should provide possibilities to acquire general learning and thinking skills (including heuristic methods, metacognitive knowledge and strategies (Boekaerts, 1999a, 1999b)) embedded in different subject-matter (De Corte, 1995) and assessment should be congruent with the learning.

Based on recent insights in cognitive psychology and instructional science (Poikela & Poikela, 1997), many educational innovations are implemented in the hope of achieving the aforementioned goals more effectively (Segers, 1996)—educational achievements that might become regular issues in the future for decades. Already within several international evaluation projects, such as TIMSS or the 2003 OECD PISA international survey, it is seen that complex problem solving will be directly assessed (Salganik, Rychen, Moser, & Konstant, 1999). Also within the DeSeCo project of the OECD, different types of competencies are developed (that might e.g. require new educational learning environments) (Owen, Stephens, Moskowitz, & Guillermo, 2000). One of these innovations is problem-based learning (PBL) (Barrows, 1984). If one ponders the implementation of PBL, a major question is: do students from PBL reach the goals (knowledge and skills, i.e., knowledge application) in a more effective way than students who receive conventional instruction?

Albanese and Mitchell (1993, p.56) pose this question as follows:

"*Stated bluntly, if problem-based learning is simply another route to achieving the same product, why bother with the expense and effort of undertaking a painful curriculum revision?*"

In order to find an answer to this question, a meta-analysis was conducted.

2. Problem-based learning versus conventional lecture-based instruction

Although new in some aspects, problem-based learning (PBL) is generally based on ideas that originated earlier and have been nurtured by different researchers (Ausubel, Novak, & Hanesian, 1978; Bruner, 1959, 1961; Dewey, 1910, 1944; Piaget, 1954; Rogers, 1969). PBL, as it is known today, originated in the 1950s and 1960s. It grew from dissatisfaction with the common medical education practices in Canada (Barrows, 1996; Neufield & Barrows, 1974). Nowadays PBL is developed and implemented in a wide range of domains. In spite of the many variations of PBL that have evolved, a basic definition is needed to which other educational methods can be compared. Six core characteristics of PBL are distinguished in the core model described by Barrows (1996). The first characteristic is that learning needs to be student-centered. Second, learning has to occur in small student groups under the guidance of a tutor. The third characteristic refers to the tutor as a facilitator or guide. Fourth, authentic problems are primarily encountered in the learning sequence, before any preparation or study has occurred. Fifth, the problems encountered are used as a tool to achieve the required knowledge and the problem-solving skills necessary to eventually solve the problem. Finally, new information needs to be acquired through self-directed learning. It is generally recognized that a seventh characteristic should be added: Essential for PBL is that students learn by analysing and solving representative problems. Consequently, a valid assessment system evaluates students' competencies with an instrument based on real life, i.e. authentic problems (Baxter & Shavelson, 1994; Birenbaum, 1996; Shavelson, Gao, & Baxter, 1996). The assessment of the application of knowledge when solving problems is the heart of the matter. Therefore, test items require examinees to apply their knowledge to commonly occurring and important problem-solving situations (Segers, Dochy, & De Corte, 1999).

It should be noted that just as the definition of PBL is ambiguous, the definition of what constitutes a conventional lecture-based program is also ambiguous. For the most part, conventional instruction is marked by large group lectures and instructor-provided learning objectives and assignments (Albanese & Mitchell, 1993).

3. Research questions

Two sets of research questions guided this meta-analysis. First, we addressed the main effects of PBL on two broad categories of outcomes: knowledge and skills (i.e., application of knowledge). Secondly, potential moderators of the effect of PBL are addressed. A first category of moderators are design aspects of the reviewed research. In the second category of moderators, we examined whether the effect of PBL differs according to various levels of student expertise. Third, we looked more closely at different types of assessment methods. Fourth, we investigated the influence of the insertion of a retention period.

4. Method

4.1. Criteria for inclusion

Before searching the literature for work pertaining to the effects of PBL, we determined the criteria for inclusion in our analysis.

1. The work had to be empirical. Although non empirical literature and literature reviews were selected as sources of relevant research, this literature was not included in the analysis.
2. The characteristics of the learning environment had to fit the previously described core model of PBL.
3. The dependent variables used in the study had to be an operationalization of the knowledge and/or skills (i.e., knowledge application) of the students.
4. The subjects of study had to be students in tertiary education.
5. To maximize ecological validity, the study had to be conducted in a real-life classroom or programmatic setting rather than under more controlled laboratory conditions.

4.2. Literature search

The review and integration of research literature begins with the identification of the literature. Locating studies is the stage at which the most serious form of bias enters a meta-analysis (Glass, McGaw, & Smith, 1981): *"How one searches determines what one finds; and what one finds is the basis of the conclusions of one's integration"* (Glass, 1976, p. 6).

The best protection against this source of bias is a thorough description of the procedure used to locate the studies.

A first literature search was started in 1997. A wide variety of computerized databases were screened: the Educational Resources Information Center (ERIC) catalogue, PsycLIT, ADION, LIBIS. Also, the Current Contents (for Social Sciences) was searched. The following keywords were used: *problem-solving*, *learning*, *problem-based learning*, *higher education*, *college(s)*, *high school*, *research*, and *review*. The literature was selected based on reading the abstracts. This reading resulted in the selection of 14 publications that met the above criteria. Next, we employed the "snowball method" and reviewed the references in the selected articles for additional works. Review articles and theoretical overviews were also gathered to check their references. This method yielded 17 new studies. A second literature search, started in 1999, followed the same procedure. In addition we contacted several researchers active in the field of PBL and asked them to provide relevant studies or to identify additional sources of studies. This second search yielded 12 studies.

4.3. Coding study characteristics

Using other literature reviews as a guide (Albanese & Mitchell, 1993; Dochy, Segers & Buehl, 1999; Vernon & Blake, 1993), we defined the characteristics central to our review and analyzed the articles we selected on the basis of these characteristics. Specifically, the following information was recorded in tables:

1. first author and the year of publication;
2. study domain;
3. number of subjects;
4. dependent variable (i.e., method of assessment) and independent variable;
5. principal outcomes of the research; and
6. method of analysis and the statistical values.

As a result of a first analysis of the studies, it became clear that other variables could also be of importance. The coding sheet was completed with the following information:

1. the year of the study in which the assessment of the dependent variable was done;
2. if there was a retention period;
3. the name of the PBL institute.

With respect to the dependent variable, we must note that only the outcomes related to knowledge and skills (i.e., knowledge application) were coded. Some studies have examined other effects of PBL, but those were not included in the analysis.

The dependent variable was used to distinguish tests that assess knowledge from tests that assess knowledge application. The following operational definitions were used in making this distinction. A knowledge test primarily measures the knowledge of facts and the meaning of concepts and principles (Segers, 1997). This type of knowledge is often defined as declarative knowledge (Dochy & Alexander, 1995). A test that assesses skills (i.e. knowledge application) measures to what extent students can apply their knowledge (Glaser, 1990). It is important to remark that there is a continuum between knowledge and skills rather than a dichotomy. Some studies treat both aspects. In coding those studies, both aspects were separated and categorized under different headings.

Two condensed tables were created (one for knowledge and one for skills) that contain potential critical characteristics. These tables are included in Appendices A and B (legend in Appendix C). In the tables, the statistical values were, as much as possible, summarized and reported as effect size (ES) and *p*-values.

4.4. Synthesizing research

There are three methods to review literature: narrative reviews, quantitative methods, and statistical meta-analysis. In a narrative review, the author tries to make sense of the literature in a systematic and creative way (Van Ijzendoorn, 1997). Quantitative methods utilize elementary mathematical procedures for synthesizing research studies (e.g., counting frequencies into box scores). These methods are more objective but give less in-depth information than a narrative review (Dochy, Segers, & Buehl, 1999).

Glass (1976) systematized the approach of quantitative procedures and introduced the term meta-analysis: the analysis of analyses, i.e., the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings (Kulik & Kulik, 1989).

For our purposes, a statistical meta-analysis was conducted. This analysis was supplemented by more inclusive vote counts and the associated sign test.

4.4.1. Vote-counting methods

The simplest and most conservative methods for combining results of independent comparisons are the vote-counting methods. Only limited information is necessary. To do a vote count of directional results, the reviewer must count the number of comparisons that report significant results in the positive direction and compare this to the number of comparisons reporting significant results in the negative direction (Cooper, 1989). Once counted, a sign test is performed to discover if the cumulative results suggest that one direction occurs more frequently than chance would suggest (Cooper, 1989; Hunter & Schmidt, 1990). In performing this procedure, one assumes that under the null hypothesis of no relation in the population of any study, the frequency of significant positive results and negative results are expected to be equal (Hedges & Olkin, 1980).

In performing the vote count, the number of experiments with significant positive and negative findings was counted. If one study contained multiple experiments, they were all counted.

4.4.2. Statistical meta-analysis

A statistical meta-analysis is the quantitative accumulation and analysis of effect sizes and other descriptive statistics across studies (Hunter & Schmidt, 1990).

4.4.2.1. Metric for expressing effect sizes The metric that we used to estimate and describe the effects of PBL on knowledge and skills was the standardized mean difference (d-index) effect size. This metric is appropriate when the means of two groups are being compared (Cooper, 1989; Glass, McGaw, & Smith, 1981). The *d*-index expresses the distance between the two group means in terms of their common standard deviation. This common standard deviation is calculated by using the standard deviation of the control group since it is not affected by the treatment.

4.4.2.2. Identifying independent hypothesis tests One of the assumptions underlying meta-analysis is that effects are independent from one another. A problem arising from calculating average effect sizes is deciding what will be considered as an independent estimate of effect when a single study reports multiple outcomes. This meta-analysis used the shifting units method from Cooper (1989). Each statistical test is initially coded as if it were an independent event. However, when examining potential moderators of the overall relation, a study's results are only aggregated within the separate categories of the influencing variable. This strategy is a compromise that allows studies to retain their maximum information value, while keeping to a minimum any violation of the assumption of independence of hypothesis tests.

4.4.2.3. Combining effect sizes across studies Once an effect size had been calculated for each study or comparison, the effects testing the same hypothesis were averaged. Unweighted and weighted procedures were used. In the unweighted procedure, each effect size was weighted equally in calculating the average effect. In the weighted procedure, more weight is given to effect sizes with larger samples (factor *w*=inverse of the variance), based on the assumption that the larger samples more closely approximate actual effects (Cooper, 1989; Hedges & Olkin, 1985). These weighted combined effect sizes were tested for statistical significance by calculating the 95% confidence interval (Cooper, 1989).

4.4.2.4. Analyzing variance in effect sizes across studies The last step was to examine the variability of the effect sizes via a homogeneity analysis (Cooper, 1989; Hedges & Olkin, 1985; Hunter & Schmidt, 1990). This can lead to a search for potential moderators. So, we can gain insight into the factors that affect relationship strengths even though these factors may have never been studied in a single experiment (Cooper, 1989).

Homogeneity analysis compares the variance exhibited by a set of effect sizes with the variance expected by sampling error. If the result of homogeneity analysis suggests that the variance in a set of effect sizes can be attributed to sampling error alone, one can assume the data represent a population of students (Hunter, Schmidt, & Jackson, 1982).

To test whether a set of effect sizes is homogeneous, a Qt statistic (Chi-square distribution, N-1 degrees of freedom) is computed. A statistically significant Qt suggests the need for further grouping of the data. The between-groups statistic (Qb) is used to test whether the average effect of the grouping is homogeneous. A statistically significant Qb indicates that the grouping factor contributes to the variance in effect sizes, in other words, the grouping factor has a significant effect on the outcome measure analyzed (Springer, Stanne, & Donovan, 1999).

5. Results

Forty-three studies met the inclusion criteria for the meta-analysis. Of the 43 studies, 33 (76.7%) presented data on knowledge effects and 25 (58.1%) reported data on effects concerning the application of knowledge. These percentages add up to more than 100 since several studies presented outcomes of more than one category.

5.1. Main effects of PBL

The main effect of PBL on knowledge and skills is differentiated. The results of the analysis is summarized in Table 1.

In general, the results of both the vote count and the combined effect size were statistically significant. These results suggest that students in PBL are better in applying their knowledge (skills). None of the studies reported significant negative findings.

However, Table 1 would indicate that PBL has a negative effect on the knowledge base of the students, compared with the knowledge of students in a conventional learning environment. The vote count shows a negative tendency with 14 studies yielding a significant negative effect and only seven studies yielding a significant positive effect. This negative effect becomes significant for the weighted combined effect size. However, this significant negative result is mainly due to two outliers (Eisenstaedt, Bary, & Glanz, 1990; Baca, Mennin, Kaufman, & Moore-West, 1990). When these two studies are left aside, the combined effect sizes approaches zero (unweighted *ES*=-0.051; weighted *ES*=−0.107, CI:+/− 0.058).

5.1.1. Distribution of effect sizes

The results of the homogeneity analysis reported in Table 1 suggest that further grouping of the knowledge and skills data is necessary to understand the moderators of the effects of PBL. As indicated by statistically significant Qt statistics, one or more factors other than chance or sampling error account for the heterogeneous distribution of effect sizes for knowledge and skills.

Table 1
Main effects of PBL

Outcome[b]	Sign.+[c]	Sign.−[c]	Studies N[d]	Average *ES*		*Qt*
				Unweighted	Weighted (CI 95%)	
Knowledge	7	15	18	−0.776	−0.223 (+/−0.058)	1379.6 (p=0.000)
Skills	14	0[a]	17	+0.658	+0.460 (+/−0.058)	57.1 (p=0.000)

[a] Two-sided sign-test is significant at the 5% level.
[b] All weighted effect sizes are statistically significant.
[c] +/− number of studies with a significance (at the 5% level) positive/negative finding.
[d] the number of total nonindependent outcomes measured.

5.2. Moderators of PBL

5.2.1. Methodological factors

A statistical meta-analysis investigates the methodological differences between studies a posteriori (Cooper, 1989; Hunter & Schmidt, 1990). This question about methodological differences will be handled through two different aspects: the way in which the comparison between PBL and the conventional learning environment is operationalized (research design) and the scope of implementation of PBL.

5.2.1.1. Research design The studies included in the meta-analysis can all be categorized as quasi-experimental (cf., criteria for inclusion). Studies with a randomized design deliver the most trustworthy data. Studies based on a comparison between different institutes or between different tracks are less reliable because randomization is not guaranteed. Some studies attempt to compensate for this shortcoming by controlling (e.g., Antepohl & Herzig, 1997; Lewis & Tamblyn, 1987) or matching the subjects (Anthepohl & Herzig, 1997; Baca, Mennin, Kaufman & Moore-West, 1990) for substantial variables. Most problematic are those studies having a historical design (Martenson, Eriksson, & Ingelman-Sundberg, 1985). Some studies comparing the PBL-outcomes with national means were also included.

The results of the homogeneity analysis reported in Table 2 suggest no significant variation in effect sizes for knowledge-related outcomes can be attributed to method-related influences (Qb=7.261, p=0.063). However, the most reliable comparisons (random) suggest that there is almost no negative effect on knowledge acquisition.

Contrary to the data concerning knowledge, the variation in effect sizes for skills outcomes was associated with the methodological factor research design (Qb=7.177, p=0.027). The weighted combined effect sizes of the designs “between institutes” or “elective tracks” are higher than the combined effect size emanating from a historical-controlled research design.

5.2.1.2. Scope of Implementation PBL is implemented in environments varying in scope from one single course (e.g., Lewis & Tamblyn, 1987) up to an entire curriculum (e.g., Kaufman et al., 1989). While the impact of PBL as a curriculum is certainly going to be more profound, a single course can offer a more controlled environment to examine the specific effects of PBL (Albanese & Mitchell, 1993; Schmidt, 1990).

Table 3 presents the result of the analysis with scope of implementation as the moderating variable. No significantly different effects on achievement were recognized between a single course (ES=0.187) and a curriculum-wide (ES=0.311) implementation of PBL (Qb=4.213, p=0.120). In both cases a clear positive effect (see vote count and combined effect sizes) is established.

The analysis of studies examining the effect on knowledge shows that scope of implementation is associated with the variation in effect sizes (Qb=13.150, p=0.001). If PBL is implemented in a complete curriculum, there is a significant negative effect (see vote count and ES=−0.339, CI:+/− 0.099). No appreciable effects can be found in a single course design.

Table 2
Research design as moderating variable

	Sign.+	Sign. −	Studies *N*	Combined *ES*		*Qb*
				Unweighted	Weighted (CI 95%)[b]	
Knowledge						7.261 (p=0.063)
Between	0	3	2	−0.242	−0.049 (+/−0.152)[ns]	
Random	3	3	4	−1.277	−0.085 (+/−0.187)[ns]	
Historical	1	2	2	−0.680	−0.202 (+/−0.082)	
Elective	3	6	10	−0.722	−0.283 (+/−0.112)	
National	0	1				
Skills						7.177 (p=0.027)
Between	4	0	4	+0.864	+0.360 (+/−0.137)	
Elective	8	0[a]	10	+0.567	+0.317 (+/−0.103)	
Historical	2	0	3	+0.685	+0.173 (+/−0.083)	

[a] Two-sided sign-test is significant at the 5% level.
[b] Unless noted [ns] , all weighted effect sizes are statistically significant.

5.2.2. Expertise-level of students

The analysis of the moderators of PBL suggests that significant variation in effect sizes exists for knowledge (Qb=125.845, p=0.000) and skills (Qb=20.63, p=0.009). The related outcomes are associated with the expertise level of the students. The results of these analyses are summarized in Table 4. It should be noted that when conventional curricula are compared with PBL, the conventional curriculum tends to be characterized by a two-year basic science segment composed of formal courses drawn from various basic disciplines (Albanese & Mitchell, 1993; Richards et al., 1996). On the other hand, in a problem-based learning environment, the students are immediately compelled to apply their knowledge to the problems that they confront. After the first two years of the curriculum, the conventional curriculum emphasizes the application of knowledge. The conventional and the problem-based learning environment become more similar (Richards et al., 1996).

The differences of the effect sizes between the expertise levels of the students are remarkable, especially for knowledge-related outcomes. In the second year (ES=−0.315), the negative trend of the first year (ES=−0.153) becomes significant. This is also shown in the vote count. The picture changes completely in the third year. In the third year, both the vote-counting method (two significant positive effects

Table 3
Scope of Implementation as moderating variable[b]

	Sign.+	Sign. −	Studies *N*	Combined *ES*		*Qb*
				Unweighted	Weighted (CI 95%)	
Knowledge						13.150 (p=0.001)
Single course	6	4	9	−0.578	−0.113 (+/−0.071)	
Curriculum	1	10[a]	9	−0.974	−0.339 (+/−0.099)	
Skills						4.213 (p=0.120)
Single course	4	0[a]	6	+0.636	+0.187 (+/−0.081)	
Curriculum	9	0[a]	10	+0.660	+0.311 (+/−0.085)	

[a] Two-sided sign-test is significant at the 5% level.
[b] All weighted effect sizes are statistically significant.

vs zero negative) and the combined effect size (*ES*=0.390) suggest a positive effect. Students in the fourth year show a negative effect of PBL on knowledge: a negative tendency in the vote-counting method and a negative combined effect size (*ES*=−0.496). On the contrary, this negative effect is not found for students who graduated.

These results suggest that the differences arising in the first and the second year disappear if the reproduction of knowledge is assessed when the broader context asks all the students to apply their knowledge (both in the conventional and the PBL environment). The only exception is the results in the last year of the curriculum.

The effects of PBL on skills (i.e., application of knowledge), differentiated for expertise-level of students give a rather consistent picture. On all levels, there is a strong positive effect of PBL on the skills of the students.

5.2.3. *Retention period*

Table 5 summarizes the results of dividing the studies into those that have a retention period between the treatment and the test and those that do not.

If the test measures knowledge, the division leads to more homogeneous groups (*Qb*=28.683, p=0.000). The experiments with no retention period show a significant negative combined effect size (*ES*=−0.209). The vote count also supports this conclusion. On the other hand, experiments using a retention period have the tendency to find more positive effects.

These results suggest that students in PBL remember more of the acquired knowledge. A possible explanation is the attention on elaboration in PBL (Schmidt, 1990): elaboration promotes the recall of declarative knowledge (Gagné, 1978; Wittrock, 1989). Although the students in PBL would have slightly less knowledge (they do

Table 4
Expertise-level of students as moderating variable

	Sign.+	Sign. −	Studies *N*	Combined *ES*		*Qb*
				Unweighted	Weighted (CI 95%)[b]	
Knowledge						125.845 (p=0.000)
1e year	1	1	3	−0.205	−0.153 (+/−0.186)[ns]	
2e year	0	6[a]	12	−1.489	−0.315 (+/−0.067)	
3e year	2	0	5	+0.338	+0.390 (+/−0.129)	
4e year	0	1	2	−1.009	−0.138 (+/−0.199)[ns]	
5e year	0	0	1	−0.037	−0.037 (+/−0.233)[ns]	
Last year	0	4[a]	3	−0.523	−0.496 (+/−0.166)	
All	0	1	1	−0.919	−0.919 (+/−0.467)	
Graduated	2	0	4	+0.193	+0.174 (+/−0.204)[ns]	
Skills						20.630 (p=0.009)
1e year	1	0	2	+0.414	+0.433 (+/−0.340)	
2e year	1	0	4	+0.473	+0.318 (+/−0.325)[ns]	
3e year	4	0[a]	11	+0.280	+0.183 (+/−0.093)	
4e year	1	0	1	+0.238	+0.235 (+/−0.512)[ns]	
5e year	1	0	1	+0.732	+0.722 (+/−0.536)	
Last year	4	0[a]	3	+0.679	+0.444 (+/−0.174)	
All	0	0	1	+0.310	+0.310 (+/−0.161)	
Graduated	1	0	1	+1.193	+1.271 (+/−0.630)	

[a] Two-sided sign-test is significant at the 5% level.
[b] Unless noted [ns] , all weighted effect sizes are statistically significant.

Table 5
Retention period as moderating variable

	Sign.+	Sign. −	Studies *N*	Combined *ES*		*Qb*
				Unweighted	Weighted (CI 95%)[b]	
Knowledge						28.683 (p=0.000)
Retention	4	2	9	+0.003	+0.139 (+/−0.116)	
No Retention	3	13[a]	24	−0.826	−0.209 (+/−0.053)	
Skills						1.474 (p=0.223)
Retention	3	0	5	+0.511	+0.320 (+/−0.198)	
No Retention	11	0[a]	22	+0.500	+0.224 (+/−0.057)	

[a] Two-sided sign-test is significant at the 5% level.
[b] All weighted effect sizes are statistically significant.

not know as many facts), their knowledge has been elaborated more and consequently they have better recall of that knowledge.

For tests assessing skills, the results suggest that no significant variation in effect sizes can be attributed to the presence or absence of a retention period. The positive effect of PBL on the skills (knowledge application) of students seems to be immediately and lasting.

5.2.4. *Type of assessment method*

The authentic studies assessed the effects of PBL on the knowledge and skills of students in very different ways. A description of the results of contrasting the knowledge- and skills-related outcomes by the type of assessment method follows. In other contexts, research has shown that assessment methods influence the effects findings (Dochy, Segers, & Buehl, 1999).

The following assessment tools were used in the studies included in this meta-analysis:

- *National Board of Medical Examiners: United States Medical Licensing*
- *Step 1*: MCQ about basic knowledge
- *Step 2*: MCQ about diagnosing
- *Modified Essay Questions* (MEQ): The MEQ is a standardized series of open questions about a problem. The information on the case is ordered sequentially: the student receives new information only after answering a certain question (Verwijnen et al., 1982). The student must relate theoretical knowledge to the particular situation of the case (Knox, 1989).

- *Essay questions*: A question requiring an elaborated written answer is asked (Mehrens & Lehmann, 1991).
- *Short-answer questions*: Compared to an essay question, the length of the desired answer is restricted.
- *Multiple-choice questions*
- *Oral examinations*
- *Progress tests*: The progress test is a written test consisting of about 250 true-false items sampling the full domain of knowledge a graduate should master (Verwijnen, Pollemans, & Wijnen, 1995). The test is constructed to assess "rooted" knowledge, not details (Mehrens & Lehmann, 1991).
- *Performance-based testing:Rating* Standardized rating scales are used to evaluate the performance of the students (performance assessment) (Shavelson et al., 1996; Birenbaum and Dochy, 1996). It can be used to evaluate knowledge as well as higher cognitive skills (Santos-Gomez, Kalishman, Resler, Skipper, & Minnin, 1990).
- *Free recall* Students are asked to write down everything they can remember about a certain subject. This task makes a strong appeal to the students' retrieval strategies.
- *Standardized patient simulations* These tests are developed by the OMERAD Institute of the University of Michigan. A patient case is simulated and students' knowledge and clinical skills are assessed by asking the student specific questions (Jones, Bieber, Echt, Scheifley, & Ways, 1984).
- *Case(s)* Students have to answer questions about an authentic case.

If the "*Key feature approach*" is used, questions are asked on only the core aspects of the case (Bordage, 1987).

The results of the statistical meta-analysis are presented in Table 6. In this analysis, we did not use the "shifting units" method to identify the independent hypothesis tests, but the "samples as units" method (Cooper, 1989). This approach permits a single study to contribute more than one hypothesis test, if the hypothesis test is carried out on separate samples of people. In this way it was possible to gain more information on certain operationalizations of the dependent variable.

The results of the homogeneity analysis (Table 6) suggest that significant variation in effect sizes as well as effects on knowledge (Qb=254.501, p=0.000) and skills (Qb=25.039, p=0.001) can be attributed to the specific operationalization of the dependent variable.

The results in the domain of the effects on skills are more coherent than the results for knowledge. The effects found with the different operationalizations of skills are all positive. A ranking of the operationalization based on the size of the weighted combined effect sizes, gives the following:

NBME Step II (0.080); Essay (0.165); NBME III (0.263); Oral (0.366); Simulation (0.413); Case(s) (0.416); Rating (0.431); MEQ (0.476).

If this classification is compared with a continuum showing to what degree the tests assess the application of knowledge, rather than the reproduction of knowledge, the following picture emerges: the better an instrument is capable of evaluating stu-

Table 6
Type of assessment method as moderating variable

	Sign.+	Sign.−	Studies N	Combined ES		Qb
				Unweighted	Weighted (CI 95%)[b]	
Knowledge						254.501 (p=0.000)
NBME part I	0	6[a]	5	−1.740	−0.961 (+/−0.152)	
Short-answer	2	1	3	+0.050	−0.123 (+/−0.080)	
MCQ	3	7	12	−1.138	−0.309 (+/−0.109)	
Rating	1	0	4	+0.209	−0.301 (+/−0.162)	
Oral	0	0	2	−0.334	−0.350 (+/−0.552)[ns]	
Progress	0	1	6	+0.011	−0.005 (+/−0.097)[ns]	
Free recall	1	0	1	+2.171	+2.171 (+/−0.457)	
Skills						25.039 (p=0.001)
NBME part II	1	0	4	+0.094	+0.080 (+/−0.125)[ns]	
NBME part III	1	0	2	+0.265	+0.263 (+/−0.153)	
Case(s)	5	0[a]	11	+0.708	+0.416 (+/−0.119)	
MEQ	1	0	1	+0.476	+0.476 (+/−0.321)	
Simulation	1	0	2	+0.854	+0.413 (+/−0.311)	
Oral	1	0	2	+0.349	+0.366 (+/−0.554)[ns]	
Essay	2	0	3	+0.415	+0.165 (+/−0.083)	
Rating	2	0	3	+0.387	+0.431 (+/−0.182)	

[a] Two-sided sign-test is significant at the 5% level.
[b] Unless noted [ns], all weighted effect sizes are statistically significant.

dents' skills (i.e., application of knowledge), the larger the ascertained effects of PBL (compared with a conventional learning environment).

Effects found with the NBME Step II are negligible. However, the NMBE Step II is also the least suitable instrument to examine the skills of the students. It assesses clinical knowledge rather than clinical performance (Vernon & Blake, 1993). The essay questions give some opportunities to evaluate the integration of knowledge (Swanson, Case, & van der Vleuten, 1997), but they are not often used to make an application of knowledge. This is also the case for oral examination. In this case, however, there was a clear distinction between questions that examined knowledge and questions that assessed problem-solving skills (Goodman et al., 1991). Only the latter are categorized as skills-related outcomes.

Excepting the NBME Step II, essay questions, and the oral examination of NBME Step III, all the other instruments can be classified as measuring the skills of the students to apply their knowledge in an authentic situation. On these tests, the students in PBL score consistently higher (*ES* between 0.416 and 0.476). The only exception is the results on the NBME step III (*ES*=0.265). It should be noted that the exam consists only partially of authentic cases.

A rating was made for the knowledge-related outcomes:

NBME I (−0.961); Oral (−0.350); MCQ (−0.309); Rating (−0.301); Short-answer (−0.123); Progress test (−0.005); Free recall (+2.171)

The results suggest a similar conclusion as the result presented in the Retention period section. If the test makes a strong appeal to retrieval strategies, students in PBL do at least as well as the students in a conventional learning environment. A rating context, short-answer questions, or free recall tests make a stronger appeal to retrieval strategies than a recognition task (NBME step I and MCQ) (Tans, Schmidt, Schade-Hoogeveen, & Gijselaers, 1986). Also the progress test examines “rooted” knowledge. The fact that students in a conventional learning environment score better on the NBME step I and on the MCQ (see vote count), suggests that they have more knowledge. The fact that the difference between students in conventional learning environments and students in PBL diminishes or even disappears on a test appealing to retrieval strategies, suggests a better organization of the students’ knowledge in PBL. However, this conclusion is rather tentative.

6. Conclusion

6.1. Main effects

The first research question in this meta-analysis dealt with the influence of PBL on the acquisition of knowledge and the skills to apply that knowledge. The vote count as well as the combined effect size (*ES*=0.460) suggest a robust positive effect from PBL on the skills of students. Also no single study reported negative effects.

A tendency to negative results is discerned when considering the effect of PBL on the knowledge of students. The combined effect size is significantly negative (*ES*=−0.223). However, this result is strongly influenced by two studies. Also the vote count does not reach a significant level.

Evaluating the practical significance of the effects requires additional interpretation. Researchers in education and other fields continue to discuss how to evaluate the practical significance of an effect size (Springer, Stanne & Donovan, 1999).

Cohen (1988) and Kirk (1996) recommend that d=0.20 (small effect), d=0.50 (moderate effect) and d=0.80 (large effect) serve as general guidelines across disciplines. Within education, conventional measures of the practical significance of an effect size range from 0.25 (Tallmadge, 1977 in Springer, Stanne & Donovan, 1999) to 0.50 (Rossi & Wright, 1977). Many education researchers (Gall, Borg, & Gall, 1996) consider an effect size of 0.33 as the minimum to establish practical significance.

If we compare the main effects with these guidelines, it can be concluded that the combined effect size for skills is moderate, but of practical significance. The effect on knowledge, already described as non-robust, is also small and not practically significant.

6.2. Moderators of PBL effects

The moderator analysis is presented as exploratory because of the relatively small number of independent studies involved.

6.2.1. Methodological factors

The most important conclusion resulting from the analysis of the methodological factors seems to be the diminished negative effect of PBL on knowledge, if the quality of the research is categorized as higher.

6.2.2. Expertise-level of students

The analysis suggested that both for knowledge- and skills-related outcomes the expertise-level of the student is associated with the variation in effect sizes. Nevertheless, the results for skills give a consistent positive picture. For knowledge-related outcomes the differences of the effects between the expertise levels of the students are remarkable. The results suggest that the differences encountered in the first and the second year disappear if the reproduction of knowledge is assessed in a broader context that asks all the students to apply their knowledge.

6.2.3. Retention Period

This moderator analysis indicates that students in PBL have slightly less knowledge, but remember more of the acquired knowledge. A possible explanation is the attention for elaboration in PBL: the knowledge of the students in PBL is elaborated more and, consequentially, they have a better recall of their knowledge (Gagné, 1978; Wittrock, 1989). For skills-related outcomes, the analysis indicates no significant variation in effect sizes. The positive effect of PBL on the skills of students seems immediate and lasting.

6.2.4. Type of assessment method

In other contexts, research has shown that assessment methods influence the findings (Dochy, Segers, & Buehl, 1999). In this review, the effects of PBL are moderated by the way the knowledge and skills were assessed. The results seem to indicate that the more an instrument is capable of evaluating the skills of the student, the larger the ascertained effect of PBL.

Although it is not so clear, an analogue tendency is acknowledged for the knowledge-related outcomes. Students do better on a test if the test makes a stronger appeal on retrieval strategies. This could be due to a better structured knowledge base, a consequence of the attention for knowledge elaboration in PBL. This is in line with the conclusion presented previously in the Retention period.

6.3. Results of other studies: different methods, same results?

The interest in the effects of PBL has already produced two good and often cited reviews (Albenese & Mitchell, 1993; Vernon & Blake, 1993). These reviews were published in a short period and mostly rely on the same literature. The two reviews used a different methodology. Albanese and Mitchell relied on a narrative integration of the literature, while Vernon and Blake used statistical methods. Methodologically, this analysis is more similar to Vernon and Blake. Both reviews, however, concluded that at that moment there was not enough research to draw reliable conclusions.

The main results of this meta-analysis are similar to the conclusions of the two reviews. They had found a robust positive effect of PBL on skills. Vernon and Blake (1993, p. 560) express it as follows:

> "Our analysis suggests that the clinical performance and skills of students exposed to PBL are superior to those of students educated in a traditional curriculum."

The reviews also drew similar conclusions about the effects of PBL on the knowledge base of students. Albanese and Mitchell (1993, p.57) concluded very carefully:

> "While the expectation that pbl students not do as well as conventional students on basic science tests appears to be generally true, it is not always true."

Vernon and Blake (1993) specified this doubt with their statistical meta-analysis:

> "Data on the NBME I …suggest a significant trend favoring traditional teaching methods. However, the vote count showed no difference between Problem-bases learning and traditional tracks" (p.555).

And

> "Several other outcome measures that appeared to be primarily tests of basic science factual knowledge. The trend in favor of traditional teaching approaches was not statistically significant" (p.556).

This meta-analysis also made similar conclusions about the effect of PBL on knowledge and provides a further validation of the findings from the two mentioned reviews. This meta-analysis then went further by analyzing potential moderators of the main effects.

Finally, a remark should be made concerning the limitations of this review. Perhaps the greatest limitation of this meta-analysis is strongly related to its greatest strength. By including only field studies (quasi-experimental research), the meta-analysis gains a lot of ecological validity, but sacrifices some internal validity relative to more controlled laboratory studies (Springer, Stanne & Donovan, 1999). As a consequence, its results should be interpreted from this perspective, from which we try to bridge the gap between research and educational practice (De Corte, 2000).

Acknowledgements

The authors are grateful to Eduardo Cascallar at the University of New York and Neville Bennett at the University of Exeter, UK for their comments on earlier drafts.

Appendix A. Studies measuring knowledge

See Table 7

Appendix B. Studies measuring skills

See Table 8

Appendix C. Legend for the tables of Appendix A and B

Study

First author and the year of publication.

Pbl-Insitute

Institute in which the experimental condition has taken place.

University of New Mexico; Universiteit van Maastricht; McMaster University; University of NewCastle; Temple University; Michigan State University; Karolinska Institutet; University of Kentucky; Rush Medical College; Mercer University; McGill University; University of Rochester; Universiteit van Keulen; University of New Brunswick; Michener Institute; University of Alberta; Harvard Medical School; Wake Forest University; Southern Illinois University.

When no institute was mentioned, the institute was described.

Bv. ‘Institute for higher professional education’ (Tans, Schmidt, Schade-Hoogeveen, & Gijselaers, 1986).

Level

Participants’ level.

1=first year

Table 7
Studies measuring knowledge

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result *ES*	Result *p*-value
Eisenstaedt et al., 1990	Temple	2	S	R	32 / 58	N	MCQ (trad. ex.)	−8.291	$p<0.001$
		4				Y	MCQ	−2.211	$p<0.001$
Mennin et al., 1993	New Mexico	2	C	K	167/508	N	NMBE I	−7.908	$p<0.0001$
Baca et al., 1990	New Mexico	A	C	K	37 / 41	N	NBME I	−0.919	$p<0.0001$
Verwijnen et al., 1990	Maastricht	A	C	I	266/1253	N	MCQ (64 questions)	-	
					471/894		MCQ (70)	-	
					565/1234		MCQ (64)	/	
					565/167		MCQ (264)	/	
Saunders et al., 1990	Newcastle	L (5)	C	I	45/243	N	MCQ (40, conv.)	−0.716	$p<0.001$
					47/242		MCQ (40, Pbl)	−0.476	$p<0.01$
Morgan, 1977	Rochester	2	C in 2nd year	R, K		N	Subjects 2nd year		
					15 / 82		NBME I	+	
					15 / 76		NBME I	+	
					16 / 81		NBME I	+	
Farquhar et al., 1986	Michigan	2	C	K	40 / 40	N	NBME I	-	ns
							anatomy	+	ns
							Physiology	+	ns
							biochemistry	+	ns
							pathology	-	ns
							microbiology	-	$p<0.05$
							Pharmacology	-	ns
							Behavorial	+	ns

(*continued on next page*)

Table 7 (*continued*)

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Kaufman et al., 1989	New Mexico	2	C	K		N	NBME I	-	$p<0.0001$
							Class of		
							1983	-	$p<0.05$
							1984	-	$p<0.01$
							1985	–	$p<0.1$
							1986	–	ns
							1987	–	$p<0.1$
							1988	–	ns
							1989	–	$p<0.05$
							1990	–	$p<0.01$
Santos-Gomez et al., 1990	New Mexico	Af	C	K	41 / 78	Y	rating (supervisors)	+0.257	$p=0.21$
					39 / 71		rating (nurses)	−0.446	$p=0.09$
					43 / 70		rating (self)	+0.525	$p=0.05$
Martenson et al., 1985	Karolinska Institutet	Af	S	H		Y	Short answer	+	$p<0.001$
		2			1651/818	N	Short answer	−0.15	$p<0.00003$
Schwartz et al., 1997	Kentucky	3	C	K		N	MCQ test 1	-	
							MCQ test 2	/	
							MCQ test 3	-	
							MCQ final exam	/	
Goodman et al., 1991	Rush	2	C	K	72/501	N	NBME I	−0.044	$p=0.40$
							Pathology	−0.242	$p=0.01$
					12 / 12		Oral in 1985	0.0	$p=0.97$
					15 / 13		Oral in 1987	−0.667	$p=0.06$
Van Hessen and Verwijnen, 1990	Maastricht	6	C	N	/179	Y	MCQ (247 questions; progress test)	-	$p<0.05$

Table 7 (*continued*)

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Bickley et al., 1990	Mercer	2	C	N	23/	N	NBME I	-	
					24/		Pathology	+	
					20/			+	
					24/			-	
					20/			+	
Anthepol and Herzig, 1999	Köln	3	S	R, C	55/57	N	–MCQ	−0.125	p=0.4
							–Short answer questions	0.424	p=0.07
							–Total	0.167	p=0.43
Verhoeven et al., 1998	Maastricht	1	C	I	190/124	N	Progress tests (242 items)	−0.203	n.s.
		2			146/104			0.211	n.s.
		3			135/87			0.288	n.s.
		4			188/151			0.193	n.s.
		5			144/140			−0.037	n.s.
		6			135/122			−0.385	p<0.01
Lewis and Tamblyn, 1987	New Brunswick	2	S	K, C	22/20	N	MCQ (100 items)	0.024	p=0.479
Antepohl and Herzig, 1997	Köln	3	S	K, M	110/110	N	Short answer	0.603	p=0.0013
Tans et al., 1986	Institute for higher vocational education	1	S	R	77/45	N	MCQ (60 items)	−2.583	p=0.0013
						Y	Free-recall test	2.171	p<0.005
Finch, 1999	Michener Institute	3	C	H	21/26	N	MCQ (60 items)		p>0.05

Table 7 (*continued*)

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Son and Van Sickle, 2000	2 colleges economics	?	S	Non-equivalent com-parison group design	72/68	N	Instrument:	0.381	p=0.05
							–16 MCQ –8 correct/incorr –1 short answer		
						Y	idem	0.384	p=0.05
					72/80				
Aaron et al., 1998	Alberta	2	S	H	113/121	N	MCQ	−0.440	$p<0.05$
					17/12	N	MCQ	−0.769	$p>0.05$
Block and Moore, 1994	Harvard Medical School	2	C	R	62/63	N	NBME part I	=	n.s.
							–behavorial science subtest	+	sign
Richards et al., 1996	Wake Forest	3	C	K	88/364	Y	Clinical rating scale 1) amount of factual knowledge	0.5	p=0.0001
Doucet et al., 1998	Dalhousie University, Halifax	CME	S	K, C	34/29	N	MCQ (40 items)	0.434	p=0.05
Distlehorst and Robbs, 1998	Southern Illinois	2	C	K, C	47/154	N	NBME part I	0.18	p=0.6528
Imbos et al., 1984	Maastricht	A	C	I		N	Anatomy (progress test)	Van jaar 1 tot 4:= Jaar 5 en 6:+	

Table 7 (*continued*)

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Jones et al., 1984	Michigan State	2	C	K	63/138	N	NBME part I		
							Overall	+	n.s.
							–anatomy	-	n.s.
							–physiology	+	$p<0.05$
							–biochemistry	+	$p<0.05$
							–pathology	-	n.s.
							–microbiology	-	n.s.
							–pharmacology	+	n.s.
							–behavorial	+	$p<.004$
					170/331	N	Subject matter part ‘clerckships exams’ (pretest)		
							–OB/Gyn	-	n.s.
							–Pediatrics	-	$p<.01$
							–Surgery	-	n.s.
							–Medicine	-	$p<0.05$
Moore et al., 1994	Harvard Medical School	2	C	K[4]	60/61	N	NBME part I		
							–Anatomy	−0.257	$p=0.16$
							–Behavorial	0.455	$p=0.01$
							–Biochemistry	−0.138	$p=0.50$
							–Microbiology	0.323	$p=0.10$
							–Pathology	0.029	$p=0.89$
							–Pharmacology	−0.037	$p=0.71$
							–Physiology	−0.159	$p=0.43$
							Total	−0.01	$p=0.96$
		4					Free recall (preventive medicine and biochemistry)	=	

Table 7 (*continued*)

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Imbos and Verwijnen, 1982	Maastricht	A	C	I		N	Progress test	=	
Donner and Bickley, 1990	Mercer	1	C	N		N	First try pass rate NBME part I	=	
Neufeld and Sibley, 1989	McMaster	L	C	N		N	First try pass rate Examination Medical Council of Canada	–	
Albano et al., 1996	Maastricht	6 (L)	C	I		N	Progress test	=	

Table 8
Studies measuring skills

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Mennin et al., 1993	New Mexico	3	C	K	144/447	N	NBME II	+0.046	p=0.29
		L			103/313		NBME III	+0.307	$p<0.001$
Schmidt et al., 1996	Maastricht	A	C	I	tot=612	N	30 cases	+0.310 /	
Baca et al., 1996	New Mexico	3	C	I,M	36 / 36	N	NBME II	/	/
Patel et al., 1990	McMaster	1	C	I	12 / 12	N	1 case	-	/
		3			12 / 12			-	/
		L			12 / 12			-	/
Hmelo, 1998	/	1	C,S	K	39 / 37	N	6 cases	+0.768	
							-accuracy	+0.521	$p<0.5$
							-length reasoning	+0.762	$p<0.005$
							-# findings	+0.547	$p<0.05$
							-use scientific concepts	+1.241	$p<0.001$
Saunders et al., 1990	NewCastle	L	C	I	45 / 240	N	MEQ1	−0.066	ns
					44 / 243		MEQ2	+1.017	$p<0.001$
								+0.4755	
Hmelo et al., 1997	/	1,2	S	K	20 / 20	N	1 case	+0.7305	
							-length reasoning	+0.883	$p<0.01$
							-use scientific concepts	+0.578	$p<0.1$
Barrows and Tamblyn, 1976	McGill	2	S	K	10 / 10	N	Standardized Patient Simulation	+1.409	$p<0.005$

(*continued on next page*)

Table 8 (*continued*)

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Kaufman et al., 1989	New Mexico	3	C	K	Total: 120/318	N	NBME II	+0.224	$p<0.01$
							in 1983	-	ns
							1984	-	ns
							1985	+	$p<0.1$
							1986	+	$p<0.1$
							1987	+	ns
							1988	+	ns
							clinical rotations	/	ns
							in 1983	−	ns
							1984	−	ns
							1985	−	ns
							1986	+	ns
							1987	+	ns
							1988	+	ns
							1989	+	ns
							clinical subscores of clin. rot.	+	sign.
							in 1983	−	ns
							1984	+	ns
							1985	/	ns
							1986	+	ns
							1987	+	$p<0.1$
							1988	+	ns
							1989	+	ns
Schwartz et al., 1997	Kentucky	3	C	K		N	Standardized Patient simulation	+	/
							MEQ	+	/
							NBME II	/	ns
							surgery subsection	+	sign.

Table 8 (*continued*)

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Goodman et al., 1991	Rush	3	C	K	36 / 297	N	NBME II	−0.133	p=0.73
							Oral (problem solving)		
		3			12 / 12		in 1985	−0.071	p=0.89
		3			15 / 13		in 1987	+0.769	p=0.04
Schuwirth, 1996	Maastricht	1	C	I	30 / 32	N	60 short cases	+0.06	
		2			30 / 30			+0.25	
		3			30 / 30			−0.114	
		4			29 / 30			+0.238	
		5			27 / 30			+0.732	$p<0.01$
		L (6)			32 / 25			+1.254	$p<0.001$
Martenson et al., 1985	Karolinska Institutet	/	S	H	818/1651	N	Essay questions (part of trad. ex.)	+0.15	$p<0.00003$
Lewis and Tamblyn, 1987	New Brunswick	2	S	K	22/20	N	Clinical performance	+0.234	p=0.2256
Finch, 1999	Michener Institute	3 (L)	C	H	21/26	N	Essay questions	+1.904	$p<0.0005$
Aaron et al., 1998	Alberta	2	S	H	17/12	N	Essay questions	0.0	$p>0.95$
Block and Moore, 1994	Harvard Medical School	3	C	K, R	62/63	N	NBME part II	/	n.s.
							Public health	+	sign.
Boshuizen et al., 1993	Maastricht	3+4	C	I	4/4	N	1 case	+2.268	p=0.024

Table 8 (*continued*)

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Richards et al., 1996	Wake Forest	3	C	K, C	88/364	Y	- Clinical rating scale	+0.426	
							2) take history and perform physical	+0.425	p=0.002
							3) derive differential diagnosis	+0.462	p=0.0005
							4) organize and express information	+0.390	p=0.004
							–NBME medicine shelf test (~part II)	+0.073	p=0.80
Doucet et al., 1998	Continuing Medical Education	CME	S	K	21/26	Y	Key Feature Problem examination (28 cases)	+1.293	p=0.001
Distlehorst and Robbs, 1998	Southern Illinois	3	C (first 2 years)	K, C	47/154	Y	–USMLE step II	+0.390	p=0.0518
							–Rating	+0.5	p=0.0028
							–Standardized patient simulations:		
							–Overall	+0.3	p=0.0596
							–post station encounter	+0.33	p=0.0742
							–patient checklist ratings	+0.14	p=0.1669
Jones et al., 1984	Michigan State	4	C (first 2 years)	K	60/142	Y	FLEX weighted average	+	n.s.

Table 8 (*continued*)

Study	Pbl-institute	Level	Scope	Design	Subj. pbl/conv	Ret.	Operat. AV	Result	
								ES	*p*-value
Moore et al., 1994	Harvard Medical School	2	C	K, R	60/61	N	Diagnostic and clinical tasks	Geen verschil	
							In 1989		
							In 1990		
Neufeld et al., 1989	McMaster	L	C	N		N	First-try pass rate		
							–Exam Medical Council van Canada	-	
							–Canadian specialty board examinations	+	

2=second year
3=third year
4=fourth year
5=fifth year
L=last year
A=in every year
Jg=Just graduated
Scope
Scope of PBL-implementation
C=curriculum-wide
S=single course
Design

- I: comparison between two institutions
- K: -curriculum-wide: Elective track within one institution
- -single course: PBL-optional course
- R: random assignment into two groups
- H: historical control
- N: control is based on a national composite
- M: matched controls
- C: controlled for substantial variables

Subjects (subj.)
Number of subjects in the experimental condition (PBL) / number of subjects in the control condition (conv)
Retention (Ret.)
Is there a retention period between treatment and test?
Y=Yes
N=No
Operationalization dependent variable (Operat. AV)
MCQ=Multiple choice question
Ratings, questionnaires
NBME (USMLE) I, II of III=National Board of Medical Examiners Part I, Part II of Part III exam
Case
Standardized Patient (simulation)
Clinical Rotations
Oral
Essay questions
Key feature problem examination
First try pass rate
Progress testing
Result
Effect size (*ES*): The sign of the *ES* shows if the Pbl-result is greater (+) or smaller (−).

If it was not possible to compute an *ES*, than only the sign of the results is given. If there was no effect found than '/' is indicated.

p-value: ns=not significant /=no *p*-value given.

References

Aaron, S., Crocket, J., Morrish, D., Basualdo, C., Kovithavongs, T., Mielke, B., & Cook, D. (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.

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

Albano, M. G., Cavallo, F., Hoogenboom, R., Magni, F., Majoor, G., Mananti, F., Schuwirth, L., Stiegler, I., & Van Der Vleuten, C. (1996). An international comparison of knowledge levels of medical students: The Maastricht Progress Test. *Medical Education*, *30*, 239–245.

Antepohl, W., & Herzig, 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*, R18.

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

Ausubel, D., Novak, J., & Hanesian, H. (1978). *Educational psychology: A cognitive view (2nd ed.)*. New York: Holt, Rinehart & Winston.

Baca, E., Mennin, S. P., Kaufman, A., & Moore-West, M. (1990). Comparison between a problem-based, community-oriented track and a traditional track within one medical school. In Z. M. Nooman, H. G. Schmidt, & E. S. Ezzat (Eds.), *Innovation in medical education: An evaluation of its present status* (pp. 9–26). New York: Springer.

Barrows, H. S. (1984). A Specific, problem-based, self-directed learning method designed to teach medical problem-solving skills, self-learning skills and enhance knowledge retention and recall. In H. G. Schmidt, & M. L. de Volder (Eds.), *Tutorials in problem-based learning. A new direction in teaching the health profession*. Assen: Van Gorcum.

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: a brief overview. In L. Wilkerson, & W. H. Gijselaers (Eds.), *New directions for teaching and learning, Nr.68* (pp. 3–11). San Francisco: Jossey-Bass Publishers.

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*, 52–54.

Baxter, G. P., & Shavelson, R. J. (1994). Science performance assessments: Benchmarks and surrogates. *International Journal of Educational Research*, *21*(3), 279–299.

Bickley, H., Donner, R. S., Walker, A. N., & Tift, J. P. (1990). Pathology education in a problem-based medical curriculum. *Teaching and Learning in Medicine*, *2*(1), 38–41.

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum, & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston/Dordrecht/London: Kluwer Academic Publishers.

Birenbaum, M., & Dochy, F. (Eds.) (1996). *Alternatives in assessments, learning processes and prior knowledge*. Boston: Kluwer Academic.

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

Boekaerts, M. (1999a). Self-regulated learning: Where are we today? *International Journal of Educational Research*, *31*, 445–457.

Boekaerts, M. (1999b). Motivated learning: The study of student x situation transactional units. *European Journal of Psychology of Education*, *14*(4), 41–55.

Boradge, G. (1987). An alternative approach to PMP's: The 'key-features' concept. In I. R. Hart, & R.

Harden (Eds.), Proceedings of the second ottawa conference. *Further developments in assessing clinical competences* (pp. 59–75). Montreal: Can-Heal Publications Inc.

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, & H. J. M. van Berkel (Eds.), *Problem-based learning as an educational strategy* (pp. 33–41). Maastricht: Network Publications.

Bruner, J. S. (1959). Learning and thinking. *Harvard Educational Review*, *29*, 184–192.

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

Chi, M. T., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–76). Hillsdale, NJ: Erlbaum.

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

Cooper, H. M. (1989). *Integrating research. A guide for literature reviews (Applied Social Research Methods Series, Vol. 2)*. London: Sage Publications.

De Corte, E. (1990a). A State-of-the-art of research on learning and teaching. Keynote lecture presented at the first European Conference on the First Year Experience in Higher Education, Aalborg University, Denmark, April 23-25.

De Corte, E. (1990b). Toward powerful learning environments for the acquisition of problem-solving skills. *European Journal of Psychology of Education*, *5*(1), 5–19.

De Corte, E. (1995). Fostering cognitive growth: A perspective from research on mathematics learning and instruction. *Educational Psychologist*, *30*(1), 37–46.

De Corte, E. (2000). Marrying theory building and the improvement of school practice: A permanent challenge for instructional psychology. *Learning & Instruction*, *10*(3), 249–266.

Dewey, J. (1910). *How we think*. Boston: Health & Co.

Dewey, J. (1944). *Democracy and education*. New York: Macmillan Publishing Co.

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., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, *69*(2), 145–186.

Dochy, F. J. R. C., & Alexander, P. A. (1995). Mapping prior knowledge: A framework for discussion among researchers. *European Journal for Psychology of Education*, *X*(3), 225–242.

Donner, R. S., & Bickley, H. (1990). Problem-based learning: An assessment of its feasibility and cost. *Human Pathology*, *21*, 881–885.

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.

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, September suppl), 11–12.

Engel, C. E. (1997). Not just a method but a way of learning. In D. Bound, & G. Feletti (Eds.), *The challenge of problem based learning (2nd ed.)* (pp. 17–27). London: Kogan Page.

Farquhar, L. J., Haf, J., & Kotabe, K. (1986). Effect of two preclinical curricula on NMBE part I examination performance. *Journal of Medical Education*, *61*, 368–373.

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

Gagné, E. D. (1978). Long-term retention of information following learning from prose. *Review of Educational Research*, *48*, 629–665.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research (6th ed.)*. White Plains, NY: Longman.

Glaser, R. (1990). Toward new models for assessment. *International Journal of Educational Research*, *14*, 475–483.

Glass, G. V. (1976). Primary, secondary and meta-analysis. *Educational Researcher*, *5*, 3–8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. London: Sage Publications.

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. *Journal of the American Medical Education*, *265*, 2373–2376.

Hedges, L. V., & Olkin, I. (1980). Vote counting methods in research synthesis. *Psychological Bulletin*, *88*, 359–369.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hmelo, C. E. (1998). Problem-based learning: Effects on the early acquisition of cognitive skill in medicine. *The Journal of the Learning Sciences*, *7*, 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*, 387–408.

Honebein, P. C., Duffy, T. M., & Fishman, B. J. (1993). Constructivism and the design of learning environments: Context and authentic activities for learning. In T. M. Duffy, J. Lowyck, & D. H. Jonassen (Eds.), *Designing environments for constructive learning*. Berlin: Springer Verlag.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis. Correcting error and bias in research findings*. California: Sage Publications.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverley Hills, CA: Sage.

Imbos, T., & Verwijnen, G. M. (1982). Voortgangstoetsing aan de medische faculteit Maastricht [Progress testing on the Faculty of Medicine of Maastricht]. in Dutch In H. G. Schmidt (Ed.), *Probleemgestuurd Onderwijs: Bijdragen tot onderwijsresearchdagen 1981* (pp. 45–56). Stichting voor Onderzoek van het Onderwijs.

Imbos, T., Drukker, J., van Mameren, H., & Verwijnen, M. (1984). The growth in knowledge of anatomy in a problem-based curriculum. In H. G. Schmidt, & M. L. de Volder (Eds.), *Tutorials in problem-based learning. New direction in training for the health professions* (pp. 106–115). Assen: Van Gorcum.

Jones, J. W., Bieber, L. L., Echt, R., Scheifley, V., & Ways, P. O. (1984). A Problem-based curriculum, ten years of experience. In H. G. Scmidt, & M. L. de Volder (Eds.), *Tutorials in problem-based learning. New direction in training for the health professions* (pp. 181–198). Assen: Van Gorcum.

Kaufman, A., Mennin, S., Waterman, R., Duban, S., Hansbarger, C., Silverblatt, H., Obenshain, S. S., Kantrowitz, M., Becker, T., Samet, J., & Wiese, W. (1989). The New Mexico experiment: Educational innovation and institutional change. *Academic Medicine*, *64*, 285–294.

Knox, J. D. E. (1989). What is… a modified essay question? *Medical Teacher*, *11*(1), 51–55.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746–759.

Kulik, J. A., & Kulik, C. L. (1989). The concept of meta-analysis. *International Journal of Educational Research*, *13*(3), 227–234.

Lewis, K. E., & Tamblyn, R. M. (1987). The problem-based learning approach in Baccalaureate nursing education: How effective is it? *Nursing Papers*, *19*(2), 17–26.

Mandl, H., Gruber, H., & Renkl, A. (1996). Communities of practice toward expertise: Social foundation of university instruction. In P. B. Bates, & U. M. Staudinger (Eds.), *Interactive minds. Life-span perspectives on the social foundation of cognition* (pp. 394–412). Cambridge: Cambridge University Press.

Martenson, D., Eriksson, H., & Ingelman-Sundberg, M. (1985). Medical chemistry: Evaluation of active and problem-oriented teaching methods. *Medical Education*, *19*, 34–42.

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

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology*. New York: Holt, Rinehard and Winston.

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

Morgan, H. R. (1977). A problem-oriented independent studies programme in basic medical sciences. *Medical Education*, *11*, 394–398.

Neufeld, V., & Sibley, J. (1989). Evaluation of health sciences education programs: Program and student assessment at McMaster University. In H. G. Scmidt, M. Lipkinjr, M. W. Vries, & J. M. Greep (Eds.), *New directions for medical education: Problem-based learning and community-oriented medical education* (pp. 165–179). New-York: Springer Verlag.

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

Neufeld, V. R., Woodward, C. A., & MacLeod, S. M. (1989). The McMaster M.D. Program: A case study of renewal in medical education. *Academic Medicine*, *64*, 423–432.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company*. New York: Oxford University Press.

Owen, E., Stephens, M., Moskowitz, J., & Guillermo, G. (2000). From 'horse race' to educational improvement: The future of international educational assessments. In: INES (OECD), The INES compendium (pp. 7-18), Tokyo, Japan. (http://www.pisa.oecd.org/Docs/Download/GA(2000)12.pdf)

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

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Poikela, E., & Poikela, S. (1997). Conceptions of learning and knowledge-impacts on the implementation of problem-based learning. *Zeitschrift fur Hochschuldidactik*, *1*, 8–21.

Quinn, J. B. (1992). *Intelligent enterprise, a knowledge and service based paradigm for industry*. New York: The Free Press.

Richards, B. F., Ober, P., Cariaga-Lo, L., Camp, M. G., Philp, J., McFarlane, M., Rupp, R., & Zaccaro, D. J. (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.

Rogers, C. R. (1969). *Freedom to learn*. Colombus, Ohio: Charles E. Merill Publishing Company.

Rossi, P., & Wright, S. (1977). Evaluation resarch: An assessment of theory, practice and politics. *Evaluation Quarterly*, *1*, 5–52.

Salganik, L. H., Rychen, D. S., Moser, U., & Konstant, J. W. (1999). *Projects on competencies in the OECD context: Analysis of theoretical and conceptual foundations*. Neuchâtel: SFSO, OECD, ESSI.

Santos-Gomez, L., Kalishman, S., Rezler, A., Skipper, B., & Mennin, S. P. (1990). Residency performance of graduates from a problem-based and a conventional curriculum. *Medical Education*, *24*, 366–377.

Saunders, N. A., Mcintosh, J., Mcpherson, J., & Engel, C. E. (1990). A comparison between University of Newcastle and University of Sydney final-year students: Knowledge and 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.

Schmidt, H. G. (1990). Innovative and conventional curricula compared: What can be said about their effects? In Z. H. Nooman, H. G. Schmidt, & E. S. Ezzat (Eds.), *Innovation in medical education: An evaluation of its present status* (pp. 1–7). New York: Springer.

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

Schwartz, R. W., Burgett, J. E., Blue, A. V., Donnelly, M. B., & Sloan, D. A. (1997). Problem-based learning and performance-based testing: Effective alternatives for undergraduate surgical education and assessment of student performance. *Medical Teacher*, *19*, 19–23.

Segers, M. S. R. (1996). Assessment in a problem-based economics curriculum. In M. Birenbaum, & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior learning* (pp. 201–226). Boston: Kluwer Academic Press.

Segers, M. S. R. (1997). An alternative for assessing problem-solving skills: The OverAll Test. *Studies in Educational Evaluation*, *23*(4), 373–398.

Segers, M., Dochy, F., & De Corte, E. (1999). Assessment practices and students' knowledge profiles in a problem-based curriculum. *Learning Environments Research*, *12*(2), 191–213.

Shavelson, R. J., Gao, X., & Baxter, G. P. (1996). On the content validity of performance assessments: Centrality of domain specification. In M. Birenbaum, & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior learning* (pp. 131–143). Boston: Kluwer Academic Press.

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

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, *69*(1), 21–51.

Swanson, D. B., Case, S. M., & van der Vleuten, C. P. M. (1997). Strategies for student assessment. In D. Boud, & G. Feletti (Eds.), *The challenge of problem-based learning [2nd ed.]* (pp. 269–282). London: Kogan Page.

Tans, R. W., Schmidt, H. G., Schade-Hoogeveen, B. E. J., & Gijselaers, W. H. (1986). Sturing van het onderwijsleerproces door middel van problemen: een veldexperiment [Guiding the learning process by means of problems: A field experiment]. *Tijdschrift voor Onderwijsresearch*, *11*, 35–46.

Tynjälä, P. (1999). Towards expert knowledge? A comparison between a constructivist and a traditional learning environment in the University. *International Journal of Educational Research*, *33*, 355–442.

Van Hessen, P. A. W., & Verwijnen, G. M. (1990). Does problem-based learning provide other knowledge. In W. Bender, R. J. Hiemstra, A. J. J. A. Scherpbier, & R. P. Zwierstra (Eds.), *Teaching and assessing clinical competence* (pp. 446–451). Groningen: Boekwerk Publications.

Van Ijzendoorn, M. H. (1997). Meta-analysis in early childhood education: Progress and problems. In B. Spodek, A. D. Pellegrini, & O. N. Saracho (Eds.), *Issues in early childhood education yearbook in early childhood education*. New York: Teachers College Press.

Verhoeven, B. H., Verwijnen, G. M., Scherpbier, A. J. J. A., Holdrinet, R. S. G., Oeseburg, B., Bulte, J. A., & Van Der Vleuten, C. P. M. (1998). An analysis of progress test results of PBL and non-PBL students. *Medical Teacher*, *20*(4), 310–316.

Vernon, D. T. A., & Blake, R. L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine*, *68*, 550–563.

Verwijnen, G. M., Pollemans, M. C., & Wijnen, W. H. F. W. (1995). Voortgangstoetsing [Progress testing]. In J. C. M. Melz, A. J. J. A. Scherpbier, & C. P. M. Van der Vleuten (Eds.), *Medisch Onderwijs in de Praktijk* (pp. 225–232). Assen: Van Gorcum.

Verwijnen, M., Imbos, T., Snellen, H., Stalenhoef, B., Sprooten, Y., & Van der Vleuten, C. (1982). The evaluation system at the medical school of Maastricht. In H. G. Schmidt, M. Vries, & J. M. Greep (Eds.), *New directions for medical education: Problem-based learning and community-oriented medical education* (pp. 165–179). New York: Springer.

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 Z. H. Nooman, H. G. Schmidt, & E. S. Ezzat (Eds.), *Innovation in medical education: An evaluation of its present status* (pp. 41–49). New York: Springer.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*, 345–376.