# <u>Effectiveness of Problem-based learning</u>

**draft paper October 8, 2008**

**David Gijbels, University of Antwerp, Belgium**

**In cooperation with:**
**Piet Van den Bossche, University of Maastricht, The Netherlands**
**Mien Segers, University of Leiden, The Netherlands**
**Filip Dochy, University of Leuven, Belgium**

**More than ever, successful functioning in society demands more than the understanding of the basic knowledge of a students' domain of study. An important challenge for today's higher education remains the development and implementation of learning and teaching practices that will foster in students the skill to acquire and apply their knowledge efficiently, think critically, analyse, synthesise and make inferences (Segers, Dochy & Cascallar, 2003). Overall, it is claimed that 'student-centred' or 'new' learning environments have the potential to improve these educational outcomes for students in higher education (Lea, Stephenson & Troy, 2003; Simons, van der Linden & Duffy, 2000). New learning environments are rooted in constructivist theory and intend to develop an educational setting to meet the challenge for today's higher education, making the students' learning the core issue and defining instruction as enhancing the learning process.**

**A well known example from higher education of such a new learning environment is problem-based learning (Birenbaum, 2003; Hendry, Frommer & Walker, 1999; Russel, Creedy & Davis, 1994; Savery & Dufy, 1995; Segers, Dochy & De Corte, 1999). Although originally developed for medical training in Canada at McMaster university, the orthodox version of problem-based learning (PBL) has been modified and applied globally in many disciplines (Gijselaers, 1995). Many curricula or parts thereof are modelled on the basis of problem-based learning. The desire to implement problem-based learning as an alternative for existing teaching practices inevitably raises the question as to whether problem-based learning is an alternative capable of effectively replacing conventional instruction (Albanese & Mitchell, 1993). Although contemporary educational practice increasingly consists of a blend of conventional and student-centred methods, there seems to be a need to compare the merits of the two approaches.**

**The core issue in this paper is: "To what extent is PBL an effective learning environment?" The question itself is indicative of a clear description of educational interventions that can be labelled 'PBL' (Newman, et al., 2004). In spite of the many variations of PBL that aim to match PBL with the specific educational or discipline context, for comparative research, a core model or basic definition is needed to which other educational methods can be compared. The six core characteristics of PBL as distinguished in Barrows' (1996) core model are used as a frame of reference in this dissertation. They can be described as follows. The first characteristic is that learning needs to be student-centred. Secondly, learning has to occur in small student groups. The third characteristic refers to the presence of a tutor as a facilitator or guide. Fourthly, authentic problems are primarily encountered in the learning sequence, before**

**any preparation or study has occurred. Fifthly, the problems encountered are used as a tool to achieve the required knowledge and the problem-solving skills necessary to eventually solve the problem. Finally, new information needs to be acquired through self-directed learning. It should be noted that just as the definition of PBL is ambiguous, the definition of what constitutes conventional instruction is also ambiguous. For the most part, conventional instruction is marked by large group lectures and instructor-provided learning objectives and assignments (Albanese & Mitchell, 1993).**

**If one ponders the implementation of PBL, a major question is: Do students from PBL reach the goals in a more effective way than students who receive conventional instruction?**

**Albanese and Mitchell (1993, p.56) pose this question as follows: "Stated bluntly, if problem-based learning is simply another route to achieving the same product, why bother with the expense and effort of undertaking a painful curriculum revision?"**

**The main scope in this paper is focussing on two cognitive learning outcomes: the acquisition as well as the application of knowledge. The main aim of problem-based learning environments in higher education is to guide students to become experts in a certain field of study, so-called professionals: Graduates who can identify the problems of different disciplines and who are capable of analysing and contributing to the solutions of these problems. The findings of cognitive psychological research, especially results from expert versus novice studies have contributed to the insights in the nature of expertise (Feltovich, Spiro & Coulson, 1993; Gagné, Yekovich & Yekovich, 1993; Glaser, 1990). These findings have provided a basis for unravelling the general goal of PBL, the development of successful problem-solving into two dimensions i.e. the acquisition as well as the application of knowledge.**

**PBL environments claim to have the potential to meet the challenges of today's educational needs. However, the question is whether or not these claims are justified. Are the potentials reached? And is there room for improvement? This paper will start with reviewing the evidence concerning the effects of problem-based learning. Recent reviews report some optimistic results but also point to challenges yet to overcome in order to meet the high expectations of such problem-based learning environments. Thereupon, this paper will give an overview of exemplary research in our team that has been done focussed on the opportunities for improvement.**

## Have the expectations been met?

**Two sets of research questions guided our first meta-analyses on the effects of problem-based learning (Dochy, Segers, Van den Bossche & Gijbels, 2003) First, we addressed the main effects of PBL on two broad categories of outcomes: knowledge and skills (i.e., application of knowledge). Secondly, potential moderators of the effect of PBL are addressed. A first category of moderators are design aspects of the reviewed research. In the second category of moderators, we examined whether the effect of PBL differs according to various levels of student expertise. Thirdly, we looked more closely at different types of assessment methods. Fourthly, we investigated the influence of the insertion of a retention period.**

**Before searching the literature for work pertaining to the effects of PBL, we determined the criteria for inclusion in our analysis. First, each study had to be**

**empirical, meaning that some data collection on students had to be included. Although more non-empirical literature and literature reviews were selected as sources of relevant research, this literature was not included in the analysis. Second, the characteristics of the problem-based learning environment, had to fit the previously described core model of PBL (Barrows, 1996). Third, each study had to include some course or curriculum comparison. Specifically, it had to compare students in a PBL environment with students in a more conventional educational setting. The dependent variables used in the studies had to be operationalized aspects of the main goals of PBL (i.e., knowledge acquisition and knowledge application). Fourth, the subjects of study had to be students in higher education (including college and university students in all possible domains of interest). Finally, to maximize ecological validity, each study had to be conducted in a real-life classroom or programmatic setting rather than under more controlled laboratory conditions.**

### ***Literature Search***

**The review and integration of research literature begins with the identification of the literature. Locating studies is the stage at which the most serious form of bias enters a meta-analysis (Glass, McGaw & Smith, 1981). As Glass (1976, p. 6) stated: "How one's search determines what one finds; and what one finds is the basis of the conclusions of one's integration". The best protection against this source of bias is a thorough description of the procedure used to locate the studies.**

**A literature search was started in 1997 which included both published and unpublished studies. A wide variety of computerized databases were utilized including: the Educational Resources Information Center (ERIC) catalogue, PsycLIT, ADION, and LIBIS as well as the Current Contents (for Social Sciences). The following keywords were used: problem-solving, learning, problem-based learning, higher education, college(s), research, and review. The literature was selected based on the abstracts. This reading resulted in the selection of 14 publications that met the above criteria. Next, we employed the 'snowball method' and reviewed the references in the selected articles for additional works. Review articles and theoretical overviews were also gathered to check their references. This method yielded 17 new studies.**

**A second literature search that began in 1999 followed the same procedure. In addition, we contacted several researchers active in the field of PBL and asked them to provide relevant studies or to identify additional sources of studies. This second search yielded 9 additional studies.**

**Although our search for literature was not limited to one single domain of interest, almost all studies meeting the criteria for inclusion were conducted in the domain of medical education. Only one study (Son & VanSickle, 2000) was situated outside the medical domain, in the field of economics. The strategies used to search for literature were meant to uncover both published and unpublished studies to prevent for publication bias. A great deal of papers was traced, but further reading revealed that eventually all papers had been published as an article in a peer-reviewed journal (e.g., Schmidt et al., 1996) or as chapter in an edited book (e.g., Boshuizen, Schmidt & Wassamer 1990).**

**From the studies that met the criteria for inclusion, 33 studies (76.7%) presented data on knowledge effects and 25 studies (58.1%) reported data on**

effects concerning the application of knowledge. These percentages add up to more than 100 since several studies presented outcomes of more than one category.

### *Main effects of PBL*

The main effect of PBL on knowledge and skills is differentiated. The results of the analysis are summarised in Table 1. In general, the results of both the vote count and the combined effect size were statistically significant. These results suggest that students in PBL are better in applying their knowledge (skills). None of the studies reported significant negative findings.

**Table 1: Main effects of PBL**

| Outcome | Sign. + | Sign. - | Studies N | Average ES Unweighted | Weighted (CI 95%) | Qt |
|---|---|---|---|---|---|---|
| Knowledge | 7 | 15 | 18 | -0.776 | -0.223 (+/- 0.058) | 1379.6 ($p$=0.000) |
| Skills | 14 | 0* | 17 | +0.658 | +0.460 (+/- 0.058) | 57.1 ($p$=0.000) |

*= Two-sided sign-test is significant at the 5% level
*Note :* All weighted effect sizes are statistically significant.
*Sign. + / - :* number of studies with a significance (at the 5% level) positive / negative finding.
*Studies N :* the number of total nonindependent outcomes measured

However, table 1 would indicate that PBL has a negative effect on the knowledge base of the students, compared with the knowledge of students in a conventional learning environment. The vote count shows a negative tendency with 14 studies yielding a significant negative effect and only 7 studies yielding a significant positive effect. This negative effect becomes significant for the weighted combined effect size. However, this significant negative result is mainly due to two outliers (Eisenstaedt, Bary & Glanz, 1990; Baca, Mennin, Kaufman & Moore-West, 1990). When these two studies are left aside, the combined effect sizes approaches zero (unweighted ES = -0.051; weighted ES = -0.107).

### Distribution of effect sizes

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

### *Moderators of PBL*

#### Methodological factors

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

*Research design.*

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

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

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

Table 2: Research design as moderating variable

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

*= Two-sided sign-test is significant at the 5% level

*Note :* Unless noted $^{ns}$, all weighted effect sizes are statistically significant.

*Scope of implementation.*

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

Table 3: Scope of Implementation as moderating variable

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

*= Two-sided sign-test is significant at the 5% level

*Note :* All weighted effect sizes are statistically significant.

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

**The analysis of studies examining the effect on knowledge shows that scope of implementation is associated with the variation in effect sizes (Qb= 13.150, *p*= 0.001). If PBL is implemented in a complete curriculum, there is a significant negative effect (see vote count and ES= -0.339). Within a single-course design, the negative effect becomes smaller (ES = 0.113) and the vote-counting suggests a tendency towards more positive effects (6 positive and 4 negative significant effects).**

### Expertise-level of students

**The analysis of the moderators of PBL suggests that significant variation in effect sizes exists for knowledge (Qb= 125.845, *p*= .000) and skills (Qb= 20.63, *p*= .009). The related outcomes are associated with the expertise level of the students. The results of these analyses are summarised in Table 4. It should be noted that when conventional curricula are compared with PBL, the conventional curriculum tends to be characterised by a two- year basic science segment composed of formal courses drawn from various basic disciplines (Albanese & Mitchell, 1993; Richards et al., 1996). On the other hand, in a problem-based learning environment, the students are immediately compelled to apply their knowledge to the problems that they confront. After the first two years of the curriculum, the conventional curriculum emphasises the application of knowledge. The conventional and the problem-based learning environment become more similar (Richards et al., 1996).**

**Table 4: Expertise-level of students as moderating variable**

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

*= Two-sided sign-test is significant at the 5% level

*Note :* Unless noted ns , all weighted effect sizes are statistically significant.

**The differences of the effect sizes between the levels of expertise of the students are remarkable, especially for knowledge-related outcomes. In the second year (ES= -0.315), the negative trend of the first year (ES= -0.153)**

**becomes significant. This is also shown in the vote count. The picture changes completely in the third year. In the third year, both the vote-counting method (two significant positive effects versus zero negative) and the combined effect size (ES= 0.390) suggest a positive effect. Students in the fourth year show a negative effect of PBL on knowledge: a negative tendency in the vote-counting method and a negative combined effect size (ES= -0.496). On the contrary, this negative effect is not found for students who graduated.**

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

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

### Retention period

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

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

**These results suggest that students in PBL remember more of the acquired knowledge. A possible explanation is the attention on elaboration in PBL (Schmidt, 1990): Elaboration promotes the recall of declarative knowledge (Gagné, 1978; Wittrock, 1989). Although the students in PBL would have slightly less knowledge (they do not know as many facts), their knowledge has been elaborated more and consequently they have better recall of that knowledge.**

**Table 5: Retention period as moderating variable**

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

*= Two-sided sign-test is significant at the 5% level
*Note :* All weighted effect sizes are statistically significant.

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

### Type of assessment method

**The authentic studies assessed the effects of PBL on the knowledge and skills of students in very different ways.**

**Table 6: Type of assessment method as moderating variable**

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

*= Two-sided sign-test is significant at the 5% level

*Note :* Unless noted $^{ns}$ , all weighted effect sizes are statistically significant.

The results of the statistical meta-analysis are presented in Table 6. In other contexts, research has shown that assessment methods influence the findings (Dochy, Segers & Buehl, 1999). In this review, the effects of PBL are moderated by the way the knowledge and skills were assessed. The results seem to indicate that the more an instrument is capable of evaluating the skills of the student, the larger the ascertained effect of PBL. Although it is not so clear, an analogue tendency is acknowledged for the knowledge-related outcomes. Students do better on a test if the test makes a stronger appeal on retrieval strategies.

In order to further investigate the moderating effect of the method of assessment on the effects of PBL a second meta analysis was set up: In this second study, we wanted to go a step further and investigate the influence of assessment as the main independent variable. The goal of this study was to describe these effects of PBL from the angle of the underlying focal constructs being measured with the assessment. Using Sugrue's model (1993, 1995) as a frame of reference, the research questions can be formulated as follows. What are the effects of PBL when the assessment of its main goals focuses on respectively (1) the understanding of concepts, (2) the understanding of the principles that link concepts and (3) the linking of concepts and principles to conditions and procedures for application?

From the described main goals of PBL, and the suggestion made from the moderator analysis in the review by Dochy et al. (2003), it is expected that compared to conventional educational methods, the effect of PBL should increase with each level of the knowledge structure.

Forty studies met the inclusion criteria for the meta-analysis, 31 studies were published in peer-reviewed journals, 9 studies were published in edited books. Of the 40 studies, 31 (77%) presented data on knowledge-concepts effects, 17 studies (42%) presented data on knowledge principles-effects and 8 (20%) reported data on effects concerning the application of knowledge (conditions and procedures). These percentages add up to more than 100 since several studies presented outcomes of more than one category (see also appendix D).

Plotting the effect sizes by study, three studies show to be serious outliers (Eisenstaedt, Barry & Glanz, 1990; Mennin et al. 1993; Tans et al., 1986). When these three studies (all situated at the concept-level of the knowledge structure) are left aside, the main effect of PBL on the three levels of the knowledge structure measured emerge to be different. The results of the analysis are summarised in Table 1.

In general, the results of the vote count were statistically significant, except for the assessment of the first level of the knowledge structure. These results suggest that students in PBL perform better at the second and third level of the knowledge structure. None of the studies reported significant negative findings at the third level of the knowledge structure. Only one study reported negative findings at the second level of the knowledge structure. At the level of understanding concepts (first level), the vote count shows a negative tendency with 5 studies yielding a significant negative effect and only 3 studies yielding a significant positive effect. However, the latter difference is not significant at the 5%-level. If we look to the weighted combined effect sizes, the latter negative effect is close to zero but positive (ES= 0.068) based on 21 studies. Based on 15 studies, students studying in PBL classes demonstrated better understanding of the principles that link concepts (ES= 0.795) than students who were exposed to conventional instruction. Based on 13 studies, students in PBL were better at the third level of the knowledge structure (ES= 0.339) than students in conventional classes. Importantly, the average weighted effect size of 0.795 belonging to the second level of the knowledge structure was the only statistically significant result.

Table 7: Main effects of PBL

| Outcome | Sign. + | Sign. - | Studies N | Average ES Unweighted | Average ES Weighted (CI 95%) | Qb | Qw |
|---|---|---|---|---|---|---|---|
| Concepts | 3 | 5 | 21 | -0.042 | 0.068 (+/- 0.864)$^{ns}$ | 18.998$^{**}$ | 113.563$^{**}$ |
| Principles | 17 | 1$^{*}$ | 15 | +0.748 | + 0.795 (+/- 0.782) | | 82.196$^{**}$ |
| Application | 6 | 0$^{*}$ | 13 | +0.401 | +0.339 (+/- 0.662)$^{ns}$ | | 23.356$^{**}$ |

$^{*}$Two-sided sign-test is significant at the 5% level
$^{**}$ $p < 0.05$
*Sign.* + / - : number of studies with a significant (at the 5% level) positive / negative finding.
*Studies N :* the number of total independent outcomes measured
*Note :* Unless noted $^{ns}$ all weighted effect sizes are statistically significant (the 95% confidence intervals do not include zero). Unweighted effect sizes were not tested for significance.

As can be seen from the statistically significant Qb statistics reported in Table 7, the grouping into three levels of assessment, allows to get a better insight into the effects of PBL. However, the results of the homogeneity analysis suggest that further grouping of the data is necessary to fully understand the moderators of the effects of PBL. As indicated by statistically significant Qw statistics, one or more factors other than chance or sampling error account for the heterogeneous distribution of effect sizes.

The results of the meta-analysis showed a difference in the reported effects of PBL between each of the three levels. Different from expectations that the effects of PBL are larger when the method of assessment is more capable of evaluating complex levels, the effect size for the third level of the knowledge structure was smaller compared to the effect size of the second level and not statistically significant. These results imply an implicit challenge for new learning environments to pay more attention to this third level of the knowledge

**structure, both during the learning activities that take place and students' assessment.**

## Opportunities for improvement

**Over the past few years, empirical research has been conducted to identify effective design variables in PBL environments. Basing their studies on empirical work, the model of Gijselaers and Schmidt (1990) demonstrated the importance of problem descriptions and social interaction for determining students' behaviour and learning outcomes. More recent empirical studies, using causal PBL models, have led to similar conclusions. Schmidt and Moust (2000), for instance, showed that, apart from the social functioning of the group, the quality of PBL-problems substantially affects the amount of self-study that is needed and the level of the students' interest. In analyzing curricula from a theoretical point of view, researchers cast social and task related aspects in a similarly prominent role (e.g., Brown et al., 1989; Williams, 1992).**

**Recently, more attention is also asked for the role of assessment in the instructional system. One of the main arguments for stressing the importance of assessment is the general belief and the empirical evidence from various studies that assessment has an important impact on instruction and learning (Gibbs, 1999; Scouller, 1998). It is argued that, in order to make new learning environments effective, the 'constructive alignment' (Biggs, 2003) between the learning environments' characteristics and the assessment is a 'magic bullet' in improving learning (Cohen, 1987). The main purpose is to make assessment congruent with the instruction and align assessment to what students should be learning (Biggs, 2003).**

### Research on team learning as source of inspiration

**PBL capitalizes strongly on small-group work. It is assumed that the collaboration of students will lead to deeper elaboration of the subject matter (e.g. Barron, 2000). Next to that, this small-group work is used to develop teamwork skills (Druskat & Kayes, 2000).**

**However, research and practice shows that this potential effectiveness is not always reached. Research has revealed cases in which large variation in group-work interaction and performance is encountered between teams that seem not to differ in composition and assigned task (Barron, 2000). This research indicates that fruitful collaboration is not merely a case of putting people with relevant knowledge together. Participation in groups often creates more frustration and dislike of groupwork than appreciation for the diversity of perspectives and improved learning which it can result in (Druskat & Kayes, 2000; Salomon & Globerson, 1989).**

**The fact that groupwork does not always reach the potential, begs the need for further understanding of the factors that drive its success. Hereto, we turn to research on team-learning as this research has tried to tackle the essence of collaborative learning: how do groups establish common frames of reference, resolve discrepancies in understanding, negotiate issues of individual an collective action, and come to joint understanding (Barron, 2000; Roschelle, 1992, Van den Bossche, 2006). In a collaborative learning environment, participants are brought together to simultaneously work on a task, in order to learn from this task. This study focuses on groups for which this task**

performance is the primary objective and in which the learning is considered a product of this collaboration for task performance. In this way, learning through collaboration is primarily a group-level phenomenon (Dillenbourg et al., 1996). Collaboration is defined as the process of building and maintaining a shared conception of a problem or task, distributing responsibility across members of the group, sharing expertise and mutually constructing and negotiating cognition (Roschelle, 1992).

In our view the development of shared mental models is more than a cognitive process of integrating and coordinating perspectives. We stress the relevance of socio-cognitive processes that serve as mediators promoting the development of shared mental models. We define these socio-cognitive processes as 'learning behaviors' (see figure 1), stressing how characteristics of the behaviors interact with knowledge building processes that lead to shared mental models (14). Achieving a shared mental model is not only a matter of understanding each other's representation (mutual understanding, but also of accetpting and incorporating each other's ways of seeing (mutual agreement) (Baker, 1995; Alpay et al., 1998). To determine the team interactions that can be considered as team learning behavior we refer to the processes of construction, co-construction and constructive conflict to reach the necessary mutual understanding and agreement.

First, meaning or understanding needs to be (co-)constructed. This process starts when one of the team members inserts meaning by describing the problem situation and how to deal with it, hereby tuning in to fellow team-members. These processes of construction of meaning can evolve into collaborative construction (co-construction), which is a mutual process of building meaning by refining, building on, or modifying the original offer in some way (Baker, 1994). This can lead to 'new' meanings that were not previously available to the group.

Second, agreement needs to be established about the (co-)constructed understandings. It is not sufficient that the inserted meanings are clarified and that there is mutual understanding. They must also be accepted before they form the basis for action (Alpay et al., 1998). However, team members may diverge in their interpretation and tackle the situation from another point of view or perspective. This can lead to a further elaboration through the negotiation of the different meanings. The team will only benefit if divergence in meaning leads to deep-level processing of the diverse information and viewpoints in the team. Through this negotiation by argument and clarification, this is constructive conflict, the team works towards a convergence of meaning in order to reach shared mental models.

It follows from our argumentation that it is important to determine under which conditions the described interactions occur. Roschelle and Teasley (1995) conclude that "collaboration does not just happen because individuals are co-present; individuals must make a conscious, continued effort to coordinate their language and activity with respect to shared knowledge" (p. 94).

The identification of the social conditions under which teams make this effort to reach shared knowledge is an essential prerequisite for developing enhanced understanding of successful collaboration. Viewing collaborative learning as reaching mutually shared cognition, and thus as fundamentally social, stresses the need to take into account the social context in which these processes take place.

**Van den Bossche, Gijselaers, Segers and Kirschner (2006) found evidence supporting their claim that both interpersonal and socio-cognitive processes have to be taken into account to understand the formation of shared mental models, resulting in higher perceived team performance. Interdependence, task cohesion, psychological safety, and group potency, as aspects of the interpersonal context, appeared to be crucial for the engagement in team learning behaviors. The identified team learning behaviors, in turn, give rise to shared mental models. Also it was found that shared mental models are an important factor to understand perceived team effectiveness. One of the most striking findings was that development of shared mental models was largely dependent on building a climate of group trust or psychological safety. This finding confirmed earlier findings of Edmondson (1999) about her research in hospitals showing that team performance was strongly related to psychological safety.**

### Lessons for PBL

**The productivity of the small-group component of pbl can probably be enhanced by taking into account some of the lessons that are drawn from team-learning-research. This is not to say that these groups are similar to teams, but the bottom-line can be that by making these groups more "team-like", their learning processes can be enhanced. Based on the conclusions of the above presented research, both social and cognitive factors can be identified that are crucial in fostering learning. Implications can be drawn for the design of the task environment as for the social and cognitive demands for effective team-learning.**

**The design of the task environment needs to take into account the conditions for the creation of a favourable learning environment. Points of attention can be the creation of interdependence, and capitalization upon task commitment and group potency. Only if students perceive their tasks as interdependent, they will engage in collaborative learning behaviour. This is not automatically implied in the tasks that pbl put forward. Making this interdependence more clear to students and incorporating in the task design by providing more complex tasks or assigning sub-tasks to students, can enhance the collaborative learning process. Moreover, tasks need to be selected which are sufficiently challenging for students, hereby building on task commitment and group potency. This also implies that tasks better are not too complex.**

**The social demands of group learning require that investments are made in the development of a beneficial personal context (e.g., psychological safety). Because of the high time pressure, in a lot of cases no (or little) attention is paid to group developmental processes. Moreover, our experience is that students do not want to (learn) to deal with these interpersonal issues. However, team research shows that investments in this interpersonal context can pay off later in the process. Students and teachers need to pay explicit attention to the basic requirements for fostering interpersonal processes and beliefs that promote learning (e.g., Smith, 1996). This entails that students and professionals need (to learn how) to cope with these interpersonal beliefs and processes. Being able to manage this social side of learning and working should be a goal of staff and management development programs, moreover, it should be a part of each student's curriculum.**

**Small groups in pbl need also have to pay explicit attention to their socio-cognitive processes in order to promote collaborative learning. This means that there needs to be room for construction, co-construction and constructive conflict. This can involve slowing down the interaction in order to inquire about meanings and test understandings (Argyris & Schön, 1996; Marsick, Watkins, & Wilson, 2002). Also, conflicts need to be seen as windows of opportunity instead of threats to progress. The results underline the power of disagreement or conflict (Jehn, 1994), but even more they stress the potential and need of dealing constructively with different opinions that may arise in a team. This is already recognized by staff and students, however, opportunities still remain to be taken in this field.**

## Literature


**Baker, M. J. (1995). Negotiation in collaborative problem-solving dialogues. In R.-J. Beun, M. Baker & M. Reiner (Eds.), *Dialogue and instruction. Modeling interaction in intelligent tutoring systems* (pp. 39-55). Berlin: Springer-Verlag.**

**Alpay, L., Giboin, A., & Dieng, R. (1998). Accidentology: An example of problem solving by multiple agents with multiple representations. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen & T. de Jong (Eds.), *Learning with multiple representations* (pp. 152-174). Amsterdam: Pergamon.**

**Baker, M. (1994). A model for negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence in Education, 5*(2), 199-254.**

**Van den Bossche, P. (2006). Minds in teams. The influence of social and cognitive factors on team learning. Maastricht: Datawyse.**

**Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2003). Effects of short term implementations of problem-based learning within teacher training. *International Journal of Educational Policy, Research & Practice, 3* (4), 55-72.**

**Newman, M., Van den Bossche, P., Gijbels, D., McKendree, J., Roberts,T., Rolfe, I., Smucny, J., & De Virgilio, G. (2004). Responses to the pilot systematic review of problem-based learning. *Medical Education, 38*, 921-923.**

**Gijbels, D., Dochy, F, Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75* (1), 27-61.**

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

**Barrows, H.S. (1996). Problem-based learning in medicine and beyond. In L. Wilkerson, & W.H. Gijselaers (Eds.), *Bringing problem-based learning to higher education: Theory and practice. New directions for teaching and learning, No. 68*. (pp. 3-13). San Francisco: Jossey-Bass Inc. Publishers.**

**Biggs, J. (2003). *Teaching for quality learning at university* (Second edition). Buckingham: Open University Press.**

**Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy & E. Cascallar (Eds.), *Optimising new modes of assessment: In search for qualities and standards* (pp. 13-36). Dordrecht: Kluwer Academic Publishers.**

**Brekelmans, M., van den Eeden, P., Terwel, J., & Wubbels, T. (1997). Student characteristics and learning environment interactions in mathematics and**

physics education: A resource perspective. *International Journal of Educational Research, 27* (4), 283-292.

Cohen, S.A. (1987). Instructional alignment: Searching for a magic bullet. *Educational Researcher, 16* (8), 16-20.

Den Brok, Bergen, T., Stahl, R.J., & Brekelmans, M. (2004). Students' perceptions of teacher control behaviours. *Learning and Instruction, 14* (4), 425-443.

Dochy, F., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation, 23* (4), 279-298.

Entwistle, N.J., & Tait, H. (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education, 19* (2), 169-194.

Feltovich, P.J., Spiro, R.J., & Coulson, R.L. (1993). Learning, teaching, and testing for complex conceptual understanding. In N. Frederksen, R. J. Mislevey & I.I. Bejar (Eds.), *Test theory for a new generation of tests*. Hilsdale, NJ: Erlbaum.

Fraser, B.J., Walberg, H.J., Welch, W.W., Hattie, J.A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11*, 145-252.

Gagné, E. D., Yekovich, C. W., & Yekovich, F. R. (1993). *The cognitive psychology of school learning* (2nd ed.). New York: HarperCollins College publishers.

Gibbs, G. (1999). Using assessment strategically to change the way students learn. In S. Brown & A. Glasner (Eds.), *Assessment matters in higher education: Choosing and using diverse approaches* (pp.41-53). Buckingham: Open University Press.

Gijselaers, W. (1995). Perspectives on problem-based learning. In W. Gijselaers, D. Tempelaar, P. Keizer, J. Blommaert, E. Bernard & H.Kasper (Eds.), *Educational innovation in economics and business administsration: The case of problem-based learning* (pp. 39 - 52). Norwell, Mass.: Kluwer.

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

Harris, K. R., & Alexander, P. A. (1998). Integrated, constructivist education: Challenge and reality. *Educational Psychology Review, 10* (2), 115-127.

Hendry, G. D., Frommer, M., & Walker, R.A. (1999). Constructivism and Problem-based learning. *Journal of Further and Higher Education, 23* (3) 359-371.

Lea, S. J., Stephenson, D., & Troy, J. (2003). Higher education students' attitudes toward student-centred learning: beyond 'educational bulimia'? *Studies in Higher Education, 28* (3), 321-334.

Newman, M., Van den Bossche, P., Gijbels, D., McKendree, J., Roberts,T., Rolfe, I., Smucny, J., & De Virgilio, G. (2004). Responses to the pilot systematic review of problem-based learning. *Medical Education, 38*, 921-923.

Russell, A. L., Creedy, D. & Davis, J. (1994). The use of contract learning in PBL. In: S. E. Chen, S. E. Cowdroy, A. J. Kingsland & M. J. Ostwald (Eds.), *Reflections on problem based learning*, (pp. 57-72). Sydney: Australian Problem Based Network.

Sambell, L., McDowell, L., & Brown, S. (1997). "But is it fair?": An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation, 23* (4), 349-371.

Savery, J. R., & Duffy, T. M. (1995). Problem-based learning: An instructional model and its constructivist framework. *Educational Technology, 35*, 31-38.

Segers, M., Dochy, F., & Cascallar, E. (2003). The era of assessment engineering: Changing perspectives on teaching and learning and the role of new modes of assessment. In M. Segers, F. Dochy & E. Cascallar (Eds.), *Optimising new modes of assessment: In search for qualities and standards* (pp. 1-12). Dordrecht: Kluwer Academic Publishers.

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

Simons, R.J., van der Linden, J., & Duffy, T. (2000). New learning: Three ways to learn in a new balance. In R.J. Simons, J. van der Linden & T. Duffy (Eds.), *New learning* (pp. 1-20). Dordrecht: Kluwer Academic Publishers.

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