

## An international comparison of knowledge levels of medical students: the Maastricht Progress Test

*M G Albano,<sup>1</sup> F Cavallo,<sup>2</sup> R Hoogenboom,<sup>3</sup> F Magni,<sup>4</sup> G Majoor,<sup>3</sup> F Manenti,<sup>5</sup> L Schuwirth,<sup>6</sup> I Stiegler<sup>7</sup> & C van der Vleuten<sup>6</sup>*

1Institute of Clinical Medicine I, Faculty of Medicine, University of Bari, Bari, Italy; 2Department of Hygiene and Community Medicine, Faculty of Medicine, University of Torino, Torino, Italy; 3Dean's Office, Faculty of Medicine, University of Limburg, Maastricht, The Netherlands; 4Institute of General Pathology, Faculty of Medicine and Surgery, University of Perugia, Perugia, Italy; 5Department of Gastro-enterology, University of Modena, Modena, Italy; 6Department of Educational Research and Development, Faculty of Medicine, University of Limburg, Maastricht, The Netherlands; and 7Dean's Office, Faculty of Medicine, University of Witten/Herdecke, Witten, Germany

### SUMMARY

The increasing international mobility of medical students has inspired the search for an international assessment format. As one step along this line, kinetics of knowledge acquisition and final cognitive levels of students were compared among one Dutch, one German and four Italian medical faculties. For this comparison, the Maastricht Progress Test (MPT) was used. For four out of the six participating faculties, it was possible to compare the level of knowledge of sixth-year students. These data showed no significant differences on the test as a whole. On the other hand, as judged from cross-sectional data on students from all study years, the kinetics of knowledge acquisition showed different trends. In one school applying problem-based learning, acquisition of knowledge by students occurred almost linearly. In another school, over the first 2 years, acquisition of knowledge occurred only in the basic sciences but not in clinical or public health/behavioural sciences. In two other schools over that same period, students seemed to gain no knowledge at all. In some faculties, a marked boost in knowledge was noted with third- or fourth-year students. These findings may be explained by peculiarities of the respective curricula, selection of students during their studies, and national or local assessment procedures. It is preliminarily concluded that the different educational approaches and assessment systems in medical education in Europe seem to have only limited influence on the final level of knowledge of the graduates. On the other hand, these differences may influence the kinetics of knowledge acquisition, especially in distinct domains like basic or clinical sciences. Therefore, the MPT may not be suitable to solve the problem of assessment of individual international exchange students, but it may be helpful in identifying corresponding cogni-

tive levels on, for example, basic sciences for students in different curricula.

### Keywords

Clinical competence; \*cognition; curriculum; International Educational Exchange; Netherlands; School Admission Criteria; students medical/\*psychology

### INTRODUCTION

Over the last decade, there has been a growing interest in the international positioning of institutions for higher education, especially with respect to participation in student mobility programmes. Within the European Union, this interest has been strengthened by financial support from programmes like ERASMUS, LINGUA and TEMPUS. However, the participation of medical students in these programmes has lagged behind, as compared to most other academic disciplines (Bourgeois 1992). One of the explanations provided for this phenomenon derived from the fact that the curricula of European medical schools are differently scheduled. Consequently, most students participating in international exchange programmes had to accept that the match between their home and host faculties would be imperfect and, as a consequence, that they would not receive credits for all study time abroad. Furthermore, these students may be requested to re-sit for exams missed when on leave. Since these unwanted effects also occurred in other disciplines, the European Commission created a special pilot programme to experiment with a transfer system of course credit points (the European Community Course Credit Transfer System, ECTS). For medicine, another approach could be to consider a trans-European assessment system akin to the United States Medical Licensing Examination.

Against this background, the present study aimed to investigate whether differences exist with respect to final levels of knowledge and the kinetics of knowledge acquisition of medical students in three European countries. Furthermore, the outcome of the study could indicate whether the MPT is applicable to assess individual students after a period of study in another European country.

## MATERIALS AND METHODS

### Participating medical faculties

Six European faculties of medicine participated in this study: one Dutch, one German, and four Italian schools. In this report, each faculty will only be designated by a capital letter. Curricula of all participating faculties had a nominal study time of 6 years. Two of the participating faculties had fully integrated, problem-based medical curricula; one other faculty had some students participating in such programme in a 'parallel track', whereas all other students followed a conventional curriculum. Students in three schools exclusively followed conventional, discipline-orientated curricula. In this study, students were classified according to the curriculum year they actually followed, i.e. if a second-year student had to redo the second year, he or she was again considered a student in year 2.

### The Maastricht Progress Test

The format of the test used for the comparison was the Maastricht Progress Test (MPT; Van der Vleuten & Verwijnen 1990). In brief, the MPT is a written test consisting of about 250 true-false items sampling the full domain of knowledge that a medical graduate should master. For each issue of the MPT, a stratified random selection is made from a computerized item bank containing some 16,000 items. The stratification is based on the International Classification of Diseases (ICD, WHO 1992). Data on reliability and validity of the MPT have been summarized by Van der Vleuten & Verwijnen (1990). At the Maastricht Faculty of Medicine (MFM), a new issue of the MPT is given to all students of all classes four times a year.

One of these tests, the MPT sat by MFM students in June 1992, was used for the international comparison. Originally, this test included 217 items. Twelve items were removed after the Maastricht students had completed the test, based on valid comments by these students. Furthermore, for the international comparison, seven items were removed that specifically addressed the situation in The Netherlands (e.g. on Dutch health insurance systems). For the remaining 198 items, eight

departments (including Microbiology and Pathology) contributed 72 questions to the basic sciences domain, 15 departments (including Rehabilitation Medicine) contributed 99 questions to the clinical sciences domain, and six departments (including Health Economics, Health Law, and Epidemiology) contributed 27 questions to the public health/behavioural sciences domain. These 198 items were translated from Dutch into German and Italian, and checked on the wording by medical experts.

### Sample selection

One of the practical drawbacks of the MPT is that, ideally, all students within one school should sit for the test at the same time. For the MFM, this was organized as usual. All 811 participants were seated for a maximum of 4 h in one big hall. For some of the other faculties, because of the large size of the classes, it proved impossible to have all students participating. In these cases, it was advised that entire classes should be dropped (e.g. years 2 and 4) rather than any attempt made to select a sample of participants from each class. At some faculties with large classes, however, a random sub-division within classes already existed. In that case, some faculties preferred to include one sub-class rather than the full class of students of a given year (Table 1). Another problem in organizing this comparison was matching as closely as possible the motivation of the students to participate in the MPT. At MFM, the MPT is a major element in the students' summative assessment scheme. Therefore, every student considers the test as an examination and will try to achieve the best score possible. At faculty U, the test was also included as an element in the summative assessment programme for students in years 1-4. In most of the other faculties, however, summative assessment is mainly based on disciplinary exams. Therefore, the MPT could not be inserted in the regular scheme of exams, particularly because it was hard to get the participation of full classes. On the other hand, faculties S and T found teachers who agreed to marginally raise the marks on their disciplinary exam if the student would pass the MPT (see below). Nevertheless, students failing the test would not lose any credit points.

### Scoring of the MPT

All participating students used optically readable scoring forms, as developed and used at MFM. Forms used abroad were returned to MFM for reading and computer processing.

Students' individual scores were calculated by subtracting the number of incorrect answers from the number of correct answers, ignoring the items for which no

**Table 1** Numbers of students participating in the international MPT comparison (full class size in parentheses)

Faculty	Months in training [+ (Y-1) × 12]	Study year (Y)					
		1	2	3	4	5	6
Q*	4	48 (300)	0	0	12	0	121 (130)
R*	6	82 (214)	82 (213)	92 (191)	54 (94)	27 (75)	2
S†	7	85 (124)	84 (115)	86 (108)	57 (86)	39 (66)	35 (73)
T†	8	128 (360)	174 (360)	5*	0	9	23 (140)
U‡	9	30 (30)	29 (29)	24 (24)	21 (21)	3*	0
V‡	10	164 (164)	140 (140)	115 (115)	155 (155)	117 (117)	120 (120)

\* Voluntary student participation without credit reward.

† Voluntary student participation with credit reward on passing the MPT.

‡ Mandatory student participation (included in the faculty's summative assessment scheme).

answer was provided (C-I scores). This method is applied to discourage guessing on the items by students (Verwijnen, Van der Vleuten & Imbos 1990). The resulting score was converted to a percentage of the number of questions in the test, i.e. 198. To decide whether a student had passed the test for each class, the mean score and the standard deviation (SD) were calculated. Any student scoring below the mean score of his or her own class minus 1.0 SD was considered to have failed the test.

Each student participating in this comparison received two feed-back forms on their achievement as compared to the class. On one form, their individual scores and the mean scores of their peer group were given grouped according to the 15 test strata. On the second form, the results of the student and his or her peer group were split up over the 29 departments that had contributed items to this test, distributed over the three domains. For each domain, a separate sub-score was calculated.

### Statistics

For each participating class in all faculties, descriptive statistics of the mean C-I scores on the full test and on the three domains were calculated (Table 2). To determine whether differences in mean C-I scores could be considered statistically significant, standard errors were calculated and 95% confidence intervals created. Thus, differences between mean C-I scores were considered significant at the  $P < 0.05$  level.

Since the number of participating students per class per faculty differed considerably, a decision had to be taken with respect to the minimum number of students per class that still allowed comparisons. A compromise had to be made between the number of students per class and the number of possible comparisons. Therefore, comparisons between classes were only made provided the

number of participating students in both classes exceeded 20.

## RESULTS

### Comparison of sixth-year cognitive levels

For the purpose of comparison of knowledge levels of final-year students, only data from faculties Q, S, T and V could be used. As shown in Fig. 1A, the range of mean C-I scores of sixth-year students (neglecting the timepoint in the sixth year at which the MPT was taken) on the test as a whole was small (41.47-44.39%) and there were no significant differences. However, if the same comparison was made for each of the three test domains, significant differences between faculty V on one hand, and faculties Q, S and T on the other became apparent. Students of faculty V obtained a mean C-I score which was 5.03-7.50% higher than scores of students of faculties Q, S and T in the domain of clinical sciences (Fig. 1C) and 12.45-18.63% higher in the domain of public health/behavioural sciences (Fig. 1D). In the domain of basic sciences, however, students of faculty V had a mean C-I score that was 9.35-11.67% lower than that of students of faculties Q, S and T (Fig. 1B).

### Comparison of cross-section kinetics of knowledge acquisition

In this paper, kinetics of knowledge acquisition are defined by the differences in class mean C-I scores on the MPT or its domains for a series of four to six consecutive classes. The mean scores on the test as a whole and on each of the three domains for each cohort of students of each faculty are given in Table 2. From a review of the plotted mean scores with the 95% confidence limits, significant differences emerged most consistently between

Table 2 Means and standard deviations of the percentual correct-minus-incorrect scores over the total test and per cluster

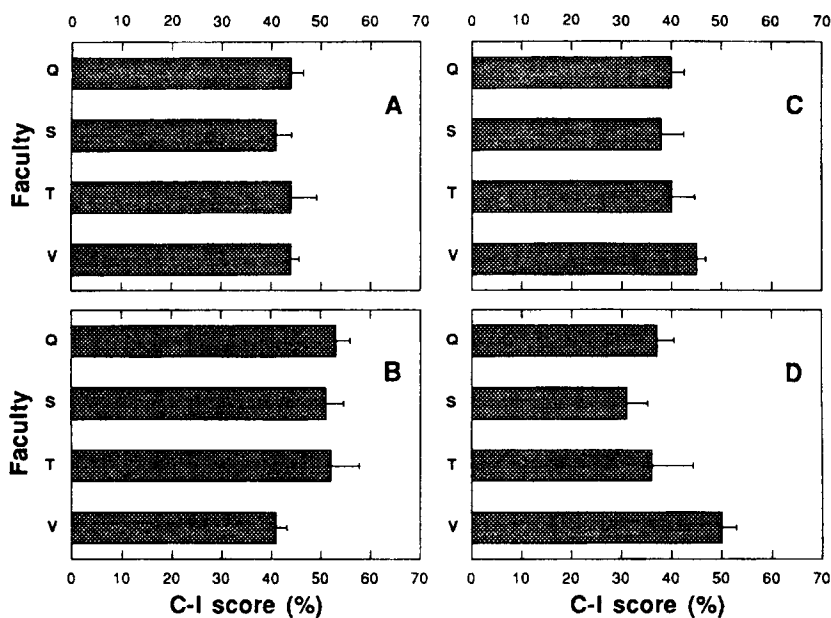
Year of training (Y)	Domain*	K <sup>†</sup>	Faculty [months in training + (Y-1) × 12]																	
			Q [4]	R [6]		S [7]		T [8]		U [9]		V [10]								
			Mean	SD <sup>‡</sup>	No. <sup>§</sup>	Mean	SD	No.	Mean	SD	No.	Mean	SD	No.	Mean	SD	No.			
1	A	198	4.01	3.19	48	4.07	3.79	82	8.97	5.72	85	8.25	3.63	128	10.79	5.79	30	11.79	4.60	164
	B	72	3.47	5.20	-	5.47	6.98	-	13.97	8.67	-	14.42	7.35	-	14.44	9.85	-	5.67	7.44	-
	C	99	2.40	3.71	-	1.84	3.97	-	6.00	6.74	-	2.70	3.56	-	7.31	5.49	-	4.77	4.48	-
	D	27	11.34	10.37	-	8.54	12.21	-	6.54	12.77	-	12.12	11.11	-	13.83	9.17	-	27.21	13.04	-
2	A	198	-	-	0	11.29	4.93	82	8.58	5.11	84	8.87	5.38	174	15.05	7.70	29	17.73	6.20	140
	B	72	-	-	-	22.36	9.17	-	16.04	9.10	-	15.71	10.24	-	19.35	14.46	-	21.81	10.26	-
	C	99	-	-	-	5.49	5.73	-	4.74	5.55	-	4.54	5.39	-	12.85	6.01	-	10.32	5.14	-
	D	27	-	-	-	3.03	11.94	-	2.78	13.33	-	6.51	12.76	-	11.62	14.55	-	34.02	13.92	-
3	A	198	-	-	0	21.93	7.55	92	14.30	5.74	86	17.68	5.56	5	29.80	8.98	24	23.54	7.09	115
	B	72	-	-	-	37.56	9.68	-	24.82	10.13	-	28.06	12.40	-	38.02	14.15	-	25.45	10.50	-
	C	99	-	-	-	13.90	7.73	-	8.57	7.14	-	12.32	7.83	-	25.84	8.91	-	18.28	7.66	-
	D	27	-	-	-	9.66	12.67	-	7.11	14.84	-	9.63	3.78	-	22.38	14.08	-	37.75	14.47	-
4	A	198	28.75	8.78	12	29.86	9.08	54	35.47	9.72	57	-	-	0	41.37	10.28	21	32.99	9.30	155
	B	72	38.19	12.97	-	43.13	11.16	-	45.61	11.52	-	-	-	-	49.21	14.96	-	32.44	12.47	-
	C	99	20.28	7.85	-	21.38	10.27	-	32.69	10.36	-	-	-	-	38.38	8.84	-	31.09	10.89	-
	D	27	34.57	16.72	-	25.58	10.04	-	18.58	16.08	-	-	-	-	31.39	20.37	-	41.46	14.68	-
5	A	198	-	-	0	56.87	6.17	27	45.82	9.23	39	34.57	10.34	9	46.97	13.60	3	39.84	9.30	117
	B	72	-	-	-	64.10	4.81	-	55.24	12.04	-	47.53	9.82	-	53.70	21.07	-	39.53	12.08	-
	C	99	-	-	-	52.75	7.66	-	41.85	9.86	-	32.66	13.14	-	40.74	12.79	-	38.27	10.27	-
	D	27	-	-	-	52.68	8.57	-	35.23	16.24	-	7.00	13.80	-	51.85	5.24	-	46.44	14.79	-
6	A	198	44.39	13.42	121	30.56	8.33	2	41.47	9.26	35	43.98	12.67	23	-	-	0	44.36	9.72	120
	B	72	52.95	16.04	-	28.48	2.09	-	50.63	10.66	-	52.42	13.98	-	-	-	-	41.28	11.94	-
	C	99	40.07	14.37	-	34.85	13.63	-	37.60	13.59	-	39.96	11.38	-	-	-	-	45.10	10.69	-
	D	27	37.40	19.19	-	20.37	5.56	-	31.22	12.44	-	36.23	20.25	-	-	-	-	49.85	16.32	-

\* (A) overall test scores; (B) basic sciences; (C) clinical sciences; (D) public health/behavioural sciences.

† K, number of test items.

‡ SD, one standard deviation.

§ No., number of participants.



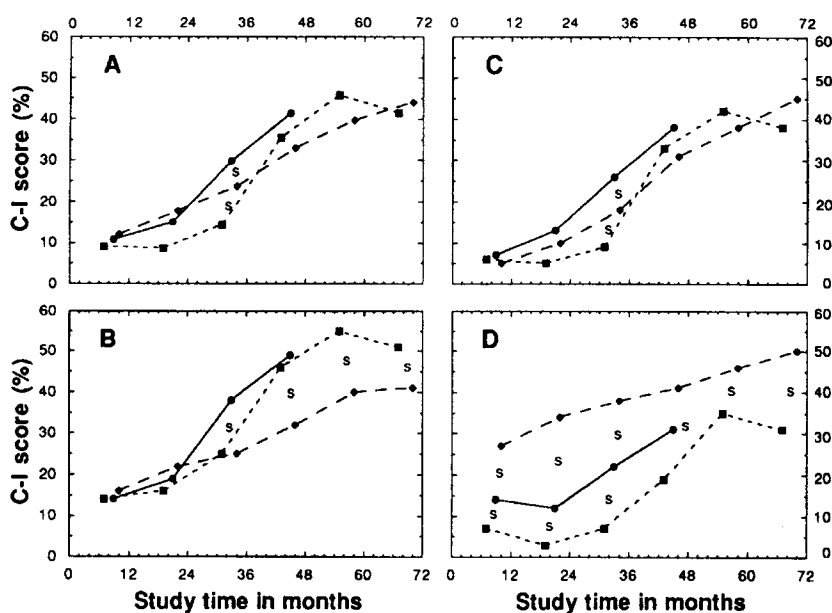
**Figure 1** Mean per cent correct-minus-incorrect (C-I) scores of sixth-year students of faculties Q, S, T and V: (A) scores on the full test; (B) scores on the basic sciences domain; (C) scores on the clinical sciences domain; and (D) scores on the public health/behavioural sciences domain. Scores are not corrected for differences in total study time. Error bars represent upper 95% confidence limits.

faculties S and U, S and V, and U and V. All plots for these three faculties are presented in Fig. 2.

First-year students of faculties S, U and V showed no significant differences in initial knowledge on the test as a whole (Fig. 2A). With respect to the kinetics of overall knowledge acquisition, three patterns emerged. Knowledge acquisition of students of faculty V seemed to proceed almost linearly over the whole period of study. Knowledge acquisition in faculty U developed parallel to V from year 1 to year 2, but showed a significant boost after year 2. Students of faculty S showed lit-

tle gain in overall knowledge from years 1 to 2, an observation also made for faculty T (Table 2). For faculty S, this lag in knowledge acquisition is apparently compensated for in years 3-5, after which no further gain in knowledge was detected. However, significant differences between faculties S, U and V were only detected for third year students.

In the basic sciences domain, marked increases in knowledge were noted for students of faculty S from years 3 to 4, and for students of faculty U from years 2 to 3 (Fig. 2B). Significant differences between mean scores of



**Figure 2** Cross-section kinetics of mean per cent correct-minus-incorrect (C-I) scores of students of faculties S (■), U (●) and V (◆): (A) scores on the full test; (B) scores on the basic sciences domain; (C) scores on the clinical sciences domain; and (D) scores on the public health/behavioural sciences domain. An 'S' between points indicates statistical significance at the level  $P < 0.05$ .

students of faculties U and V were found in years 3 and 4, and between faculties U and S in years 4, 5 and 6.

A marked knowledge boost was also detected in the clinical sciences domain for students of faculty S from years 3 to 4. In year 3, the mean score of students of faculty S was still significantly lower than those of faculties U and V; the mean scores of faculties U and V were also mutually significantly different (Fig. 2C).

The most consistent differences were observed in the public health/behavioural sciences domain (Fig. 2D). It is noteworthy that the mean scores of the students of these three faculties were already significantly different in year 1. The mean score of faculty U in year 4 was *not* significantly different from those of faculties S and V, respectively; all other differences were statistically significant.

## DISCUSSION

In this study, knowledge levels of sixth-year students could be compared for students from four out of the six participating faculties. Students' mean scores on the full test were surprisingly similar. On analysis of the scores on the basic, clinical and public health/behavioural sciences domains, however, the mean score of students of faculty V on the basic sciences domain turned out to be significantly lower than that of the other faculties, which is compensated for by the higher scores of faculty V students on the clinical and public health/behavioural sciences domains. The validity of the latter observation will be discussed below. Another peculiar finding is that the final-year students of faculties Q, S and T scored significantly better on the basic sciences domain than on the clinical sciences domain. However, it should be kept in mind that the circumstances under which the students participated in the MPT varied (see Table 1). Furthermore, on interpreting the data, one should also consider that faculty V students were experienced in handling this test format, whereas students from other faculties encountered this format for the first time. Nevertheless, the similarity of the scores of final-year students on the MPT suggest this test potentially to be useful as part of an international medical licensure exam. In that case, complimentary assessment formats like the Objective Standard Clinical Examinations (OSCEs) should be added (Feletti *et al.* 1983; Van der Vleuten & Verwijnen 1990).

The cross-sectional data obtained in this study indicate that the kinetics of knowledge acquisition of students of medical faculties in Europe may be quite different. In general, these differences are most likely explained by pluriformity of medical curricula and/or the associated examination systems. The significant increase in knowledge on basic sciences for students of faculty U (as compared to faculties S and V) from years 2 to 3, for example,

may be explained by a national licensure examination on basic sciences held by the end of year 3. The marked increase in knowledge observed for students of faculty S in basic and especially clinical sciences from years 3 to 4 coincides with the start of their intense training in clinical sciences in year 4. However, the accompanying boost in knowledge in the basic sciences of faculty S students is not explained in that way. An alternative explanation could be that, at this faculty (as in faculties R and T), there is a strong selection of students during their studies (see Table 1), which may yield an increase of the mean scores on the test as a whole in the higher classes.

The validity of the significant differences between the mean scores of students from faculties S, U and V on the public health/behavioural sciences domain may be low. The fact that significant differences already existed for first-year students may strongly indicate that part of these items were formulated in a specific cultural context unfamiliar to students from other countries. However, the influence of the scores on these items on the score on the full test is rather limited as this domain contains only 13% of the items of the full test.

The data presented in this paper do not yield evidence that curricula which are quite different with respect to teaching methodology (e.g. integrated, problem-based versus disciplinary, lecture-based) yield different overall knowledge levels for the final-year students. In Dutch national studies like the one described in the present paper, Bender *et al.* (1984) and Verwijnen *et al.* (1990) found comparable graduate knowledge levels and even knowledge acquisition kinetics for students from MFM and other Dutch medical faculties. At that time, all latter faculties had discipline-orientated, lecture-based curricula that were strongly different from the integrated, problem-based curriculum at MFM. Both types of curricula were also often found to yield comparable knowledge levels in graduates in other countries (Berkson 1993).

In conclusion, for three medical faculties from three European countries, different trends in the kinetics of students' knowledge acquisition over the full curriculum were observed. Consequently, progress testing may not be a suitable solution for the student assessment problems encountered in international student exchange. On the other hand, cross-section knowledge acquisition graphs, as presented in Fig. 2, could be used to identify corresponding cognitive levels of students on individual domains (e.g. basic sciences) in different curricula.

## ACKNOWLEDGEMENTS

The collaboration of the Dutch Faculties of Medicine of the Erasmus University in Rotterdam and the University of Leiden, as well as the advice given by M. Verwijnen

MD of the University of Limburg, are gratefully acknowledged.

This study was supported by grant SWP.92.IOP.RuLi.003 from the 'Stimuleringsfonds internationalisering (STIR)' of The Netherlands Ministry for Education and Sciences.

## REFERENCES

- Bender W, Cohen-Schotanus J, Imbos T, Versfelt W A & Verwijnen M (1984) Medische kennis bij studenten uit verschillende medische faculteiten: van hetzelfde laken een pak? *Nederlands Tijdschrift voor Geneeskunde* **128**, 917–21.
- Berkson L (1993) Problem-based learning: Have the expectations been met? *Academic Medicine* **68** (Supplement 10), S79–S88.
- Bourgeois M (1992) *ERASMUS and mobility in medical studies*. Commission of the European Communities, Task Force for Human Resources, Education, Training, and Youth, Brussels.
- Feletti G I, Saunders N A, Smith A J *et al.* (1983) Comprehensive assessment of final-year medical student performance based on undergraduate programme objectives. *Lancet* **ii**, 34–7.
- Van der Vleuten C & Verwijnen M (1990) A system for student assessment. In *Problem-Based Learning: Perspectives from the Maastricht Experience* (ed. by C Van der Vleuten & W Wijnen), pp. 27–49. Thesis Publishers, Amsterdam.
- 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: *Innovation in Medical Education: An Evaluation of its Present Status* (ed. by Z M Nooman, H G Schmidt & E S Ezzat), pp. 40–9. Springer-Verlag, Berlin.
- World Health Organization (1992–1994) ICD10. *International statistical classification of diseases and related health problems*. WHO, Geneva.

Received 5 September 1995; editorial comments to authors 24 October 1995; accepted for publication 6 December 1995