

# Differential Student Attrition and Differential Exposure Mask Effects of Problem-Based Learning in Curriculum Comparison Studies

Henk G. Schmidt, PhD, Arno M. M. Muijtjens, PhD, Cees P. M. Van der Vleuten, PhD, and Geoffrey R. Norman, PhD

## Abstract

### Purpose

Studies that compare the effects of problem-based and conventional medical curricula on student performance generally show no significant differences. However, curriculum comparison studies are at best quasi-experimental, so they are prone to forms of selection bias. The purpose of this study was to reanalyze data from such comparisons by controlling for two potential biases: differential student attrition and differential exposure.

### Method

The authors reanalyzed 104 previously published comparisons involving a single, problem-based medical school in the

Netherlands (Maastricht University's medical school), using student attrition and study duration data from this school and the schools with which it was compared. The authors removed bias by reequalizing the comparison groups in terms of attrition and study duration.

### Results

The uncorrected data showed no differences between problem-based and conventional curricula: Mean effect sizes as expressed by Cohen *d* were 0.02 for medical knowledge and 0.07 for diagnostic reasoning. However, the reanalysis demonstrated medium-level effect sizes favoring the problem-based curriculum. After corrections for attrition

and study duration, the mean effect size for knowledge acquisition was 0.31 and for diagnostic reasoning was 0.51.

### Conclusions

Effects of the Maastricht problem-based curriculum were masked by differential attrition and differential exposure in the original studies. Because this school has been involved in many studies included in influential literature reviews published in the past 20 years, the authors' findings have implications for the assessment of the value of problem-based learning put forward by these reviews.

Since problem-based learning (PBL) was introduced into the medical curriculum at the McMaster University Faculty of Health Sciences in 1969, the teaching method has enjoyed considerable success. Seventy percent of

U.S. medical schools include in their curricula small-group tutorial sessions organized around patient cases, and 6% describe their curricula as problem based outright.<sup>1</sup> The majority of Australian medical schools have adopted PBL as an important aspect of their teaching, and PBL has also made inroads in European and Asian medical education.<sup>2–6</sup>

Despite PBL's apparent popularity among medical educators and students,<sup>7–10</sup> its effectiveness has become the subject of substantial debate.<sup>11–14</sup> Colliver,<sup>11</sup> for instance, discussed a number of studies that compared medical knowledge test performance of students in problem-based medical curricula with the performance of students in conventional medical curricula. On the basis of his interpretation of three 1993 reviews<sup>7,10,15</sup> and eight additional curriculum comparison (CC) studies published after 1993, Colliver<sup>11</sup> concluded that evidence is lacking that PBL improves learning, because differences on knowledge tests, if any, are small and do not consistently favor problem-based curricula.

Additional potential support for Colliver's<sup>11</sup> claim comes from two more

recent reviews. Dochy and colleagues<sup>16</sup> surveyed 43 CC articles and found what they called "a tendency to negative results" when considering the effect of PBL on medical students' knowledge acquisition. Schmidt and colleagues<sup>9</sup> reviewed 108 CC studies that involved a specific, problem-based medical school and spanned a period of almost 30 years. They found an almost negligible overall effect of PBL on students' knowledge acquisition.

### The Possibility of Selection Bias

Do the findings reported in the aforementioned reviews provide direct evidence for Colliver's<sup>11</sup> claim that PBL has no particular effect on medical students' knowledge acquisition? CC studies are at best quasi-experimental studies—they fall short in terms of several of the requirements of true experiments. For instance, however desirable randomization may be, random allocation of participants to the different conditions of the experiment is virtually impossible in natural settings.<sup>17</sup> Further, quasi-experimental studies are prone to several biases threatening their *internal validity*, that is, the extent to which

**Dr. Schmidt** is professor, Department of Psychology, Erasmus University, Rotterdam, the Netherlands.

**Dr. Muijtjens** is associate professor, Department of Educational Development and Research, Maastricht University Medical Center, Maastricht, the Netherlands.

**Dr. Van der Vleuten** is professor, Department of Educational Development and Research, Maastricht University Medical Center, Maastricht, the Netherlands.

**Dr. Norman** is professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

Correspondence should be addressed to Dr. Schmidt, Department of Psychology, Erasmus University, Burgemeester Oudlaan 50, Building A, Room A 2-02, 3062 PA Rotterdam, the Netherlands; telephone: (+31) 10-408-1749; e-mail: schmidt@fsw.eur.nl.

*Acad Med.* 2012;87:463–475.  
First published online February 22, 2012  
doi: 10.1097/ACM.0b013e318249591a

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/A82>.

findings can really be attributed to the treatment studied and not to some other factor or influence.<sup>18</sup>

In this article, we will first argue that findings reported in the CC literature<sup>7–11</sup> may result from what Cook and Campbell called “selection bias.”<sup>18</sup> There are four sources of possible selection bias in these CC studies: differential enrollment, differential sampling, differential attrition, and differential exposure. We describe each briefly below, using examples drawn from the CC literature, including the reviews described above and the studies on which they were based. Subsequently, we will present a reanalysis of 104 previously published CCs involving a single, prominent, problem-based medical school in the Netherlands and various conventional Dutch medical schools. The goal of our reanalysis was to remove the influence of differential attrition and differential exposure on the data these CC studies produced in order to demonstrate that selection bias influenced the outcomes of a sizeable number of these comparisons.

We begin with some observations regarding sources of selection bias in previously published CC studies and reviews.

### Differential enrollment

An assumption central to any comparison study is that the samples or populations involved are sufficiently similar in terms of their knowledge or aptitude prior to enrolling in the experiment, because prior knowledge and aptitude both influence the rate of subsequent learning.<sup>19</sup> The effect of student differences was examined in a recent study that analyzed the performance of students from 116 U.S. medical schools on the United States Medical Licensing Examination across an 11-year period.<sup>20</sup> Performance was predicted using student variables (MCAT score, grade point average [GPA], age, minority status). Preexisting differences between students accounted for most of the variance found in performance.

In some published CC studies, students were known to be different in terms of prior knowledge and aptitude. For example, McMaster University’s medical school uses a criterion heavily weighted to “personal qualities” to select students, and those students’ GPAs tend to be

significantly lower than the GPAs of students at other medical schools.<sup>21</sup> Therefore, the results of several early CC studies in which McMaster students did not do well<sup>22–24</sup> can be attributed to differences in students’ prior knowledge rather than to curricular differences. It is, therefore, essential that studies on curriculum effects include information about students’ entry-level performance. None of the eight studies included in Albanese and Mitchell’s<sup>7</sup> review contained such information. The same was true for four of the post-1993 studies included in Colliver’s<sup>11</sup> review. Similarly, 46 of the 57 comparisons in Dochy and colleagues’<sup>16</sup> review lacked this information.

### Differential sampling

In a number of published CC studies, *populations* of medical students from problem-based schools were compared with samples of *volunteers* from conventional schools. However, students who volunteer for CC studies tend to be the better students of their class—poorly performing students do not readily volunteer to have their knowledge level tested—which may bias results in favor of the volunteer samples. Most studies of this type included in the reviews did not include data that would allow one to check whether volunteers represented the better students; the two that did<sup>25,26</sup> demonstrated that those who volunteered to participate in CC studies were indeed among the better students of their class. Of the reviews mentioned above, only Schmidt et al<sup>9</sup> provided systematic information regarding differential sampling.

### Differential attrition

In U.S. and Canadian medical schools, gaining admission to medical school is a virtual guarantee of future entry into practice. Data from the Association of American Medical Colleges<sup>27</sup> suggest that the attrition rate in U.S. schools is less than 3%. This is not the case in those countries where students enter medical school following high school graduation. In Mexico, for example, fewer than 60% of medical students graduate with their entering class, although eventually 90% will graduate.<sup>28</sup> Statistics in Sweden are similar, with 50% of students graduating “on-time” and 85% to 90% eventually graduating.<sup>29</sup>

Recently, studies have begun to demonstrate that schools with problem-based curricula have higher graduation rates and shorter study duration than schools with conventional curricula.<sup>30–34</sup> The case of Harvard’s dental school is instructive<sup>31</sup>: Before the introduction of PBL in its basic sciences curriculum, the dental school’s attrition rate was almost 27% (i.e., of every 100 students entering the school, 27 never graduated). After PBL was introduced, the attrition rate dropped to 17%. And, after PBL was implemented throughout the curriculum, the attrition rate fell to less than 7%. A large-scale study<sup>33</sup> involving all 10 classes of students who entered the eight medical schools in the Netherlands between 1989 and 1998 (in total, almost 14,000 students) demonstrated that the country’s three problem-based schools had, on average, almost 8% better graduation rates than did the country’s conventional schools, even though graduation rates in the country were already reasonably high (over 80%). If conventional schools lose more students in the course of training than do problem-based schools, then studies involving comparisons of PBL and conventional curricula are prone to attrition bias.

### Differential exposure

Poorly performing students may drop out, particularly in the early years of medical school. But, more commonly, they simply take longer to graduate. In U.S. and Canadian schools, students are more likely to repeat a year than to drop out. In European schools, it is common for individual students to extend their studies by weeks, months, or years. These students benefit from the increased time dedicated to learning; there is a well-established linear relationship between time spent on learning and learning outcomes.<sup>35,36</sup>

In the context of CC studies, if problem-based schools have a higher proportion of “on-time” graduates than do conventional schools, then effectively the PBL curriculum is shorter than the conventional curriculum; that is, it has shorter study duration or time of exposure. There is evidence that this is the case. In the large-scale Dutch study<sup>33</sup> discussed above, students in the three problem-based schools needed, on average, almost half a year less time to

graduate than did students of the five conventional schools.

### Potential Effects of Attrition and Exposure Bias in CC Studies

What are the possible effects of attrition bias, if ignored? Assume for a moment that a particular school's PBL curriculum results in greater student knowledge gains than does another school's conventional curriculum. Assume in addition that students drop out of medical school largely because they perform poorly on examinations. A direct consequence is that, of these two medical schools, the one with the conventional curriculum will lose more students than will the one with the PBL curriculum because of students' poor performance. Students who drop out are, by definition, excluded from CC studies, which involve performance only of those students who have stayed. Therefore, attrition bias in CC studies may mask a potential effect of PBL because the final student cohort in the school with the PBL curriculum contains a higher proportion of the type of students who would have dropped out had they been enrolled at the school with the conventional curriculum.

An example from the drug evaluation field may serve as an illustration: If the effects of drug A and drug B are compared in a randomized controlled trial, one possible outcome is that the surviving patients will report similar levels of recovery, leading to the conclusion that both treatments are equally good (as is the case in most CC studies). However, if patients using drug A have a mortality rate that is higher than that of patients using drug B, the actual effect of drug B is *masked* by this differential mortality. Similarly, differential attrition may mask a curriculum's genuine effects on student learning. The problematic effects of differential attrition on outcomes have been identified in randomized controlled trials<sup>37</sup> but have not yet been noted in CC studies.

The same applies to differences in time needed to complete medical school, or study duration. If students in a conventional school take longer to complete their studies than do those in a problem-based school, potential performance differences favoring the PBL curriculum may be masked by differential exposure.

At this point, we have demonstrated that at least some of the published CC studies included in reviews of the literature may have suffered from differential attrition and differential exposure, which may mask effects of PBL on knowledge acquisition in these studies. This implies that the outcomes of these studies cannot be taken at face value. Some reanalysis of these studies' data is therefore needed, ideally including performance data from students who dropped out. This, though, poses a problem; usually, no such data would be available. However, because students tend to drop out because of poor performance, one can reequalize the comparison groups by excluding the poorest-performing students from the group with the lower attrition rate.

In our reanalysis of a large number of comparisons, we took into account between-school differences in graduation rates as well as time of exposure to the curricula. We did the latter by predicting performance scores among the problem-based school's students as if they were exposed to the same number of months of the treatment as the students from the conventional schools.

### Method

We limited our reanalysis to the studies included in Schmidt and colleagues'<sup>9</sup> 2009 review that involved comparisons of Maastricht University's problem-based medical school (Maastricht) with various conventional Dutch medical schools. There were a number of advantages to using these data from the Netherlands. First, and most important, sufficiently reliable graduation rate and study duration information were available for Maastricht and for the conventional Dutch schools with which it was compared, enabling us to correct for possible attrition bias and differences in exposure to the educational treatments.

Second, Dutch students are admitted to medical schools through a centralized process, which employs a weighted lottery procedure that is based on achievement on a national entrance examination. All eight Dutch medical schools are state based and have an essentially equivalent volume of student intake (approximately 300–400 students). As in most European countries, students in the Netherlands

enter medical school after high school graduation and are about 18 to 19 years old. For these reasons, students in the eight schools are highly similar in terms of past performance, age, and gender,<sup>38</sup> which makes enrollment bias unlikely in CC studies involving these schools. In addition, all Dutch medical schools employ a six-year curriculum, and the subject matter taught is largely overlapping.<sup>39</sup> Because there are no national licensing examinations in the Netherlands, all medical schools are required to have rigorous assessment programs; these are intensively monitored and audited in the national accreditation system. Again, this facilitates comparisons between curricula types.

Third, Maastricht's medical school curriculum is by far the most extensively studied PBL curriculum in the world.<sup>9</sup> It was established in 1974, five years after McMaster University admitted its first group of students to its problem-based program. Therefore, findings based on the Maastricht PBL curriculum appear prominently in the various reviews described above and have contributed greatly to the "no difference" conclusions reached by their authors.<sup>7–12</sup> The curricula of the other Dutch medical schools in these comparisons are conventional to varying degrees, with a continued emphasis on lectures.<sup>33</sup>

### Study selection procedures

Schmidt and colleagues<sup>9</sup> initiated a literature search in 2009 that used the ISI Web of Science, PsycINFO, Educational Resources Information Center, and PubMed databases and covered the years 1974–2009. The key words they used were *problem-based learning*, *curriculum*, *comparison*, and *Maastricht*, as well as all possible combinations and permutations of these terms. The Google Scholar and Scopus search engines were used to identify additional sources. Finally, online PBL archives were searched. These searches identified 19 articles and book chapters in which a total of 278 comparisons were reported.<sup>9</sup>

For the study reported here, we selected 104 of the 278 comparisons for reanalysis. Our inclusion criteria were (1) Maastricht students were compared with students from at least one conventional Dutch medical school, (2) the comparison concerned performance on tests of medical



knowledge or diagnostic reasoning, and (3) both attrition and study duration data were available for each of the groups compared.

### The original datasets

As noted above, Schmidt and colleagues<sup>9</sup> datasets form the basis of the reanalysis of 104 comparisons reported here. Seventy-eight of these 104 comparisons were conducted with a medical knowledge test procedure called the “progress test,” which was developed originally at Maastricht and is now in use in half of Dutch medical schools<sup>40,41</sup> as well as at schools elsewhere.<sup>42–44</sup> A typical progress test consists of some 200 questions covering medicine as a whole. The test is administered four times each year to all students, regardless of their year of study. For each administration, a new version of the test is drawn from a large item bank. Students must demonstrate progress on each subsequent test, which is why it is referred to as a “progress” test. It has high reliability, with Cronbach alphas typically above 0.80 within years and above 0.90 across years. Evidence for its validity has been reported as well.<sup>45,46</sup> Of the other 26 comparisons in this reanalysis, 10 were made using a test of anatomy knowledge, and 16 focused on the application of knowledge in diagnostic reasoning tasks. These instruments all had reliability coefficients above 0.80.

For each of the comparisons, Schmidt and colleagues<sup>9</sup> identified the medical schools involved, whether populations or (random) samples were studied, and the class of the students involved. They extracted measurement information, if provided, together with relevant procedural information. Finally, they computed a measure of the strength of the effect of PBL for each comparison, expressed as Cohen effect size *d*.<sup>47</sup> See Appendix 1 and Appendix 2 for these details of the 104 comparisons we selected for our reanalysis.

Because Dutch medical students spend six years in medical school, many of the CC studies made within-class comparisons among first-, second-, third-, fourth-, fifth-, and sixth-year students. Schmidt et al<sup>9</sup> did not report all these comparisons separately. In cases where multiple comparisons were made, they reported mean *d* values (printed in

italic type in Appendix 2) weighted according to the numbers of students in the contributing comparison groups.

Comparisons were not always reported in the original articles in a way that made direct computation of *d* values possible. In two instances, Schmidt et al<sup>9</sup> had to reconstruct the data from graphs presented in the article.<sup>25,48</sup>

The attrition rate and study duration data for each of the comparisons are made available by the VSNU, the Dutch Association of Universities (see also Schmidt and colleagues,<sup>33</sup> Post et al,<sup>34</sup> and Verwijnen et al<sup>48</sup>). The means presented in columns 3 to 6 in Appendix 2 were computed on the basis of the data from each of the classes involved in the comparisons.

### Reanalysis of the Schmidt et al<sup>9</sup> datasets: The Amsterdam example

Our approach to removing attrition and exposure bias from the comparisons can best be illustrated using as an example a 1996 study<sup>49</sup> that features prominently in the various published reviews and is reanalyzed here. This study compared diagnostic reasoning competence among students at three Dutch medical schools, including the problem-based Maastricht medical school. At each of these schools, five classes of students were presented with 30 cases and required to provide a diagnosis for each case. Figure 1, panel A, displays mean scores for each of these classes at each school, as presented in the original study. The critical statistics for this study are available in Appendix 2.

For simplicity, we concentrate here on just one of the 1996 study's comparison schools: the University of Amsterdam's medical school. Comparing the performance of students in Maastricht's PBL curriculum with that of students in Amsterdam's conventional curriculum favored the latter: Schmidt and colleagues<sup>9</sup> reported the mean effect size *d* as  $-0.36$ . However, mean graduation rates of the participating classes were 93% for Maastricht versus 82% for Amsterdam. To match these groups in terms of attrition rate, we excluded from our data the poorest-performing 11% of students from Maastricht, so that both schools had the same attrition rate. Figure 1, panel B, displays the mean scores for each class at each school after attrition bias is taken into account. We

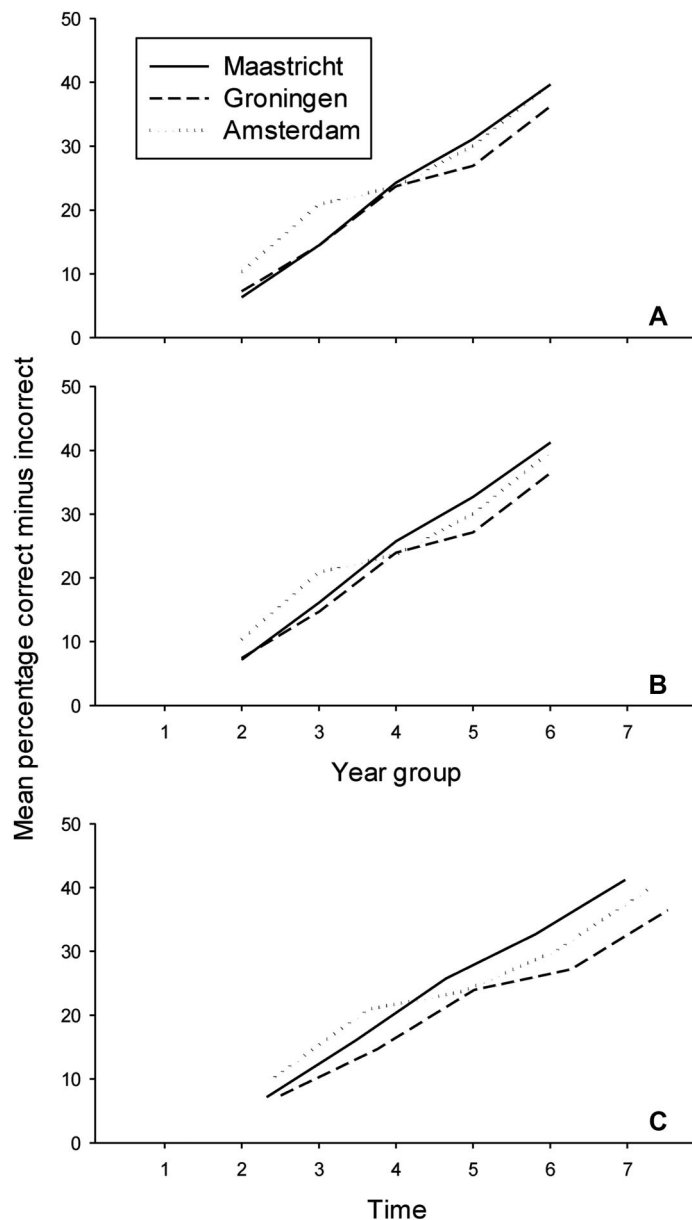
recomputed the mean effect size, which resulted in a *d* value of  $-0.15$ .

Finally, we considered differences in study duration. Maastricht students in the participating classes eventually graduated in an average of 6.97 years, whereas the Amsterdam students graduated in an average of 7.28 years. This difference was taken into account in the data as represented in Figure 1, panel C, in which the graphs have been remodeled to represent actual study duration rather than the nominal study duration of six years. (In Figure 1, panel A, results for many of the students who were actually in their seventh, eighth, or ninth year of study are displayed as if these students were in their fourth, fifth, or sixth year.) On the basis of Figure 1, panel C, we interpolated the projected mean scores for each of the classes, computed new *d* values, and recomputed the mean effect size, which was equal to 0.09. So as a result of corrections, the *d* value of this comparison changes from medium negative to slightly positive for the PBL curriculum. We reanalyzed the other comparisons in similar fashion. We explain the mathematical technicalities of the procedure in Supplemental Digital Appendix 1 (available at <http://links.lww.com/ACADMED/A82>).

### Results

Appendix 2 displays the results of our reanalysis of 104 comparisons of one Dutch medical school's problem-based curriculum with conventional curricula at other Dutch medical schools. The appendix's third through sixth columns present mean attrition and study duration data derived from various sources.<sup>33,34</sup> These data represent averages across the classes involved in the comparisons—that is, the average success across classes in terms of graduation rate and duration of study. The seventh column contains the effect sizes reported in the original review.<sup>9</sup> The eighth and ninth columns present the effect sizes we calculated in our reanalyses. Table 1 presents summary statistics from the original study and our reanalysis.

As we predicted, because the attrition rate was consistently lower and the duration of study was consistently shorter in the problem-based school (except in Verhoeven and colleagues<sup>26</sup> study), the



**Figure 1** Visualization of effects of correcting for differences in attrition and study duration on diagnostic reasoning findings.<sup>46</sup> Progress test scores (percentage correct minus incorrect) are used to compare students' knowledge acquisition at three Dutch medical schools: Maastricht, Groningen, and Amsterdam, all of which use a six-year curriculum. Panel A shows the observed mean scores of students in years 2 through 6, as reported in Schmidt and colleagues.<sup>49</sup> Panel B shows the corresponding scores after we corrected for differential attrition, and panel C shows the scores after additional correction for differential exposure. The patterns in panel C were obtained by stretching the curves to have them comply with the average time (in years) that students at each school required to graduate. (Panel A is adapted from Schmidt and colleagues.<sup>49</sup>)

differences favoring the problem-based school become larger. After we corrected for attrition, the overall effect size  $d$  for the acquisition of medical knowledge, which was 0.02 in the original review,<sup>9</sup> increased to 0.18. After we corrected for differences in study duration, the overall effect size increased to  $d = 0.31$ . The effect was similar for diagnostic

reasoning: After we corrected for attrition, the overall  $d$  value increased from 0.07 to 0.27, and then to 0.51 after we corrected for study duration. These findings suggest that in the 104 comparisons we reanalyzed, robust positive effects of PBL were indeed masked by differential attrition and differential exposure.

## Discussion

Previous reviews of curricular effects on medical knowledge acquisition and knowledge use in diagnostic reasoning generally have shown that the performance of medical students from schools with PBL curricula is not uniformly superior to that of students whose schools' curricula employ more direct forms of instruction, such as lectures.<sup>7-11,15,16</sup> As a consequence, the use of PBL in medical education continues to be regarded with skepticism. In a recent study, Kirschner and colleagues<sup>12</sup> argued that instructional approaches emphasizing minimal direct instruction (including PBL), although intuitively appealing, are less effective than those emphasizing teacher guidance, because the former type would ignore the structure of the human cognitive system and its limitations of working memory. In reaching this conclusion, Kirschner and colleagues explicitly referred to the Colliver<sup>11</sup> review.

As stated earlier, this study's central thesis is that many of the CC studies on which PBL critics have based their conclusions may have been subject to forms of selection bias. To test our hypothesis, we reanalyzed 104 curricular comparisons involving Maastricht University's medical school, which is one of the most prominent problem-based schools. This particular school has contributed a large number of CC studies to the literature, so it has had a potentially large influence on the conclusions of systematic reviews conducted to date on PBL in medical education. If these studies were shown to contain biased estimations of the effect of PBL on knowledge acquisition, then the conclusions of the reviews themselves could be called into question.

The findings of our reanalysis supported these hypotheses. Corrections for differential attrition and study duration resulted in medium-sized positive effects of PBL in most of the 104 comparisons we reanalyzed. Because effect sizes can be interpreted as  $z$  scores, the findings as summarized in Table 1 imply that the average student from the Maastricht problem-based curriculum surpasses 62% of his or her fellow Dutch medical students in conventional curricula on medical knowledge tests and 70% of them on diagnostic reasoning tests. In these previously published comparisons, effects of PBL were, therefore, clearly

Table 1

**Summary Statistics of the Original Study and the Effects of Correcting for Differences in Attrition and Study Duration, as Demonstrated by the Reanalysis**

Domain	No. of comparisons included	No. of comparison schools involved	Mean effect size d in original dataset*	Mean d after correction for attrition	Mean d after correction for attrition and study duration
Medical knowledge	88	7	0.02	0.18	0.31
Diagnostic reasoning	16	2	0.07	0.27	0.51

\* Effect sizes for the original dataset are from Schmidt et al, 2009.<sup>9</sup>

masked by differential attrition and differential exposure.

Our study has a number of limitations. The first pertains to our decision to reequalize the comparison groups by excluding the poorest-performing students from the problem-based school sample. Although the decision to leave medical school may not be simply a consequence of poor performance, research has demonstrated that there is a strong relationship between medical students' level of performance and the probability that they will drop out. Roeleveld<sup>38</sup> found that most of the poorest-performing medical students in the Netherlands had dropped out by the end of their second year. A study by Van der Vleuten and colleagues<sup>41</sup> showed that medical students who eventually dropped out of the program had a mean score on previous progress tests that was, on average, 0.65 standard deviations below the class mean. We therefore felt that our decision to exclude the poorest-performing students from the problem-based groups in order to reequalize them with the conventional groups was justified.

However, it is not entirely likely that all students who drop out are from the bottom of their class. We therefore conducted a sensitivity analysis in which we found that if half of the students who dropped out did so because of poor performance and the rest did so at random, then the effect sizes for knowledge acquisition would be  $d = 0.11$  after correction for attrition and  $d = 0.23$  after additional correction for study duration. For diagnostic reasoning, the respective corrected effect sizes would be  $d = 0.18$  and  $d = 0.41$ . If one-third of the students who dropped out did so because of poor performance and the other students did so at random, then the effect sizes for knowledge acquisition would be  $d = 0.08$  after correction for attrition and  $d = 0.20$  after additional correction for study duration. For diagnostic reasoning, the respective corrected effect sizes would be

$d = 0.15$  and  $d = 0.37$ . So, even in the highly unlikely case that two-thirds of the students who dropped out did so for reasons other than poor performance, positive effects of PBL are still visible in the data, attesting to their robustness.

A second limitation involves the generalizability of our findings, which are based on data from one problem-based school. Perhaps this school—Maastricht University's medical school—has a number of idiosyncratic features that induce better performance in its students and that may not be present at other problem-based schools. To address this possible limitation of our findings, we conducted similar analyses of the two other problem-based schools in the Netherlands—the medical schools of Groningen University and Nijmegen University—for which relevant graduation data were available<sup>33</sup> and that participated in some of the CC studies.<sup>41</sup> The results of these additional analyses confirmed the findings we report in this article, suggesting that our findings have the potential for further generalizability. Also, international data show that problem-based schools tend to have lower levels of student attrition and delayed graduation,<sup>30–32</sup> which suggests that our findings may not be limited to Dutch medical education.

Why would PBL curricula protect students against dropout and study delays better than conventional, lecture-based curricula? In our view, the tutorial group, one of PBL's distinguishing characteristics, plays a central role here.<sup>9</sup> First, the tutorial group is a source of friendships, and it enables students to develop more personal relationships with their teachers than is possible in larger classrooms; both of these factors are considered to be protective against premature dropout.<sup>50,51</sup> Second, regular small-group tutorials in problem-based schools provide peer pressure and natural

deadlines for work to be completed and, therefore, encourage students not to postpone studying. These are reasons why students in medical schools with PBL curricula graduate earlier than students from conventional schools. (See Schmidt et al<sup>52</sup> for an extended discussion on what works in PBL.)

## Conclusions

Our findings seem to have a number of implications for CC research. First, comparisons between different instructional approaches that do not explicitly take selection bias into account should be considered inappropriate. We have demonstrated here that selection bias has nontrivial effects on their outcomes. Second, we suggest that conclusions reached in the reviews of CC studies published in the past 20 years should be used cautiously.<sup>7–11,15,16</sup> We have shown here that robust effects of PBL were actually masked in many of the studies included in these reviews. The possibility that other CC studies have suffered from the same shortcomings cannot be excluded, particularly because problem-based schools other than Maastricht have also begun to report lower attrition rates than conventional comparison schools.<sup>30–32</sup> It is, therefore, unfortunate that calls to discard PBL, based on these reviews, have been relatively influential.<sup>11–13</sup> We believe we have shown that these calls lacked solid empirical foundation. Finally, our findings demonstrate that PBL, as a serious approach to improving medical education, may deserve another, less biased look.

*Acknowledgments:* The authors acknowledge the assistance of Dr. Wilco te Winkel in the collection of the data.

*Funding/Support:* None.

*Other disclosures:* Not applicable.

*Ethical approval:* Not applicable.



## References

- 1 Kinkade S. A snapshot of the status of problem-based learning in U.S. medical schools, 2003–04. *Acad Med.* 2005;80:300–301.
- 2 O'Neill PA, Morris J, Baxter CM. Evaluation of an integrated curriculum using problem-based learning in a clinical environment: The Manchester experience. *Med Educ.* 2000;34:222–230.
- 3 Antepohl W, Herzig S. Problem-based learning versus lecture-based learning in a course of basic pharmacology: A controlled, randomized study. *Med Educ.* 1999;33:106–113.
- 4 Fyrenius A, Silen C, Wirell S. Students' conceptions of underlying principles in medical physiology: An interview study of medical students' understanding in a PBL curriculum. *Adv Physiol Educ.* 2007;31:364–369.
- 5 Tiwari A, Lai P, So M, Yuen K. A comparison of the effects of problem-based learning and lecturing on the development of students' critical thinking. *Med Educ.* 2006;40:547–554.
- 6 Khoo HE. Implementation of problem-based learning in Asian medical schools and students' perceptions of their experience. *Med Educ.* 2003;37:401–409.
- 7 Albanese MA, Mitchell S. Problem-based learning: A review of literature on its outcomes and implementation issues. *Acad Med.* 1993;68:52–81.
- 8 Schmidt HG, Dauphinee WD, Patel VL. Comparing the effects of problem-based and conventional curricula in an international sample. *J Med Educ.* 1987;62:305–315.
- 9 Schmidt HG, Van der Molen HT, Te Winkel WWR, Wijnen WHFW. Constructivist, problem-based learning does work: A meta-analysis of curricular comparisons involving a single medical school. *Educ Psychol.* 2009;44:227–249.
- 10 Vernon DT, Blake RL. Does problem-based learning work? A meta-analysis of evaluative research. *Acad Med.* 1993;68:550–563.
- 11 Colliver JA. Effectiveness of problem-based learning curricula: Research and theory. *Acad Med.* 2000;75:259–266.
- 12 Kirschner PA, Sweller J, Clark RE. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ Psychol.* 2006;41:75–86.
- 13 Shanley PF. 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.* 2007;82:479–485.
- 14 Wittert GA, Nelson AJ. Medical education: Revolution, devolution and evolution in curriculum philosophy and design. *Med J Aust.* 2009;191:35–37.
- 15 Berkson L. Problem-based learning—Have the expectations been met? *Acad Med.* 1993;68(10 suppl):S79–S88.
- 16 Dochy F, Segers M, Van den Bossche P, Gijbels D. Effects of problem-based learning: A meta-analysis. *Learn Instr.* 2003;13:533–568.
- 17 Norman GR, Schmidt HG. Effectiveness of problem-based learning curricula: Theory, practice and paper darts. *Med Educ.* 2000;34:721–728.
- 18 Cook TD, Campbell DT. *Quasi-Experimentation: Design and Analysis for Field Settings.* Chicago, Ill: Rand McNally; 1979.
- 19 Alexander PA, Judy JE. The interaction of domain-specific and strategic knowledge in academic performance. *Rev Educ Res.* 1988;58:375–404.
- 20 Hecker K, Violato C. How much do differences in medical schools influence student performance? A longitudinal study employing hierarchical linear modeling. *Teach Learn Med.* 2008;20:104–113.
- 21 Neufeld VR, Barrows HS. The “McMaster philosophy”: An approach to medical education. *J Med Educ.* 1974;49:1040–1050.
- 22 Neufeld V, Sibley J. Evaluation of health sciences education programs: Program and student assessment at McMaster University. In: Schmidt HG, Lipkin M, Vries MW, Greep JM, eds. *New Directions for Medical Education: Problem-Based Learning and Community-Oriented Medical Education.* New York, NY: Springer Verlag; 1989.
- 23 Neufeld VR, Woodward CA, MacLeod SM. The McMaster M.D. program: A case study of renewal in medical education. *Acad Med.* 1989;64:423–432.
- 24 Patel VL, Groen GJ, Norman GR. Effects of conventional and problem-based medical curricula on problem-solving. *Acad Med.* 1991;66:380–389.
- 25 Prince KJ, van Mameren H, Hylkema N, Drukker J, Scherpbier AJ, van der Vleuten CP. Does problem-based learning lead to deficiencies in basic science knowledge? An empirical case on anatomy. *Med Educ.* 2003;37:15–21.
- 26 Verhoeven BH, Verwijnen GM, Scherpbier A, et al. An analysis of progress test results of PBL and non-PBL students. *Med Teach.* 1998;20:310–316.
- 27 Brewer L, Grbic D. Medical students' socioeconomic background and their completion of the first two years of medical school. *AAMC Analysis in Brief.* December 2010;9(11). [https://www.aamc.org/download/165418/data/aibvol9\\_no11.pdf](https://www.aamc.org/download/165418/data/aibvol9_no11.pdf). Accessed October 4, 2011.
- 28 Mendiola MS. Secretary of medical education, UNAM Faculty of Medicine, Mexico City, Mexico. Personal communication with G.R. Norman, August 12, 2011.
- 29 Hammar M. Dean, Faculty of Health Sciences, Linköping University, Linköping, Sweden. Personal communication with G.R. Norman, August 12, 2011.
- 30 Burch VC, Sikakana CNT, Seggie JL, Schmidt HG. Performance of academically-at-risk medical students in a problem-based learning programme. A preliminary report. *Adv Health Sci Educ Theory Pract.* 2007;12:345–358.
- 31 Howell H. Ten years of hybrid PBL at Harvard School of Dental Medicine. Paper presented at: 4th International Symposium on Problem-Based Learning in Dental Education; October 24, 2005; Nakorn Pathom, Thailand.
- 32 Iputo JE, Kwizera E. Problem-based learning improves the academic performance of medical students in South Africa. *Med Educ.* 2005;39:388–393.
- 33 Schmidt HG, Cohen-Schotanus J, Arends L. Impact of problem-based, active learning on graduation rates of ten generations of Dutch medical students. *Med Educ.* 2009;43:211–218.
- 34 Post GJ, De Graaff E, Drop MJ. Length and output of medical training in Maastricht [in Dutch]. *Ned Tijdschr Geneesk.* 1986;130:1903–1905.
- 35 Bloom BS. Time and learning. *Am Psychol.* 1974;29:682–688.
- 36 Carroll JB. A model for school learning. *Teach Coll Rec.* 1963;64:723–733.
- 37 Juni P, Egger M. Commentary: Empirical evidence of attrition bias in clinical trials. *Int J Epidemiol.* 2005;34:87–88.
- 38 Roeleveld J. Lottery Categories and Academic Achievement: Report to the Advisory Committee Admissions Numerus-Fixus Programs [in Dutch]. Amsterdam, the Netherlands: University of Amsterdam: SCO/Kohnstamm Instituut; 1997.
- 39 Schadé E, Sminia TD. Final objectives for university training of doctors: Framework 1994 for medical training [in Dutch]. *Ned Tijdschr Geneesk.* 1995;139:30–35.
- 40 Muijtjens A, Schuwirth L, Cohen-Schotanus J, van der Vleuten CPM. Differences in knowledge development exposed by multi-curricular progress test data. *Adv Health Sci Educ Theory Pract.* 2008;13:593–605.
- 41 van der Vleuten CP, Schuwirth LW, Muijtjens AM, Thoben AJ, Cohen-Schotanus J, van Boven CP. Cross institutional collaboration in assessment: A case on progress testing. *Med Teach.* 2004;26:719–725.
- 42 Blake JM, Norman GR, Keane DR, Mueller CB, Cunningham J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: Psychometric properties and effect on learning. *Acad Med.* 1996;71:1002–1007.
- 43 Coombes L, Ricketts C, Freeman A, Stratford J. Beyond assessment: Feedback for individuals and institutions based on the progress test. *Med Teach.* 2010;32:486–490.
- 44 Norman G, Neville A, Blake JM, Mueller B. Assessment steers learning down the right road: Impact of progress testing on licensing examination performance. *Med Teach.* 2010;32:496–499.
- 45 Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach.* 1996;18:103–109.
- 46 Wijnen WHFW. Final objectives tests: Why and how? [in Dutch]. *Onderzoek van Onderwijs.* 1977;6:16–19.
- 47 Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- 48 Verwijnen GM, Van der Vleuten C, Imbos T. A comparison of an innovative medical school with traditional schools: An analysis in the cognitive domain. In: Nooman Z, Schmidt HG, Ezzat E, eds. *Innovation in Medical Education: An Evaluation of Its Present Status.* New York, NY: Springer Publishing; 1990:40–49.
- 49 Schmidt HG, Machiels-Bongaerts M, Hermans H, Ten Cate O, Venekamp R, Boshuizen HPA. The development of diagnostic competence: A comparison between a problem-based, an integrated, and

- a conventional medical curriculum. *Acad Med.* 1996;71:658–664.
- 50 Severiens SE, Schmidt HG. Academic and social integration and study progress in problem based learning. *High Educ.* 2009;58:59–69.
- 51 Tinto V. Classrooms as communities—Exploring the educational character of student persistence. *J High Educ.* 1997;68:599–623.
- 52 Schmidt HG, Rotgans JI, Yew EHJ. The process of problem-based learning: What works and why. *Med Educ.* 2011;45:792–806.
- References Cited Only in Appendixes**
- 53 Albano MG, Cavallo F, Hoogenboom R, et al. An international comparison of knowledge levels of medical students: The Maastricht Progress Test. *Med Educ.* 1996;30:239–245.
- 54 Schuwirth LW, Verhoeven BH, Scherpbier AJ, et al. An inter- and intra-university comparison with short case-based testing. *Adv Health Sci Educ Theory Pract.* 1999;4:233–244.
- 55 Gijbels D, Dochy F, Van den Bossche P, Segers M. Effects of problem-based learning: A meta-analysis from the angle of assessment. *Rev Educ Res.* 2005;75:27–61.



Appendix 1	Summary of 104 Comparisons From Studies Comparing the Performance of Students of Maastricht University's Problem-Based Medical School With That of Students of Conventional Medical Schools in the Netherlands*				
Study and test	Conventional comparison medical school	Sample information by medical school	No. of classes involved and participants' level of training	Measuring instrument information†	Study previously reviewed in‡
<b>Medical knowledge</b>					
Verwijnen et al, 1990 <sup>48</sup> : Test 1	School A	<ul style="list-style-type: none"> <li>PBL school: Population (N = 266)</li> <li>Conventional school: Volunteers (N = 550)</li> </ul>	6 classes: 1st–6th year	64 items randomly stratified, selected from progress test	1, 2, 3, 4, 6, 7, 8
	School B	<ul style="list-style-type: none"> <li>PBL school: Population (N = 266)</li> <li>Conventional school: Volunteers (N = 703)</li> </ul>	6 classes: 1st–6th year	64 items randomly stratified, selected from progress test	1, 2, 3, 4, 6, 7, 8
Verwijnen et al, 1990 <sup>48</sup> : Test 2	School A	<ul style="list-style-type: none"> <li>PBL school: Population (N = 174)</li> <li>Conventional school: Volunteers (N = 196)</li> </ul>	3 classes: 4th–6th year	70 items randomly stratified, selected from progress test	1, 2, 3, 4, 6, 7, 8
	School B	<ul style="list-style-type: none"> <li>PBL school: Population (N = 471)</li> <li>Conventional school: Volunteers (N = 698)</li> </ul>	6 classes: 1st–6th year	70 items randomly stratified, selected from progress test	1, 2, 3, 4, 6, 7, 8
Verwijnen et al, 1990 <sup>48</sup> : Test 3	School A	<ul style="list-style-type: none"> <li>PBL school: Population (N = 189)</li> <li>Conventional school: Volunteers (N = 390)</li> </ul>	3 classes: 4th–6th year	64 items randomly stratified, selected from progress test	1, 2, 3, 4, 6, 7, 8
	School B	<ul style="list-style-type: none"> <li>PBL school: Population (N = 565)</li> <li>Conventional school: Volunteers (N = 677)</li> </ul>	6 classes: 1st–6th year	64 items randomly stratified, selected from progress test	1, 2, 3, 4, 6, 7, 8
	School C	<ul style="list-style-type: none"> <li>PBL school: Population (N = 565)</li> <li>Conventional school: Volunteers (N = 565)</li> <li>Conventional school: Random sample (N = 167)</li> </ul>	6 classes: 1st–6th year	64 items randomly stratified, selected from progress test	1, 2, 3, 4, 6, 7, 8
Verwijnen et al, 1990 <sup>48</sup> : Test 4	School C	<ul style="list-style-type: none"> <li>PBL school: Population (N = 565)</li> <li>Conventional school: Random sample (N = 167)</li> </ul>	6 classes: 1st–6th year	Progress test: 264 items	1, 2, 3, 4, 6, 7, 8
Verhoeven et al, 1998 <sup>26</sup> : Test December 1994	Non-PBL school	<ul style="list-style-type: none"> <li>PBL school: Population (N = 938)</li> <li>Conventional school: Volunteers (N = 728)</li> </ul>	6 classes: 1st–6th year	Progress test: 242 items	6, 7, 8
Verhoeven et al, 1998 <sup>26</sup> : Test March 1995	Non-PBL school	<ul style="list-style-type: none"> <li>PBL school: Population (N = 966)</li> <li>Conventional school: Volunteers (N = 361)</li> </ul>	6 classes: 1st–6th year	Progress test: 226 items	6, 7, 8
Albano et al, 1996 <sup>53</sup>	School Q	<ul style="list-style-type: none"> <li>PBL school: Population (N = 811)</li> <li>Conventional school: Random sample (N = 181)</li> </ul>	3 classes: 1st, 4th, and 6th year	Progress test: 198 items	6, 7, 8
	School R	<ul style="list-style-type: none"> <li>PBL school: Population (N = 811)</li> <li>Conventional school: Random sample (N = 339)</li> </ul>	6 classes: 1st–6th year	Progress test: 198 items	6, 7, 8
	School S	<ul style="list-style-type: none"> <li>PBL school: Population (N = 811)</li> <li>Conventional school: Random sample (N = 386)</li> </ul>	6 classes: 1st–6th year	Progress test: 198 items	6, 7, 8
	School T	<ul style="list-style-type: none"> <li>PBL school: Population (N = 811)</li> <li>Conventional school: Random sample (N = 339)</li> </ul>	5 classes: 1st–3rd year, 5th and 6th year	Progress test: 198 items	6, 7, 8
Van der Vlieten et al, 2004 <sup>41</sup> : Test October 2006 <sup>5</sup>	Nondisclosed school	<ul style="list-style-type: none"> <li>PBL school: Population (N = 1596)</li> <li>Conventional school: Population (N = 1420)</li> </ul>	6 classes: 1st–6th year	Progress test: 187 items	8
Van der Vlieten et al, 2004 <sup>41</sup> : Test December 2006 <sup>5</sup>	Nondisclosed school	<ul style="list-style-type: none"> <li>PBL school: Population (N = 1570)</li> <li>Conventional school: Population (N = 1350)</li> </ul>	6 classes: 1st–6th year	Progress test: 189 items	8

(Appendix continues)

Appendix 1, continued

Study and test	Conventional comparison medical school	Sample information by medical school	No. of classes involved and participants' level of training	Measuring instrument information†	Study previously reviewed in‡
Van der Vleuten et al, 2004 <sup>41</sup> ; Test February 2007 <sup>5</sup>	Nondisclosed school	• PBL school: Population (N = 1413)	6 classes: 1st–6th year	Progress test: 191 items	8
		• Conventional school: Population (N = 911)			
Van der Vleuten et al, 2004 <sup>41</sup> ; Test May 2007 <sup>5</sup>	Nondisclosed school	• PBL school: Population (N = 1267)	6 classes: 1st–6th year	Progress test: 194 items	8
		• Conventional school: Population (N = 952)			
Prince et al, 2003 <sup>25</sup>	School C <sup>1</sup>	• PBL school: Random sample of volunteers** (N = 75)	1 class: 4th year	Anatomy test: 138 items presented in the clinical context of 16 cases	8
		• Conventional school: Random sample of volunteers** (N = 36)			
	School C	• PBL school: Random sample of volunteers** (N = 75)	1 class: 4th year	Anatomy test: 48 items without clinical context	8
		• Conventional school: Random sample of volunteers** (N = 36)			
Prince et al, 2003 <sup>25</sup>	School E <sup>1</sup>	• PBL school: Random sample of volunteers** (N = 75)	1 class: 4th year	Anatomy test: 138 items presented in the clinical context of 16 cases	8
		• Conventional school: Random sample of volunteers** (N = 32)			
	School E	• PBL school: Random sample of volunteers** (N = 75)	1 class: 4th year	Anatomy test: 48 items without clinical context	8
		• Conventional school: Random sample of volunteers** (N = 32)			
Prince et al, 2003 <sup>25</sup>	School F <sup>1</sup>	• PBL school: Random sample of volunteers** (N = 75)	1 class: 4th year	Anatomy test: 138 items presented in the clinical context of 16 cases	8
		• Conventional school: Random sample of volunteers** (N = 33)			
	School F	• PBL school: Random sample of volunteers** (N = 75)	1 class: 4th year	Anatomy test: 48 items without clinical context	8
		• Conventional school: Random sample of volunteers** (N = 33)			
Prince et al, 2003 <sup>25</sup>	School G <sup>1</sup>	• PBL school: Random sample of volunteers** (N = 75)	1 class: 4th year	Anatomy test: 138 items presented in the clinical context of 16 cases	8
		• Conventional school: Random sample of volunteers** (N = 38)			
	School G	• PBL school: Random sample of volunteers** (N = 75)	1 class: 4th year	Anatomy test: 48 items without clinical context	8
		• Conventional school: Random sample of volunteers** (N = 38)			

(Appendix continues)

Appendix 1, continued

Study and test	Conventional comparison medical school	Sample information by medical school	No. of classes involved and participants' level of training	Measuring instrument information <sup>†</sup>	Study previously reviewed in <sup>‡</sup>
Prince et al, 2003 <sup>25</sup>	School H <sup>¶</sup>	<ul style="list-style-type: none"> <li>PBL school: Random sample of volunteers** (N = 75)</li> <li>Conventional school: Random sample of volunteers** (N = 49)</li> </ul>	1 class: 4th year	Anatomy test: 138 items presented in the clinical context of 16 cases	8
	School H	<ul style="list-style-type: none"> <li>PBL school: Random sample of volunteers** (N = 75)</li> <li>Conventional school: Random sample of volunteers** (N = 49)</li> </ul>	1 class: 4th year	Anatomy test: 48 items without clinical context	8
<b>Diagnostic reasoning</b>					
Schmidt et al, 1996 <sup>49</sup>	Amsterdam University	<ul style="list-style-type: none"> <li>PBL school: Volunteers (N = 200)</li> <li>Conventional school: Volunteers (N = 200)</li> </ul>	5 classes: 2nd–6th year	Providing the right diagnosis for 30 clinical cases	5, 6, 7, 8
	Groningen University	<ul style="list-style-type: none"> <li>PBL school: Volunteers (N = 200)</li> <li>Conventional school: Volunteers (N = 200)</li> </ul>	5 classes: 2nd–6th year	Providing the right diagnosis for 30 clinical cases	5, 6, 7, 8
Schuwirth et al, 1999 <sup>54</sup>	Groningen University	<ul style="list-style-type: none"> <li>PBL school: Volunteers (N = 78)</li> <li>Conventional school: Volunteers (N = 177)</li> </ul>	6 classes: 1st–6th year	Case-based computerized test: 60 cases	7, 8

\* PBL indicates problem-based learning; PBL school, Maastricht University's medical school. This table is adapted from Schmidt HG, Van der Molen HT, Te Winkel WWR, Wijnen WHFW. Constructivist, problem-based learning does work: A meta-analysis of curricular comparisons involving a single medical school. *Educ Psychol*. 2009;44:227–249. Copyright © 2009 American Psychological Association. Adapted with permission from Taylor & Francis, Ltd.

† Most often, comparisons used progress tests (see text for description). Without exception, progress tests consisted of true/false choice items.

‡ Reviews that include the particular study: 1. Schmidt et al, 1987<sup>8</sup>; 2. Albanese and Mitchell, 1993<sup>7</sup>; 3. Vernon and Blake, 1993<sup>10</sup>; 4. Berkson, 1993<sup>15</sup>; 5. Colliver, 2000<sup>11</sup>; 6. Doehy et al, 2003<sup>16</sup>; 7. Gijbels et al, 2005<sup>55</sup>; 8. Schmidt et al, 2009.<sup>9</sup>

§ Data were collected using the methodology of Van der Vleuten et al<sup>41</sup> and were made available by the same authors to Schmidt et al for their 2009 review.

¶ Participants from each school in Prince and colleagues<sup>25</sup> study were given two anatomy tests, one without clinical context and one with items embedded in clinical cases. Therefore, each school appears twice in this table.

\*\* Samples were selected at random from pools of volunteers.

Appendix 2

**Reanalysis of the Results of 104 Comparisons From Studies Comparing the Performance of Students of Maastricht University's Problem-Based Medical School With That of Students of Conventional Medical Schools in the Netherlands\***

Study and test	Conventional comparison medical school	Mean attrition rate as %, by curriculum type <sup>†</sup>		Mean study duration in years, by curriculum type <sup>†</sup>		Effect size d in original dataset <sup>‡</sup>	Effect size d after correcting for attrition <sup>‡</sup>	Effect size d after correcting for differences in attrition and study duration <sup>‡</sup>
		PBL	Conv.	PBL	Conv.			
<b>Medical knowledge</b>								
Verwijnen et al, 1990 <sup>48</sup> ; Test 1	School A	15	24	7.33	8.25	0.44	0.63	0.92
	School B	15	23	7.33	7.96	-0.10	0.02	0.17
Verwijnen et al, 1990 <sup>48</sup> ; Test 2	School A	15	24	7.33	8.25	-0.01	0.18	0.71
	School B	15	23	7.33	7.96	-1.02	-0.87	-0.75
Verwijnen et al, 1990 <sup>48</sup> ; Test 3	School A	15	24	7.33	8.25	0.16	0.34	0.66
	School B	15	23	7.33	7.96	-0.06	0.06	0.25
	School C	15	26	7.33	8.09	-0.14	0.06	0.34
Verwijnen et al, 1990 <sup>48</sup> ; Test 4	School C	15	26	7.33	8.09	-0.15	0.05	0.33
Verhoeven et al, 1998 <sup>26</sup> ; Test December 1994	Non-PBL school	7	15	6.95	6.68	0.01	0.15	0.03
Verhoeven et al, 1998 <sup>26</sup> ; Test March 1995	Non-PBL school	7	15	6.95	6.68	-0.31	-0.19	-0.27
Albano et al, 1996 <sup>53</sup>	School Q					0.46		
	School R					0.61		
	School S					0.55		
	School T					0.78		
Van der Vleuten et al, 2004 <sup>41</sup> ; Test October 2006 <sup>††</sup>	Nondisclosed school	10	20	6.84	7.59	0.06	0.26	0.52
Van der Vleuten et al, 2004 <sup>41</sup> ; Test December 2006 <sup>††</sup>	Nondisclosed school	10	20	6.84	7.59	-0.08	0.09	0.34
Van der Vleuten et al, 2004 <sup>41</sup> ; Test February 2007 <sup>††</sup>	Nondisclosed school	10	20	6.84	7.59	0.09	0.27	0.50
Van der Vleuten et al, 2004 <sup>41</sup> ; Test May 2007 <sup>††</sup>	Nondisclosed school	10	20	6.84	7.59	0.18	0.37	0.56
Prince et al, 2003 <sup>25**</sup>	School C <sup>††</sup> (clinical)	14	24			0.65	0.65	
	School C (nonclinical)	14	24			-0.16	-0.16	
Prince et al, 2003 <sup>25**</sup>	School E <sup>††</sup> (clinical)	14	24			0.60	0.95	
	School E (nonclinical)	14	24			0.05	0.33	
Prince et al, 2003 <sup>25**</sup>	School F <sup>††</sup> (clinical)	14	21			0.23	0.42	
	School F (nonclinical)	14	21			0.01	0.19	

(Appendix continues)



Appendix 2, continued

Study and test	Conventional comparison medical school	Mean attrition rate as %, by curriculum type <sup>†</sup>		Mean study duration in years, by curriculum type <sup>†</sup>		Effect size d in original dataset <sup>‡</sup>	Effect size d after correcting for attrition <sup>‡</sup>	Effect size d after correcting for differences in attrition and study duration <sup>‡</sup>
		PBL	Conv.	PBL	Conv.			
Prince et al, 2003 <sup>25**</sup>	School G <sup>††</sup> (clinical)	14	19			-0.55	-0.46	
	School G (nonclinical)	14	19			-0.03	0.09	
Prince et al, 2003 <sup>25**</sup>	School H <sup>††</sup> (clinical)	14	17			0.51	0.60	
	School H (nonclinical)	14	17			0.21	0.28	
<b>Diagnostic reasoning</b>								
Schmidt et al, 1996 <sup>49</sup>	Amsterdam University	7	18	6.97	7.28	-0.36	-0.15	0.09
	Groningen University	7	16	6.97	7.52	0.20	0.46	0.85
Schuwirth et al, 1999 <sup>54</sup>	Groningen University	9	14	6.91	7.19	0.38	0.51	0.58

\* See Table 1 for descriptions of these comparisons.

<sup>†</sup> PBL indicates problem-based learning curriculum; conv., conventional curriculum. In all comparisons, the PBL curriculum is that of Maastricht University's medical school. Italic type indicates attrition and study duration averaged over classes of students involved.

<sup>‡</sup> Effect size is expressed as Cohen d. Italic type indicates an average d, weighted over classes of students and numbers of students. Effect sizes for the original dataset are from Schmidt et al, 2009.<sup>9</sup>

<sup>§</sup> This study did not allow for correction because attrition rates and study durations of participating schools were not known.

<sup>¶</sup> Data were collected using the methodology of Van der Vleuten et al<sup>41</sup> and were made available by the same authors to Schmidt et al for their 2009 review.

\*\* This study allowed only for the estimation of d values after correction for attrition.

<sup>††</sup> Participants from each school in Prince and colleagues<sup>25</sup> study were given two anatomy tests, one with items embedded in clinical cases (clinical) and one without clinical context (nonclinical). Therefore, each school appears twice in this table.