Effectiveness of problem-based learning curricula: theory, practice and paper darts

Geoffrey R Norman[1] *& Henk G Schmidt*[2]

Summary In a recent review article, Colliver concluded that there was no convincing evidence that problem-based learning was more effective than conventional methods. He then went on to lay part of the blame on cognitive psychology, claiming that 'the theory is weak, its theoretical concepts are imprecise... the basic research is contrived and *ad hoc*'. This paper challenges these claims and presents evidence that (a) cognitive research is not contrived and irrelevant, (b) curriculum level interventions are doomed to fail and (c) education needs more theory-based research.

Keywords Education, medical, graduate, *methods; evaluation studies; *knowledge; *models, educational; problem-based learning, *standards.

Medical Education 2000;34:721–728

Introduction

In a recent review article, Colliver[1] examined the literature on the effectiveness of problem-based learning published since the last two comprehensive reviews in 1993.[2,3] The evidence is substantially consistent with the earlier reviews. On measures of knowledge, such as national licensing examinations, PBL students perform a little better or a little worse than students in conventional curricula. On measures of clinical reasoning or diagnostic ability, there is a small but significant effect of PBL and on measures of satisfaction, there is a consistent benefit of PBL. These findings lend credence to the conclusion that PBL is unlikely to make students learn more in the short term (although they may retain more[4]). However, PBL may show small and consistent short-term gains in clinical skills, and does show consistent gains in satisfaction.

Colliver's[1] interpretation of these findings is somewhat more pessimistic:

> The review of the literature revealed no convincing evidence that PBL improves knowledge base and clinical performance, at least not of the magnitude that would be expected given the resources required… (p. 259)

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
[2]Department of Psychology, Maastricht University, PO Box 616, 6200 MD Maastricht, Netherlands

Correspondence: Professor Geoffrey R Norman, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada

In a companion paper in this issue, Albanese[5] discusses the interpretation of these effects in detail, and challenges Colliver's interpretation. We concur with his challenge, and believe that it amounts to ignoring much of the data that he, himself, reviews.

What's the problem?

Despite the reservations noted above, we are not in substantial disagreement with Colliver about the evidence. While we may disagree as to whether an effect size is moderate or medium, and should be attended to or discounted, we do agree that PBL does not result in dramatic differences in cognitive outcomes. We believe that PBL *has* been oversold by its advocates, promising enormous benefits and largely ignoring the associated resource costs. We must take blame for some of the excesses in our earlier writings, and can only acknowledge that with maturity comes a degree of moderation.

But we strenuously disagree with Colliver about the cause of the problem. His case is stated succinctly:

> If PBL is based on educational principles and learning mechanisms that are supported by basic research, why isn't the PBL curriculum more effective with respect to knowledge and clinical skills? The problem is, as I see it, that the theory is weak; its theoretical concepts are imprecise, lacking explicit descriptions of their interrelationships and of their relationships with observables, such as interventions and outcomes. In addition, the basic

© Blackwell Science Ltd

> research is contrived and *ad hoc*, using manipulations that seem to ensure the expected results regardless of the theory – which is too indefinite to place any real constraints on the observables anyway. In brief, the ties between theory and research are loose at best.[1] (p. 264)

Colliver's solution to this perceived problem is:

> My recommendation is that we reconsider the value of thinking in terms of this imprecise theory about underlying mechanisms and of pursuing basic research that attempts to test its indefinite predictions.[1] (p. 266)

The ties between theory and research (by which we presume Colliver means the kind of programme evaluation he cites, not basic research) are loose indeed. But the problem lies with the programme evaluators, not the theoreticians. We believe strongly that the evidence as presented is completely expected in view of the poor understanding of learning exemplified by these curriculum level interventions, and that real progress will result from *more*, not less, theory-based research

We begin by illustrating with a parable from physics:

> Newton's law describing the action of gravity on objects with known mass was discovered about 300 years ago. It has stood the test of time, and was only superceded by Einstein's general relativity in the first half of the 20th century. Although it is 'wrong' it is a perfectly adequate approximation under most circumstances. Indeed, Newton's laws suffice to place communication satellites within metres in stationary earth orbit and to guide interplanetary probes to the limit of the solar system. While the calculations are complex, since they must consider the gravitational attraction of the sun, the earth, the moon and all other planets, they are easily performed by modern computers.
>
> However, Newton's Laws have their limits. While they can guide a spacecraft beyond Pluto, they cannot predict the path of a paper dart from hand to ground with any precision at all. The dart is subject to myriad aerodynamic forces, from winds to passing birds, interacting in complex ways with its own form, so that precise prediction is impossible.

A hundred years ago, a casual observer of Lilienthal or the Wrights, as they tried to understand the performance of their model airplanes, might well admonish them to abandon any further theory development, since the theory is inadequate to explain the flight paths. But it's not that the theory is bad. Quite the converse, it's remarkably good (and its successor, relativity theory, would be much worse, since the calculations would be far more complex). It's just that there are many more interacting forces than just gravity.

Fortunately, engineers and physicists did not abandon their efforts to model aerodynamic forces on surfaces, still using the basic theory of Newtonian mechanics, but incorporating more and more variables, whose relationships were investigated and understood using simplified shapes in laboratories and wind tunnels. And their efforts led ultimately to the spacecraft, whose flight path can be easily modeled by Newton's laws in their original form.

Similarly, we will argue that the small effects and inconclusive findings derived from research on PBL result, not from the inadequacy of the theory and its basis in the laboratory, but from the futility of conducting research on interventions which, like PBL, are inadequately grounded in theory, in real environments, which are so complex and multifactorial, with so many unseen interacting forces, using outcomes so distant from the learning setting, that any predicted effects would inevitably be diffused by myriad unexplained variables. Indeed, we believe that the fact that *any* significant effects have been observed is evidence of the effectiveness of PBL, which succeeds to show differences in a situation where the cards are stacked high against it. In elaborating these arguments, we make, and support, three claims:

1. Basic cognitive research is *not* contrived and irrelevant:
   a. The interventions are precise.
   b. The outcomes are highly relevant.
   c. The studies do not amount to self-fulfilling prophesies.
   d. The effects are non-trivial.
2. Curriculum level interventions, using simple experimental designs such as RCTs and limiting the manipulation to one variable, are doomed to fail.
3. Education needs more, not less, theory-based research conducted in relatively controlled settings, if real advances in educational practice are to result.

Cognitive research is not contrived and irrelevant

In making the point that cognitive research is irrelevant, Colliver discusses two phenomena – context dependence and activation. Context dependence is the demonstration that recall is better when the retrieval setting matches the setting of original learning. Colliver discusses in detail a study by Godden and Baddeley[6] where members of a university diving club memorized lists of

words on land and underwater. The words learned on land were better recalled on land and *vice versa*.

We completely agree that the phenomenon of context dependence is of only marginal educational relevance, since the match between learning and retrieval setting is rarely controllable. Further, since context includes both relevant variables such as the clinical problem, and irrelevant, such as the room in which the tutorial occurs, PBL does, at best, an incomplete job at matching context. We argued precisely this point in our 1992 paper.[4]

But does the Godden and Baddeley[6] paper simply 'ensure the expected results'? Hardly. We are accustomed to thinking of the mind as a biological computer, but this study clearly challenges the metaphor. The computer's recall from memory is unaffected by the room it is in and, yet, human recall clearly is. We expect that Godden and Baddeley's paper is so widely cited precisely because its results were not anticipated. It is *because of* Godden and Baddeley that the effect of context seems so commonsense, not the reverse.

Colliver assaults activation as a (partial) explanation of what works in PBL and why, while ignoring the studies that gave rise to the introduction of this concept in medical education in the first place. These studies demonstrate that discussing a problem in the small-group setting strongly activates prior knowledge of participants;[7] and that this knowledge is not simply a bag of facts that students have available but can be described as a 'naive theory' or 'mental model' that these students entertain with regard to the problem at hand.[8] This naive theory of the problem, once activated through discussion, in turn facilitates the processing of problem-relevant new information[8] (see for a recent overview[9]). A recent illustration of the effects of prior knowledge activation in learning new information is a study by De Grave,[10] who used a problem of blood pressure regulation that is actually used in a problem-based medical curriculum with students of that particular curriculum. His experimental participants greatly outperformed his controls: they produced 25% more accurate propositions on a knowledge test of blood pressure regulation. The effect size measures in all these studies were well above the level that Colliver would describe as moderate; the average effect size was 1.30.

Colliver concludes from review of these two domains that 'educational theory and its basic research… seems to be nothing more than metaphor and demonstration, nothing that taps into the underlying substrate of learning that would allow prediction and control.' Perhaps, if context dependence was all that theory had to offer. But permit us to describe two additional counter-examples from the first author's work.

(a) Forward vs. backward reasoning

Studies of expertise both within medicine[11] and in other domains[12] have identified that forward reasoning, from data to solution, is a hallmark of expertise. That is, novices tend to reason about clinical cases using reasoning such as 'Well, it might be a heart attack because the patient has chest pain', reasoning backward from the diagnosis of heart attack to the manifestation, chest pain. Expert reasoning more commonly is of the form 'The patient has chest pain with radiation and diaphoresis, so the likely diagnosis is heart attack' – forward from symptoms to diagnoses.

This finding is accepted as commonplace. It is also consistent with the admonition to medical students to 'gather all the data', 'be systematic', 'don't do pattern recognition' and 'avoid premature closure' – that is, to avoid having tentative diagnoses guide search. Indeed, Colliver mentions it as a failing of a study at Rush University[13] that, despite showing a clear superiority in diagnostic accuracy ($d = +0{\cdot}80$), the PBL students did backward reasoning. (The former finding is dismissed as unimportant because the PBL students were just 'doing what they had been doing' with similar problems.)

However, the finding is inconsistent with evidence in psychology of perception where letter recognition is influenced by memory at the word level (top-down processing[14]). In pondering this phenomenon, we noticed that all the studies of reasoning direction depend on verbal reports, thus may confound expertise and confidence with reasoning. We therefore conducted a study in a lab setting, where undergraduate psychology students were taught ECG diagnosis.[15] All were taught in exactly the same way. At the test, half the students were told to 'Carefully search the ECG for abnormal features, list them and then arrive at a diagnosis' (i.e. forward reasoning); the other half were told to come to a diagnosis as fast as they could then search for evidence to support the diagnosis (backward reasoning). The overall accuracy of the forward group was 42%; of the backward group was 59%. The forward group was gathering too much data and was left with the odious task of explaining it, which led them astray. We venture that this cannot be dismissed as a study that is 'contrived and ad hoc, using manipulations that seem to ensure the expected results', since conventional wisdom would have predicted the opposite result.

(b) Mixed/blocked practice

Traditional educational approaches tend to teach concepts in identifiable blocks. The most common example of this is the textbook. In statistics, books are organized around particular statistical tests, so that there is a chapter on 'The *t*-test', another on 'One Way ANOVA', and so on. At the end of each chapter is a series of exercises, where the student must apply the learned rules to calculate the test. Typically, the question is posed as 'Perform a *t*-test on the following data'. The student would almost never encounter a *t*-test in the ANOVA chapter or a Mann Whitney U in the *t*-test chapter.

Clearly, implicit in this approach is the view that learning amounts to assimilation of rules (equations) which must be practised in order to be learned and, once learned, can be applied appropriately in other situations. Of course, that is not what happens. Graduates of statistics courses report that they learn the equations, pass the exams, but never do discover the circumstances under which they (or more likely, their computers) should apply a particular test. When the data are ambiguous, and the rules for classification are not explicit, simple rehearsal is not sufficient. Instead, students should encounter problems from diverse categories, and practice learning the features which discriminate an example of one class from another. That is, conventionally, students learn in *blocked practice* with multiple examples from a single category. The contrast is *mixed practice* where students are required to identify examples from diverse categories.

Recently, Hatala investigated these two practice modes in a study of teaching ECG diagnosis to first year students in a cardiology unit. All had the same basic instruction with the rule and two prototypical examples from each category. The 'blocked' group saw additional examples from each category, while the 'mixed' group practised with the same examples in random order. Finally, all had the same test with six new examples. The overall accuracy of the 'blocked' group was 31% and that of the mixed group was 48%.

We suggest that these results are non-trivial – a gain in diagnostic accuracy of 17% is highly relevant. The intervention is precise and could easily be adopted in many circumstances. Are the results a self-fulfilling prophesy? One might argue so, although it seems that the self-evident nature of the manipulation has escaped generations of academic authors.

Neither these two examples, nor the two chosen by Colliver, adequately represent the diversity of findings from cognitive psychology which might have potential application to medical education. But we believe they do serve as evidence that not all basic research is trivial, self-evident and irrelevant to educational practice.

Curriculum-level interventions are doomed to fail

We continue to be astonished that researchers can attract funding for large, multicentre, multimegabuck trials of educational interventions. We are even more astonished, indeed distressed, to hear that the Campbell Collaboration, the group which is on a research synthesis mission, plans to restrict their purview to large randomized trials of curriculum level interventions. Many decades of studies in this genre have yielded a consistent result – no difference. Many years ago, one of us came across a small book called *The Teaching–Learning Paradox*,[16] which reviewed educational trials from 1900 to 1960. The final chapter was elegantly titled with the Shakespearean epithet 'Full of Sound and Fury – Signifying Nothing'. And nothing has changed since then, as Colliver's review ably attests.

Although clinical trials of drugs have their own problems,[17–19] primarily related to the extensive inclusion and exclusion criteria necessary to identify a homogeneous and responsive sample, their problems pale in comparison to educational trials. We believe that there are three main reasons why educational trials are ill-founded and ill-advised:

There is no such thing as a blinded intervention

Colliver shows a certain preoccupation with randomization as a *sine qua non* of good research. But in doing so, he forgets why random allocation is so important. The idea is to ensure that both groups are equivalent except for the operation of chance, so that any differences in the outcome can be attributed to the intervention. Unfortunately, in educational interventions, it is impossible to maintain blinding, either among students or teachers, so that it is impossible to attribute success or failure solely to the intervention. Teachers in the study group may be more enthusiastic, or less; students may be excited at the new approach, or distressed; teachers or students in the control group may try harder. So any difference or similarity, is subject to multiple interpretations.

There is no such thing as a pure outcome

Students in medical school are a highly selected and atypical group. They have demonstrated the skill to succeed in academic settings. They have survived the rigour of a highly selective admission process, and are now on the lowest rung of a ladder to social status and success. Thus, faced with a test such as a national

licensing examination, the majority have all the prerequisite skills, regardless of the curriculum they are in, to pass this hurdle (of course not all pass, but this is hardly something to blame the curriculum for). Under these circumstances it is hardly surprising that curriculum level comparisons show small effects on high stakes outcome measures.

There is no such thing as a uniform intervention

As anyone who has visited more than one PBL school will attest, the little acronym covers a multitude of sins. PBL is practised very differently in different institutions.[20] Indeed, it would be surprising if it were otherwise. But these different implementations reveal a deeper diversity. Since PBL has a number of characteristics, it can be implemented in very different ways in different places – PBL is individualized, cooperative, small group, with non-expert tutors, self-paced, using problems to list but a few.

As an exercise, we tried to match some of these terms with the equivalent term in a synthesis of 302 meta-analyses of educational and psychological interventions, involving over 10 000 studies, conducted by Lipsey and Wilson.[21] We looked for descriptions of treatments that roughly matched some of the characteristics of PBL. The results are shown in Table 1.

It is evident that the interventions associated with PBL are multiple and each has a demonstrable effect – some positive and some negative. While it may be tempting to average these effects, this would be folly indeed, first, because many of the effects are derived from very different populations and content, typically public school science instruction. Second, even if the population and content were similar, it cannot be assumed that the results are additive, or that the effects should be equally weighted. Far more likely is the possibility that there are complex interactions among many of the treatment components, so that any estimate of effectiveness must account for these interactions. The presence of these multiple components in any curriculum – level intervention like PBL will invariably confound attempts to seek cause–effect relationships, and simple experimental strategies like randomization will hardly remedy the situation.

Table 1 Effect sizes associated with various aspects of problem-based learning

Characteristic	Effect size
(a) Individualized	0·23
(b) Cooperative	0·54
(c) Small group	0·31
(d) With non-expert tutors	−0·74
(e) Self-paced	−0·07
(f) Self-directed	−0·05
(g) Using problems	0·20
(h) Inquiry based	0·16
(i) Instruction in problem solving	0·54
(j) Inductive	0·06

Note: all ES except (d) derived from Lipsey and Wilson.[21] (d) is based on analysis of Eagle and Harasym[27] and Davis *et al.*[28]

Education needs more, not less, theory-based research

This brings us to our final point. Trials of curriculum level interventions, whether they show large, small, zero or negative effect sizes, whether they are or are not randomized, whether they are done on large or small samples, whether they are of sound or unsound methodology, are, in our view, a waste of time and resources. They are a waste because, by their very nature, they are doomed to examining one variable, or more likely an unspecified combination of many variables, at a time. We may, if we are fortunate, conclude that a particular intervention, administered to a particular sample, to teach a particular content, using a particular outcome, was successful. But we will have no way of determining the generalizability of the findings to a different situation. Of course, far more likely is the conclusion that Colliver has reached, that practical interventions rarely have sizeable effects, simply because the many intervening variables mitigate against observing sizeable effects.

This is precisely the major criticism levelled at randomized controlled trials of drugs,[17–19] which commonly address the question of whether a drug works or not. But relatively few patients benefit in any particular trial, and the nature of the trial design is such that it can provide no answer to the question of who might benefit, or whether a particular patient is likely, in the future, to benefit from treatment. In the not uncommon situation that the trial is unsuccessful, it provides no insight as to why the intervention failed. The problem is that the act of randomization has eliminated from the object of inquiry precisely those variables which might provide insight into these questions. As Donald Campbell once said in an address to the American Educational Research Association:

> When a researcher says that such and such an effect is true, all other things being equal, he speaks from the experience of setting a great many other things equal.

As we have indicated, the problems with educational interventions are more serious since there is much less control in the educational setting. Does this mean, in turn, that all efforts at educational research are fruitless? Not at all. But we must take a cue from the natural sciences and move away from blind allegiance to the canons of sound methodology (randomization and all that) to recognition and support for research programmes, whose intent is to create an environment where ideas are shepherded from the basic science laboratory to the application setting.

The basic science of learning has a central role in such programmes; only through research which is conducted with careful and systematic exploration of multiple variables, directed at testing and elaboration of theories, can we ever expect to understand enough to achieve practical goals. As understanding evolves, we can then safely proceed to the messier environment of the application setting. However, instead of seeking to control the application setting so that we can manipulate variables one at a time, we must seek to capture and measure precisely those variables that the hard-core experimentalist seeks to randomize away. The advantage of a real environment is not that it is so messy with extraneous variables that we must randomize their influence away, but that it is so rich with other variables that we must capture these effects to truly understand the complexity of learning interactions. In doing so, we must use sophisticated tools, such as structural equations modelling, to capture the main effects and interactions of these variables.[22]

We have already provided examples from our own work of research conducted in the highly controlled situations directed at basic theory building. We ask the reader's indulgence as we present some of our own work which seeks to understand the complexities of the application setting through examining the interplay of multiple variables with structural equations modelling. Structural equations modelling is a multivariate statistical technique that enables investigators to test substantive and complex causal theories against correlational data, and to determine the extent to which such theories would 'fit' the empirical data.

The second author's research group has conducted a series of studies using structural equation modelling to clarify the complex relations among variables thought to be important in PBL.[23–25] One such study was aimed at testing a theoretical model of PBL against data collected over six consecutive academic years in the problem-based medical curriculum of Maastricht University.[26] Participants were approximately 1350 undergraduate medical students who enrolled in a total

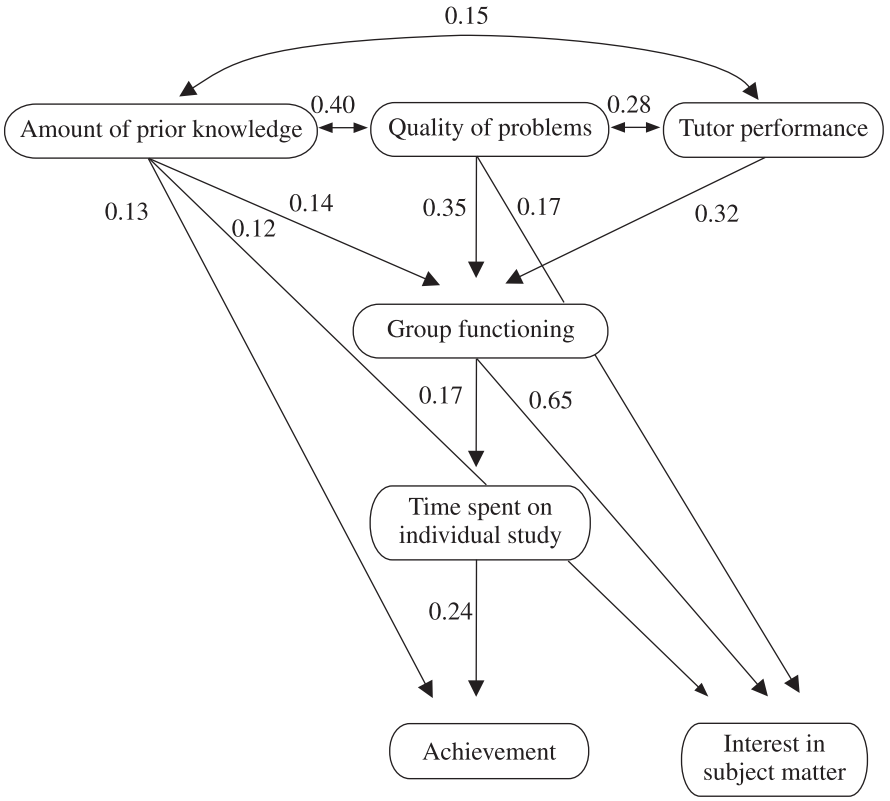

Figure 1 Path model of PBL, using data from six subsequent academic years of the Maastricht problem-based medical curriculum[25]

of 120 courses over that period. In each of these courses, data were collected regarding prior knowledge of participants, the quality of the problems presented, tutor functioning, tutorial-group functioning, self-study time, interest in subject-matter and academic achievement. These data were submitted to structural equations modelling and tested against the theory. Figure 1 summarizes the main findings. The values accompanying the arrows are path coefficients indicating the magnitude of the influence that variables exert on other variables in the model.

Findings such as these suggest both theoretical and practical avenues for further enquiry. Theoretically, they may deepen our understanding of the relative contributions of factors involved in PBL. For instance, the findings suggest a central role of high-quality problems in PBL. The question then, is what is a high-quality problem? Practically, they have already led to clarification of whether tutors should be subject-matter experts or process facilitators, an issue hotly debated in medical education.[24]

Conclusions

We completely concur with Colliver that we should rethink the promise of PBL for the acquisition of basic knowledge and clinical skills. The evidence that he and others have presented is undeniable. We can safely assume that any study that treats PBL as a single 'intervention' and examines the usual cognitive and clinical outcomes will arrive at a conclusion of minimal difference. But we differ from Colliver in two fundamental ways:

First, PBL *does* provide a more challenging, motivating and enjoyable approach to education. That may be a sufficient *raison d'être*, providing the cost of implementation is not too great. Clearly, we can use more information about the relative costs of each curriculum approach.

Second, while we agree that we should rethink the promise of PBL, our ways of rethinking will differ significantly from Colliver's. We believe that the field will advance only by a systematic research programme which encompasses all aspects from theory building and testing conducted with rigorous experimental designs in highly controlled and artificial settings to programme evaluations in realistic settings with a deliberate attempt to capture all possible variables and interactions rather than randomize them away. Theory development should be viewed as an essential and central component in the quest for prediction and control, not as a diversion from the 'real' goal.

References

1 Colliver J. Effectiveness of problem based learning curricula. *Acad Med* 2000;**75**:259–66.
2 Albanese MA, Mitchell S. Problem-based learning: a review of literature on its outcomes and implementation issues. *Acad Med* 1993;**68**:52–81.
3 Vernon DTA, Blake RL. Does problem-based learning work? A meta-analysis of evaluative research. *Acad Med* 1993;**68**:550–63.
4 Norman GR, Schmidt HG. The psychological basis of problem-based learning: a review of evidence. *Acad Med* 1992;**67**:557–65.
5 Albanese MA. Response to Colliver. *Med Educ* 2000;**34**:???.
6 Godden DR, Baddeley AD. Context-dependent memory in two natural environments: on land and underwater. *Br J Psychol* 1975;**66**:325–31.
7 Schmidt HG. Activatie van voorkennis en tekstverwerking (Activation of prior knowledge and text processing). *Nederlands Tijdschrift Voor Psychologie* 1984;**39**:335–47.
8 Schmidt HG, De Volder ML, De Grave WS, Moust JHC, Patel VL. Explanatory models in the processing of science text: The role of prior knowledge activation through small-group discussion. *J Educ psychol* 1989;**81**:610–9.
9 Schmidt HG, Moust JHC. Processes that shape small-group tutorial learning: A review of research. In: DH Evensen, CE Hmelo, eds. *Problem-Based Learning: a Research Perspective on Learning Interactions.* Mahwah, NJ: Lawrence Erlbaum; pp 19–52, 2000.
10 De Grave WS, Schmidt HG, Boshuizen HPA. Problem based learning: Effects of problem analysis in a small group on learning a text in first-year medical students. *Instructional Sci* 2000;in press.
11 Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. *Cogn Sci* 1986;**10**:91–116.
12 Larkin JH, McDermott J, Simon DP, Simon HA. Models of competence in solving physics problems. *Cogn Sci* 1980;**4**:317–45.
13 Hmelo CE. Cognitive consequences of problem-based learning for the early development of medical expertise. *Teaching Learning Med* 1998;**10**:92–100.
14 Reicher GM. Perceptual recognition as a function of meaningfulness of stimulus materials. *J Exp Psychol* 1969;**81**:274–80.
15 Norman GR, Brooks LR, Colle CK, Hatala RM. The benefit of diagnostic hypotheses in clinical reasoning: an experimental study of an instructional intervention for forward and backward reasoning. *Cognit Inst* 1999;**17**:433–48.
16 Dubin R, Taveggia TC. *The Teaching–Learning Paradox. A Comparative Analysis of college teaching methods.* University of Oregon, USA: Center for the Advanced Study of Educational Administration; 1968.
17 Tanenbaum SJ. Evidence and expertise: The challenge of the outcomes movement to medical professionalism. *Acad Med* 1999;**74**:757–63.

18 Tonnelli M. The philosophical limits of evidence based medicine. *Acad Med* 1998;**73**:1234–40.
19 Feinstein AR, Horwitz R. Problems in the 'evidence' of 'evidence-based medicine'. *Am J Med* 1997;**103**:529–35.
20 Maudsley G. Do we all mean the same thing by 'Problem-based Learning'? A review of the concepts and a formulation of the ground rules. *Acad Med* 1999;**74**:178–85.
21 Lipsey MW, Wilson DB. The efficacy of psychological, educational and behavioral treatment: confirmation from meta-analysis. *Am Psychol* 1993;**12**:1181–209.
22 Bentler PMEQS. Structural Equations Program Manual. Los Angeles, CA: BMDP Statistical Software Inc.; 1989.
23 Schmidt HG, Gijselaers WH. *Causal modelling of problem-based learning*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA, 1990.
24 Schmidt HG, Moust JHC. What makes a tutor effective? A structural-equations modeling approach to learning in Problem-based curricula. *Acad Med* 1995;**70**:708–14.
25 Van Berkel H, Schmidt HG. Motivation to commit oneself as a determinant of achievement in problem-based learning. *Higher Educ* 2000;in press.
26 Schmidt HG. *Testing a Causal Model of Problem-based Learning*. Paper presented at the Annual Meeting of the American Educational Research Association, Montréal, Canada, April 19–23, 1999.
27 Eagle CJ, Harasym PH, Mandin H. Effects of tutors with case expertise on problem-based learning issues. *Acad Med* 1992;**67**:465–9.
28 Davis WK, Nairn R, Paine ME, Anderson RM, Oh MS. Effects of expert and non-expert facilitators on the small-group process and on student performance. *Acad Med* 1992;**67**:470–4.

Received 5 *June* 2000; *accepted for publication* 5 *June* 2000