CARFAX

An analysis of progress test results of PBL and non-PBL students

B. H. VERHOEVEN[1], G. M. VERWIJNEN[1], A. J. J. A. SCHERPBIER[1], R. S. G. HOLDRINET[2], B. OESEBURG[2], J. A. BULTE[2] & C. P. M. VAN DER VLEUTEN[1]
[1]*Faculty of Medicine, University of Maastricht, The Netherlands;* [2]*Faculty of Medical Sciences, University of Nijmegen, The Netherlands*

SUMMARY *This study compares the academic achievement of students from two Dutch medical schools, one employing problem-based learning (PBL) and one non-PBL methods. Two progress tests were administered to 1904 PBL students and 1089 non-PBL students. No systematic differences were found on total test scores. After the test was split into three categories (basic, clinical and social sciences), a few, non-systematic differences were found. To some extent the basic sciences favoured the non-PBL curriculum and the social sciences the PBL curriculum, but this was not consistent for both test administrations. Also, the profiles for both schools at the level of disciplines show remarkable resemblances. Only at the level of individual questions could differences between PBL and non-PBL be demonstrated. The results obtained in this study strongly agree with those of previous comparisons between medical schools but indicate that effects of PBL and non-PBL instructional methods on medical factual knowledge are even more similar than we had previously thought.*

Introduction

Since the introduction of problem-based learning (PBL), many studies have been reported comparing the effectiveness of PBL curricula with more traditional educational programmes (Albanese, 1993; Berkson, 1993; Vernon & Blake, 1993). One aspect of interest is the level of factual medical knowledge that graduates of a problem-based learning curriculum have in comparison with graduates from more traditional schools. Most research in this respect uses student achievement results on licensure examinations such as the USMLE (formerly NBME). One part of these exams covers basic sciences and most students take this part at the end of the second year. Another part covers the clinical sciences and is most often taken during the fourth year (Schmidt *et al.*, 1987; Swanson *et al.*, 1992; NBME, 1997). In so far as differences occurred, PBL students had lower test scores on the basic sciences part and better test scores on the clinical sciences part (Albanese, 1993; Vernon & Blake, 1993). Several other examinations required for licensure and graduation, including oral exams and clinical rotation ratings, have also been used for comparisons. The differences found between PBL and non-PBL were not very consistent. The selection of students and implementation differences in PBL probably account for these inconsistent findings (Woodward, 1996). The fact that some studies mention differences between PBL and non-PBL students in knowledge of basic sciences at the end of the second year does not have to mean that PBL students eventually have less knowledge of basic sciences at the end of the curriculum (Van der Vleuten, Verwijnen & Wijnen, 1996). It is plausible that the gathering of specific medical knowledge does not occur at the same moment in time in a non-PBL curriculum as in a PBL curriculum.

In this study the medical factual knowledge of PBL and non-PBL students is compared at all levels of training. The major differences between the two medical schools involved are found in the first four years. The PBL school offers a problem-based, student-centred and integrated curriculum that is organized in themes. The programme consists of interdisciplinary units of usually six weeks (Van der Vleuten, Scherpbier *et al.*, 1996). More attention is paid to basic and social sciences during the first two years than in years 3 and 4. The clinical sciences dominate in years 3 and 4. At the time this study was conducted, the non-PBL school had a conventional discipline-oriented educational method marked by instructor-provided learning objectives and assignments, large-group lectures and structured laboratory experiences. The teaching of the principles of basic sciences (mainly years 1 and 2) preceded the instruction in clinical sciences (mainly years 3 and 4). In the last two years all Dutch medical students spend their time on clinical rotations and no major programme differences exist between the medical schools.

At the time of the study a curriculum change was in preparation in the non-PBL school. As part of the preparation, the PBL school's Progress Test (PT) was

Correspondence: B. H. Verhoeven, Skillslab, University of Maastricht, PO Box 616, 6200 MD Maastricht, The Netherlands, Tel: (31) 43 3881789. Fax: (31) 43 3618612. Email: b.verhoeven@sk.unimaas.nl

0142-159X/98/040310-07 $9.00 © 1998 Carfax Publishing Ltd

Table 1. Number of participating students on two administrations of the Progress Tests (percentages of non-PBL classes in parentheses).

	December 1994			March 1995		
Year	PBL	non-PBL	(% of class)	PBL	non-PBL	(% of class)
1	190	124	(68)	218	54	(30)
2	146	104	(64)	151	60	(37)
3	135	87	(49)	133	51	(29)
4	188	151	(58)	169	53	(20)
5	144	140	(68)	155	69	(33)
6	135	122	(59)	140	74	(36)

administered to students of all six classes of the non-PBL school. This was done to get experience with administration of this kind of test and to provide a base-line for further research guiding the curriculum change. Ideally a random sample of students had to participate, but this could, for practical reasons, not be realized. Therefore volunteers were asked to take the PTs. Volunteers are mostly the better students, but this methodological shortcoming had to be taken for granted.

We were interested whether there would be differences in students' results on the Progress Test with regard to (1) total test score (overall medical knowledge); (2) subscores in basic sciences, clinical sciences and social sciences; (3) subscores in disciplines (anatomy, surgery, ENT, etc.) in all curriculum years; and (4) scores on individual test items.

Methods

Instrument

The results on the PT were used in this comparison (Verwijnen *et al.*, 1982). The PT is best characterized as a comprehensive final examination in medicine. Each PT samples the complete domain of factual knowledge that is considered pertinent for a medical graduate to master and consists of approximately 250 true/false questions with a question mark ('I do not know') option. These questions are stratified in categories based on the International Classification of Diseases (ICD). The content of the questions may or may not be covered by the educational programme. Therefore the PT can be called a 'programme independent' test and can be administered at every medical school regardless of its specific educational programme. The PT is given four times a year to all students in the curriculum regardless of their class. An equivalent test is produced for each occasion on the basis of a blueprint. In this way it is possible to monitor progress or growth towards the overall objectives (Van Hessen & Verwijnen, 1990; Verwijnen *et al.*, 1990; Van der Vleuten, Verwijnen & Wijnen, 1996). Two progress tests were used in this study: the PT of December 1994 consisting of 242 items; the PT of March 1995 comprising 226 items.

Subjects

Students of all six classes of two medical schools in the Netherlands, one PBL and one non-PBL, were used (the medical training programme takes 6 years in The Netherlands). The freshmen of the eight medical schools are comparable in respect of age, gender and knowledge base. They enter medical school directly after secondary education. The secondary school system in The Netherlands is quite homogeneous and students only graduate after passing a national secondary school examination. Medical school admission is centralized. The selection procedure is partly based on grade-point average and partly on a lottery, and is therefore called a 'weighted' lottery system (Wijnen, 1978). After this selection, students are assigned to one of the eight medical schools in The Netherlands. No academic differences between entering medical students across schools are to be expected through this system.

Table 1 shows the number of participating students from both schools per class as well as the percentage of the participating non-PBL students in relation to the total number of students in their year.

Procedure

Two PTs were administered. At the PBL school the PT is part of the regular assessment programme and the test results had consequences for study progress. At the non-PBL school the PT was not part of the assessment programme. All participating students were volunteers and their test scores had no consequences for advancement decisions. We had no information on the representativeness of the volunteer group. The two administrations used in this study took place in December 1994 and in March 1995. On both occasions it concerned fully synchronized administrations at the same moment in time only at different places.

Statistical analysis

To discourage guessing, formula scoring was used in which the incorrect scores are subtracted from the correct scores (the 'I don't know' option receives no marks). For each student the scoring was expressed as a percentage of the total number of questions. For both PTs the means and standard deviations were calculated per class and per faculty. Total scores as well as subscores on basic sciences (anatomy, biochemistry, pharmacology, physiology, genetics and cell biology, immunology, microbiology and

Table 2. Students' mean correct minus incorrect scores and standard deviations (SD) on total test and clusters of disciplines.

	December 1994					March 1995				
		PBL		Non-PBL			PBL		Non-PBL	
Year	No. of items	Mean	SD	Mean	SD	No. of items	Mean	SD	Mean	SD
Total test	242					226				
1		9.61	4.06	10.70	5.37		11.42	4.29	12.17	6.85
2		20.68	4.94	19.27	6.68		20.64	6.43	22.26	8.27
3		26.18	6.68	23.50	9.32		26.51	6.88	30.37†	9.14
4		33.27	7.98	31.54	8.95		35.14	7.64	38.03	9.28
5		38.46	8.53	38.80	9.16		38.41	8.59	41.52	9.37
6		41.03	8.01	44.49†	8.99		42.12	6.71	44.85	11.07
Basic sciences	92					93				
1		8.43	4.77	10.68†	6.78		11.26	5.66	14.97†	8.20
2		19.09	8.27	16.66	8.59		18.78	8.19	25.48†	10.74
3		21.08	8.56	19.63	10.00		24.46	8.88	31.58†	11.06
4		25.86	9.31	25.66	10.50		32.46	8.96	37.74†	10.02
5		30.13	10.97	31.16	11.70		36.25	10.39	38.41	10.30
6		32.33	10.21	37.07†	11.00		36.49	8.24	41.08†	12.09
Clinical sciences	103					91				
1		8.24	4.74	9.23	6.65		4.77	4.57	3.74	7.90
2		19.26	6.94	20.46	8.13		14.41	7.46	13.33	8.53
3		26.39	8.62	26.09	10.60		23.10	8.69	25.02	10.99
4		37.88	9.80	36.15	10.80		34.38	10.33	35.41	11.92
5		44.56	9.71	45.30	10.40		38.72	11.06	41.77	13.35
6		48.71	10.04	52.51†	11.10		46.44	9.12	48.60	12.85
Social sciences	47					42				
1		14.93	9.74	13.98	10.30		26.18	11.90	24.21	14.33
2		26.90	11.20	21.77†	10,40		38.25	12.40	34.44	12.78
3		35.70	12.71	25.39†	13.80		38.42	12.67	39.26	12.57
4		37.65	12.86	32.93†	13.90		42.72	12.25	44.34	14.42
5		41.43	13.04	39.53	12.10		42.53	13.40	47.83†	13.19
6		41.26	11.36	41.42	12.37		45.24	11.15	45.05	16.74

Note: Statistical significance at the level $p < 0.01$ (one-way ANOVA).

pathology), clinical sciences (surgery, cardiology, dermatology, gynaecology and obstetrics, family medicine, internal medicine, paediatrics, ENT, neurology, orthopaedics, ophthalmology, pulmonology and rehabilitation medicine) and social sciences (health care economy, epidemiology, health care law, ethics and philosophy, medical psychology, medical sociology and social psychiatry) were compared. To test the significance of the differences between the mean total scores and subscores per year group of both schools, a one-way ANOVA was used. To correct for the number of comparisons, a Bonferroni correction was used and $p \leq 0.01$ was considered to be significant. The correct minus incorrect scores on disciplines and items were compared using visual inspection of plots and correlation coefficients (Pearson product-moment correlation coefficient).

Results

In Table 2 the correct minus incorrect scores of both schools are presented per year on the complete test as well as on the three clusters of disciplines (basic, clinical and social sciences). The statistically significant differences in scores found in December 1994 do not correspond with those found in March 1995. Only the differences that are found in the basic sciences scores in the first and sixth years are significantly different in both December and March. Remarkably, however, this difference does not exist for year five. In March 1995 the significant differences mainly concern the basic sciences in favour of the non-PBL group, whereas in December 1994 the significant differences are found in the social sciences domain in favour of the PBL group.

At the discipline level of comparison, each discipline may be represented by only a few test items, causing a considerable spread of scores. For this reason statistical significance testing has not been used and a visual inspection of the profile was preferred. In figures la to ld the mean test score on each discipline is depicted for the first and sixth year separately. The parallelism of the profile of the curves is striking. There is quite some variation across disciplines, but this is (almost completely) identical for

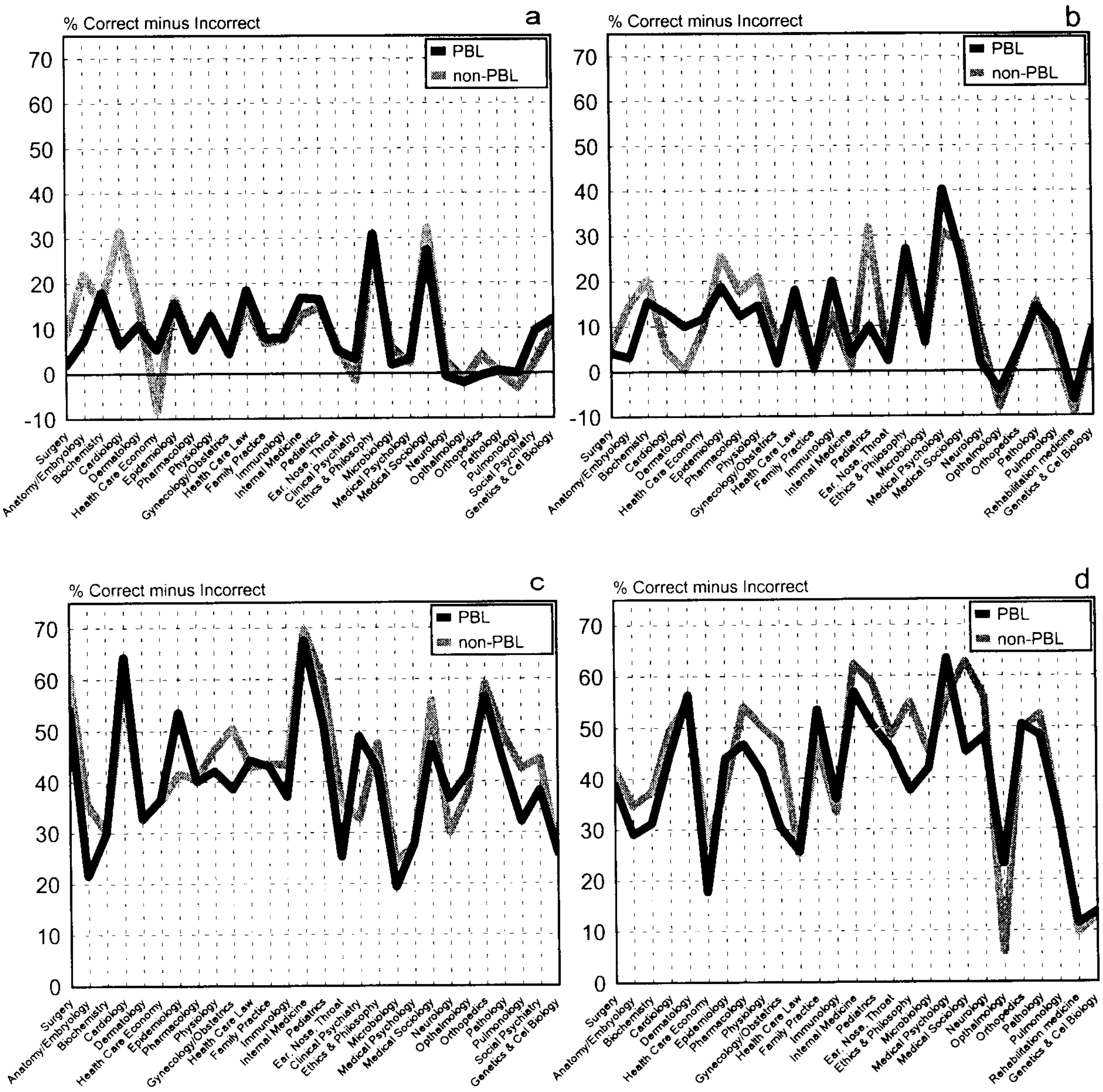

Figure 1. Disciplinary knowledge profiles (mean correct minus incorrect score on each discipline) of PBL and non-PBL students: (a) first-year students, Progress Test of December 1994; (b) first-year students, Progress Test of March 1995; (c) sixth-year students, Progress Test of December 1994; (d) sixth-year students, Progress Test of March 1995.

both schools. Interestingly, the profiles are not the same across both test administrations.

The most detailed level of comparison is a comparison of the 'item scores'. These scores represent the percentage of students who answered a question correctly minus the percentage of students who answered the same question incorrectly. These 'item scores' (242 test items in December and 226 in March) of both faculties are plotted in scatter diagrams for years one, three and six (Figures 2 to 4). Data points on the diagonal represent perfect resemblance between the scores of the PBL and non-PBL students. The more distant a data point is from the diagonal, the less resemblance is present between the two schools on that specific item, suggesting that the PBL and non-PBL students have different (curriculum-specific) knowledge. Figure 2 contains the data for year one. A cluster of data points is situated near the intersection of the two axes of the graph, meaning that many items remain unanswered or that many students in both faculties do not have the knowledge to answer those items correctly. Some items are situated near the diagonal and further away from the intersection of the axis. Students of both faculties have the same level of knowledge on these items. In addition, some items are answered correctly by either more non-PBL or more PBL students, suggesting the existence of curriculum-specific knowledge on these items. In the third year (Figure 3), fewer data points are situated near the intersection, meaning that students, as can be expected, answer more questions. However, quite a lot of items are answered correctly by either more non-PBL or more PBL students. This suggests curriculum-specific knowledge on those questions. The number of items that have a negative

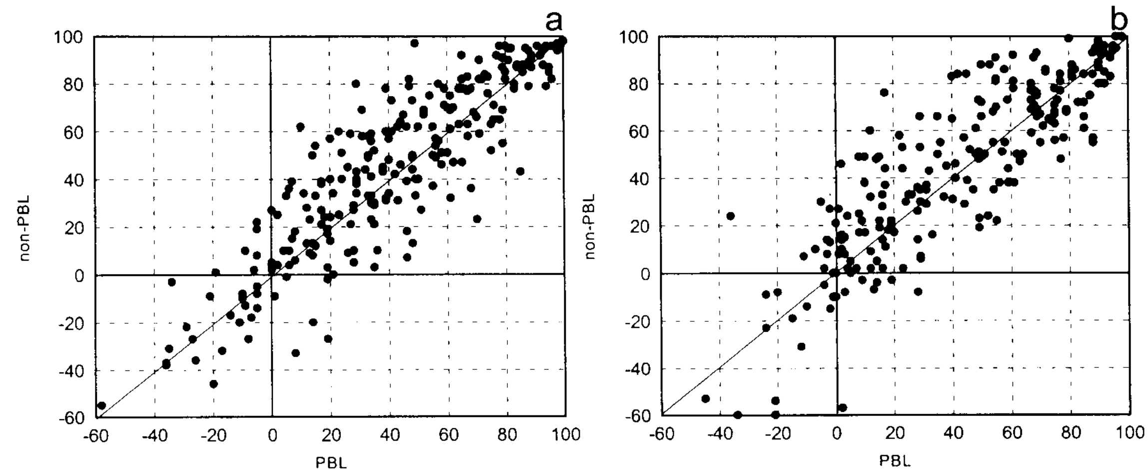

Figure 4. Comparison of correct minus incorrect 'item scores' of sixth-year PBL and non-PBL students: (a) all 242 items of Progress Test of December 1994; (b) all 226 items of Progress Test of March 1995.

questions from a discipline are more difficult or easier irrespective of the school's training method. The level of difficulty of a cluster of questions within a certain discipline or category appears to be rather a generic phenomenon, varies from test to test, and is independent of the curriculum that one follows. Only at the level of individual questions could differences between PBL and non-PBL be demonstrated. In the first year of medical school many questions can be found that are answered better by PBL or non-PBL students. However, this effect diminishes systematically with increasing seniority in the curriculum. At the end of the sixth year these differences at item level too yielded nearly total parallelism. At the end of their curricula, PBL and non-PBL students answer the same questions correctly and incorrectly. The students master the same knowledge; it is only the moment in time that they learn it that differs.

Methodological considerations

A few methodological considerations can be mentioned with regard to this study. In the non-PBL medical school all participating students were volunteers. The test results had no consequences for them in contrast with the PBL students, for whom the tests were obligatory and part of their exams. Both aspects are essential methodological shortcomings of this study and make a valid interpretation of this comparison more difficult (Schmidt, 1990). Volunteers in most cases are the better students (Ten Cate, 1985), and tests that have no consequences for the participants lead to different answer strategies, particularly more guessing. Indeed, we found that the volunteers (non-PBL students) did use significantly fewer question marks (data not shown). To correct for these different answer strategies, formula scoring was used in this comparison.

Perhaps the questions that are used in the PT are not sensitive enough to detect differences between curricula. They could be too rough, not tailored enough, asking about only those knowledge aspects that everyone will learn anyway by spending six years in a medical school. The real differences could be in topics that are not included in the PT and therefore are still to be discovered. Research, however, shows that the PBL students experience the PT as a relevant test that reflects topics covered by their curriculum (Linden *et al.*, 1995). This, plus earlier comparisons, suggests the PT is a valid and valuable instrument to compare medical curricula, even internationally (Albano *et al.*, 1996).

Despite these methodological shortcomings, this study indicates that the effects of PBL and non-PBL instructional methods on medical factual knowledge output are even more similar than we had previously thought.

Table 3. Pearson correlation coefficients of item scores of PBL and non-PBL students on two administrations of the Progress Test.

Year	December 1994	March 1995
1	0.72	0.66
2	0.69	0.69
3	0.73	0.76
4	0.83	0.78
5	0.87	0.85
6	0.88	0.86

Notes on contributors

B. H. VERHOEVEN is a Physician, a Researcher in the Skillslab and Interim Chairman of the Progress Test Review Committee at the University of Maastricht.

G. M. VERWIJNEN is a General Practitioner, senior lecturer in the Skillslab and Chairman of the Progress Test Review Committee at the University of Maastricht.

A. J. J. A. SCHERPBIER is a Physician, Associate Professor and Head of the Skillslab at the University of Maastricht.

R. S. G. HOLDRINET is Associate Professor of Internal Medicine (Hematology) and Chairman of the Curriculum Committee, Faculty of Medical Sciences, Nijmegen University.

B. OESEBURG is Professor of Physiology and Vice-chairman of the Curriculum Committee, Faculty of Medical Sciences, Nijmegen University.

J. A. BULTE is Senior Educationalist and Secretary of the Curriculum Committee, Faculty of Medical Sciences, Nijmegen University.

C. P. M. VAN DER VLEUTEN is an Educationalist and Professor and Chair of the Department of Educational Research and Development at the University of Maastricht.

References

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

ALBANO, M.G., CAVALLO, F., HOOGENBOOM, R., MAGNI, F., MAJOOR, G., MANENTI, F., SCHUWIRTH, L., STIEGLER, I. & VAN DER VLEUTEN, C. (1996) An international comparison of knowledge levels of medical students: the Maastricht Progress Test, *Medical Education*, 30, pp. 239–245.

BENDER, W., COHEN-SCHOTANUS, J., IMBOS, T., VERSFELT, W.A. & VERWIJNEN, G.M. (1984) Medische kennis bij studenten uit verschillende medische faculteiten: van hetzelfde laken een pak? [Medical knowledge in students from various medical schools: being served with the same sauce?], *Nederlands Tijdschrift voor Geneeskunde*, 128, pp. 917–921.

BERKSON, L. (1993) Problem-based learning: have the expectations been met?, *Academic Medicine*, 68 (Suppl. 10), pp. S79–S88.

FACULTEIT DER MEDISHE WETENSCHAPPEN (1994) *Studiegids Geneeskunde 1994/1995 [Study guide Medicine 1994/1995]* (Nijmegen, Universiteitsdrukkerij).

LINDEN, M.L., BRÜMMER, I., HOOGENBOOM, R.J.I., SCHERPBIER, A.J.J.A. & VERWIJNEN, G.M. (1995) De Maastrichtse voortgangstoets. Een vergelijking van drie peilingen van het consumentenoordeel [The Maastricht Progress Test. A comparison of three samplings of consumer judgment], *Bulletin Medisch Onderwijs*, 14, pp. 186–191.

NATIONAL BOARD OF MEDICAL EXAMINERS (1997) 1995 Performance on USMLE Examinations, < http//www.nbme.org/step95.htm > , accessed 25 March 1997.

SCHMIDT, H.G. (1990) Innovative and conventional curricula compared: what can be said about their effects?, in: Z.M. NOOMAN, H.G. SCHMIDT & E.S. EZZAT (Eds) *Innovation in Medical Education: An Evaluation of its Present Status*, pp. 1–7 (New York, Springer Verlag).

SCHMIDT, H.G., DAUPHINEE, W.D. & PATEL, V.L. (1987) Comparing the effects of problem-based and conventional curricula in an international sample, *Journal of Medical Education*, 62, pp. 305–315.

SNELLEN-BALENDONG, H.A.M., WIJNEN, W.H.F.W. & LANGEVOORT, H.L.L. (1994) Diversiteit van medische curricula in Nederland [Diversity of medical curricula in the Netherlands], *Nederlands Tijdschrift voor Geneeskunde*, 138, pp. 1136–1142.

SWANSON, D.B., CASE, S.M., MELNICK, D.E. & VOLLE, R.L. (1992) Impact of the USMLE Step 1 on teaching and learning of the basic biomedical sciences, *Academic Medicine*, 67, pp. 553–556.

TEN CATE, T.J. (1985) De invloed van anonimiteit op de resultaten van tijdschrijfonderzoek [The influence of anonymity on the results of time registration research], *Tijdschrift voor onderwijsresearch*, 10, pp. 263–273.

VAN DER VLEUTEN, C.P.M., SCHERPBIER, A.J.J.A., WIJNEN, W.H.F.W. & SNELLEN, H.A.M. (1996) Flexibility in learning: a case report on problem-based learning, *International Higher Education* (March), pp. 17–24.

VAN DER VLEUTEN, C.P.M., VERWIJNEN, G.M. & WIJNEN, W.H.F.W. (1996) Fifteen years of experience with progress testing in a problem-based learning curriculum, *Medical Teacher* 18, pp. 103–109.

VAN HESSEN, P.A.W. & VERWIJNEN, G.M. (1990) Does problem-based learning provide other knowledge?, in: W. BENDER, R.J. HIEMSTRA, A.J.J.A. SCHERPBIER & R.P. ZWIERSTRA (Eds) *Teaching and Assessing Clinical Competence*, pp. 446–451 (Groningen, Boekwerk Publications).

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

VERWIJNEN, M., IMBOS, T., SNELLEN, H., STALENHOEF, B., POLLEMANS, M., VAN LUYK, S., SPROOTEN, M., VAN LEEUWEN, Y. & VAN DER VLEUTEN, C. (1982) The evaluation system at the medical school of Maastricht, *Assessment and Evaluation in Higher Education*, 7, pp. 225–244.

VERWIJNEN, M., VAN DER VLEUTEN, C. & IMBOS, T. (1990) A comparison of an innovative medical school with traditional schools: an analysis in the cognitive domain, in: Z.M. NOOMAN, H.G. SCHMIDT & E.S. EZZAT (Eds) *Innovation in Medical Education: An Evaluation of its Present Status*, pp. 40–49 (New York, Springer Verlag).

WIJNEN, W.H.F.W. (1978) The Netherlands, in: B.B. BURN (Ed.) *Admission to Medical Education in Ten Countries*, pp. 105–114 (New York, International Council for Educational Development).

WOODWARD, C.A. (1996) Problem-based learning in medical education: developing a research agenda, *Advances in Health Science Education*, 1, pp. 83–94.