

Measuring knowledge and clinical reasoning skills in a problem-based curriculum

Henry P A Boshuizen, Cees P M van der Vleuten, Henk G Schmidt & Maureen Machiels-Bongaerts

Department of Educational Research and Development University of Limburg, Limburg, the Netherlands

SUMMARY

The purpose of this study was to investigate the validity of the Progress Test that was specially designed for measuring the growth of knowledge and clinical reasoning skills in a problem-based medical curriculum. Scores and subscores of students from the different categories of the Progress Test were compared with their scores on a Clinical Reasoning Tests. Both the Progress Test and the Clinical Reasoning Test revealed the same pattern of increasing scores over the years, and had a high intercorrelation. Further analyses revealed that the clinical sciences subscore in the progress test explained the variations in the clinical reasoning test scores. The knowledge of the behavioural sciences subscore made a small but independent contribution. The knowledge of the biomedical sciences subscore did not have this independent effect. These outcomes are discussed in this paper from the perspective of development of medical expertise research and theory. Some educational consequences are also discussed.

Keywords

Curriculum; *education, medical, undergraduate; *educational measurement; Netherlands; *problem-based learning; thinking

INTRODUCTION

One of the reasons for promoting problem-based learning (PBL) is that it is thought to encourage self-directed learning in students, and also that self-directed learning should be preferred to teacher-directed learning. Students themselves know better than their teachers what they know and do not know, and therefore what subjects need their attention next. Another reason concerns motivation: students should be allowed to pursue subjects they are interested in, at that particular time. Intrinsic motivation is thought to determine the time and effort put into studying and later, results. Consequently, a feature of many problem-based curricula is that students are responsible for their own learning. Teachers or book lists do not prescribe what students have to learn during a specific period; the students themselves decide what they will study. Their decisions are aided by the problems they are working on. While working on problems (that have been carefully designed by the teaching staff in order to be able to fulfil this role) they analyse what they know about the issues involved, and what they apparently do not know. The students also have to decide which level of mastery and which level of detail they want to attain at that point. Finally, students will also choose the media they want to use for learning. Possibilities are traditional books, audiovisuals, computer simulations, interviews with an expert, field work, etc. The learning objectives pursued by the individual students will be similar in many respects, but will also differ as a consequence of differences in prior knowledge and interest. This relative freedom of the students makes it very difficult for the staff to formulate rigorous course objectives.

The final consequence of this theory of teaching and learning is that examination cannot be based on a test that is designed as a traditional end-of-course test. The end-of-course test may not do justice to the individual learner and may force the students to direct themselves exclusively to the (expected) contents of the test. As an alternative the Medical School of the University of Limburg developed the Maastricht Progress Test. This Progress Test is designed as an exit level test: the cognitive curriculum objectives are translated into true–false items. Together these items should cover all medical areas a graduate is supposed to know. For each test a sample of approximately 250 items is taken. An example of a single item is:

(given) A patient with a renal disorder shows metabolic acidosis.

(question) Hypoventilation contributes to compensating for this acidosis.

~~TRUE~~/FALSE.

Correspondence: Dr Henny P A Boshuizen, University of Limburg, Department of Educational Research and Development, PO Box 616, 6200 MD Maastricht, the Netherlands

Every student, from first year to near-graduate, takes the test four times a year. Students are allowed to skip those items they have no knowledge of; guessing is not really discouraged, although students sometimes feel that the correction made for guessing is meant for that purpose. Grading is based on the number of items answered correctly minus the number of questions answered incorrectly. Per-test passing scores are calculated for each class. Passing scores set for freshmen are of course much lower than for final year students (Verwijnen *et al.* 1982). Many years of experience with the Progress Test show that the scores of the students continuously increase over the years. The educational advantage of the progress test has proven itself over the years in that it does not encourage cramming a few days before the test, which is often followed by a very rapid process of forgetting (Semb & Ellis 1994).

The Progress Test is used not only as an assessment instrument, but also as a means for feedback. Students receive detailed reports including their total score, scores on the biomedical, clinical and behavioural sciences in general and per subject. Students also receive reports on individual items and references to literature that can be checked. On the whole, the Progress Test has been shown to be a valuable instrument for assessment and feedback. Other medical faculties (e.g. McMaster University) and other professional schools (e.g. Health Sciences at the University of Limburg) have adopted the principle as part of their assessment procedures.

A potential validity problem with the Progress Test is that there is a mismatch between the kind of questions generally asked and the aim of the PBL programme. Most Progress Test questions address only the factual knowledge level. Problem-solving and clinical reasoning items are rare, while only a few questions per Progress Test can be classified as knowledge application questions pertaining to short cases. Nevertheless, the latter kind of questions would better reflect the educational goals. Furthermore, fixed response questions with few alternatives (two in this case) allow the students to rely on recognition or to reason back from the set alternatives. By doing so, students can circumvent active hypothesis generation and the necessity of forward search, i.e. working from the information provided in the case. Both strategies may lead to a good answer and students will apply them in case of uncertainty or incomplete knowledge on the subject. There is some evidence that knowledge-oriented multiple-choice tests do have predictive value for more authentic problem-solving measures (Norcini *et al.* 1985) and even for later practice performance (Norman 1991). There is also evidence suggesting that the cueing effect of multiple-choice questions works in multiple directions, i.e. cueing towards and away from the correct answer (Schuwirth *et al.* 1996). However, despite the correlations, fixed response tests do not mirror the way knowledge is applied in actual practice where hypothesis generation plays an important role.

In the present article we explore the validity of the Progress Test by comparing student scores on this instrument with scores on a Clinical Reasoning Test (Schmidt *et al.* 1996). This test consists of 30 case vignettes using an open-ended format, hence it requires active hypothesis generation as a necessary step towards the differential diagnosis that is asked for (*see* Appendix 1). The question that is the main focus of the present paper is whether the same performance patterns can be found in the Progress Test and in the Clinical Reasoning Test. Therefore the study investigates whether our students' diagnostic problem solving follows the same development path as does their knowledge measured by the Progress Test. The study also investigates how Progress Test scores and Clinical Reasoning scores are related during consecutive years.

METHODS

Data were collected in October 1993. About 40 students per year (the whole curriculum takes 6 years, 4 preclinical and 2 clinical years; graduations occur during the whole year, as soon as the student has fulfilled all requirements) were invited to participate in the study. Freshmen who had started only 6 weeks before were not included, resulting in three preclinical groups: second, third and fourth year students. In the clinical period an extra criterion was used, i.e. the number of clerkships completed. This was done because some students have to wait several months before they can begin the clinical rotations. Ro-1 students were fifth-year students who had recently started their first clerkship. Ro-2 students were sixth year students. They had recently completed their third clerkship, which could be either internal medicine, surgery or family medicine (the other two having been completed previously). These three large clerkships each take about 3 months. Ro-3 students were about to graduate or had graduated very recently. Administratively these students are also sixth-year students; they are only slightly behind. In fact, they are about 1 year ahead of the Ro-2s. Delays can be due to waiting times, extracurricular activities undertaken, etc. Normally these students do not have higher or lower marks than those who graduate before September 1. 223 students participated: 40 second-year students; 41 third-year students; 41 fourth-year students; 40 Ro-1s; 20 Ro-2s; and 41 Ro-3 students.

These students took the Clinical Reasoning Test. This test consisted of 30 vignettes with known diagnosis, cov-

ering all organ systems with the exception of psychiatry. Psychiatry was not included because psychiatric case descriptions are incompatible with the proposed length of the vignettes. Students were asked to read the cases and to come up with a differential diagnosis. Explanations or justifications were not asked for. The differential diagnoses were scored as follows: if the intended diagnosis was in the first place of the list, 2 points were given; if it was in any other place than the first, 1 point was credited. Four cases yielded diagnoses with one or more subdiagnoses (e.g. case 20 was an acute pancreatitis case with subdiagnoses of gallstones and obstruction of the bile flow). Students who included some subdiagnoses received bonus points (maximum of 7). Four cases were excluded because subject experts were of the opinion that alternative diagnoses were too plausible. The reliability across year groups was 0·91 (Cronbach's alpha) and a median reliability of 0·63 was found within the year groups.

For the students who participated in the study, percentage correct minus incorrect scores on the Progress Tests of September and December 1993 were obtained. The first test preceded the period in which data were collected, the second one followed it. The students' scores are found by taking the number of items answered correctly, corrected for guessing by subtracting the number of items answered incorrectly. Reliabilities (coefficient alpha) calculated over all students were 0·92 and 0·90, median reliabilities per year group were 0·66 and 0·65. Some students had taken only one test. In that case the score on one test was used. Those students who had not participated in both tests were excluded from the analyses. (Except for illness, the most common reason for not sitting the Progress Test is having obligations elsewhere that do not allow travel to Maastricht, e.g. because the student is abroad for electives. Most often this occurs at the end of the fourth year, after a student has finished the preclinical programme. Another reason for non-participation is found in the sixth-year students group. Students who have taken 24 tests and have obtained enough passing scores are no longer required to take the test.) As a result the samples consisted of 39 (40 originally) second-year students, 41 (41) third-year students, 31 (41) fourth-year students, 39 (40) Ro-1s, 20 (20) Ro-2s, and 25 (41) Ro-3 students. Students received one small financial remuneration for their participation.

Differences between group means were tested using analysis of variance. The associations between the scores on the different measures were analysed bivariate and multivariate. In the bivariate analysis, Pearson correlations were used. These observed correlations were statistically corrected for unreliability (attenuation correction) in each of the measures and true correlations were obtained. True correlations represent hypothetical associations when all instruments used yield perfectly reliable scores. The multivariate analysis consisted of multiple regression analysis with the observed correlations as input. The criterion variable used was the Clinical Reasoning score; the predictor variables were group (number of years in the curriculum) and the three subtest scores of the Progress Test: biomedical, clinical and behavioural sciences. The strength of the associations is expressed in the R^2 (R squared), representing the squared multiple correlation or the variance explained in the criterion variable by the predictor variables. A stepwise regression procedure was used. This means that predictors are entered one by one in the analysis, using statistical significance as entry criterion.

RESULTS AND DISCUSSION

Figure 1 shows the Progress Test results of the six groups. Groups differ significantly ($F(5, 189) = 77{\cdot}191$, $P < 0{\cdot}0001$); all Newman–Keuls comparisons except those between Ro-2 and Ro-3 are significant. The groups show the same increase over the years that is commonly found on this test. Comparisons of the group means and the population means that are depicted in the same graph (placed with parentheses) suggest that the subjects selected for this study do not deviate dramatically from their peers. Differences between sample mean and population mean are never greater than 1·9. Notice that the Ro-2 and -3 group are compared with the same population value (39·32), the mean score of the sixth year students.

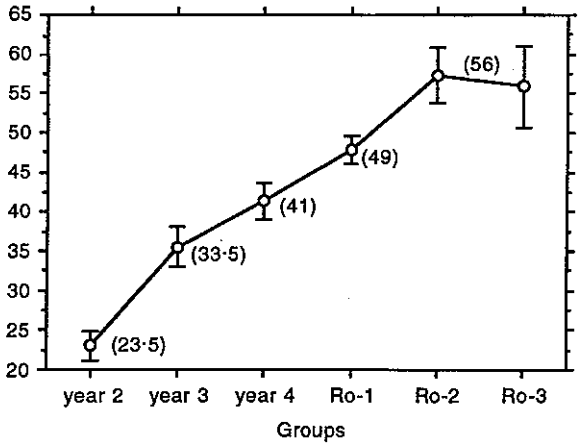

Figure 1 Mean percentage score on the two Progress Tests with 95% confidence error bars. Mean scores of the populations are placed between parentheses.

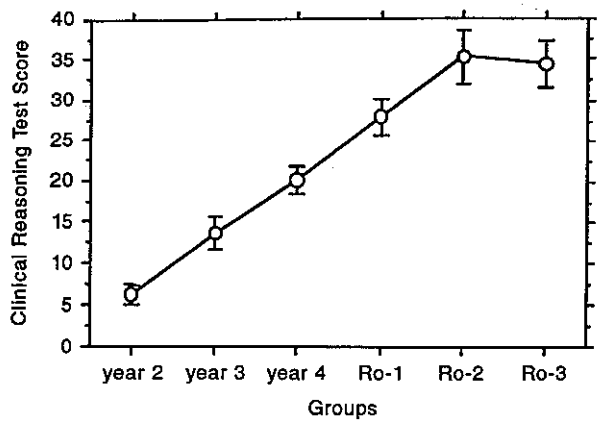

Figure 2 Mean clinical reasoning scores for the five groups with 95% confidence error bars.

Figure 2 shows the results of the same students on the Clinical Reasoning Test. Again groups differ significantly ($F(5, 189) = 119{\cdot}802$, $P < 0{\cdot}0001$). The same pattern in the Newman–Keuls comparisons were found. The curves of the Clinical Reasoning Test and the Progress Test have basically the same shape, which is expressed in an observed correlation of 0·85 between the scores on both tests (72% common variance). Using the mean value of Progress Test reliabilities on both test occasions, the (true) correlation corrected for unreliability is 0·93. Calculated at the group level this correlation drops dramatically, ranging from 0·30 (true correlation 0·46) for the second-year students, to 0·61 for the third-year group (true correlation 0·97). Mean correlation within groups is 0·49 (true correlation 0·77).

Such a discrepancy between correlations in the whole group and in subgroups should be expected as an effect of restriction of range. It might, however, also indicate that although both tests have a large common basis, they measure partly different constructs, i.e. other factors than pure knowledge (as measured by the Progress Test) play a role in clinical reasoning – at least at different stages of development. In order to investigate this explanation the multiple regression analyses per group and for total students were used. The correlations for total students are shown in Table 1. All observed correlations are high, except for those with the behavioural sciences subscores. The true correlations indicate a completely liner relation between biomedical and clinical knowledge. The pattern of true correlations between Clinical Reasoning and the other variables remains unaffected; the correlation with the clinical science subscore is now unity.

A similar picture emerges at the group level (Fig. 3). Correlations per group are much lower, again the correlations between the Clinical Reasoning Test score and the behavioural science subscore are lowest. Furthermore, correlations are remarkably low in the second year students group, as compared with the more knowledgeable groups. This might be an effect of lack of relevant knowledge: a real restriction of range. In the later years knowledge seems to have grown enough to explain at least a large part of the variance in clinical reasoning. The estimated true scores on the subtests show the same pattern of relations.

The results of the stepwise-regression analyses per group are summarised in Table 2, which shows that the amount of variance in the Clinical Reasoning Test scores that can be explained by differences in performance on the three subtests of the Progress Test increases over the years (from 0·742 in the second year to 0·974 in the final year). Another remarkable finding is that either one or two subscores in the Progress Test are needed to explain this variation. However, no clear pattern is discernible in these outcomes. In the second year it is the biomedical science subscore that explains the outcomes, in the third year we see a combination of clinical sciences and behavioural sciences subscores as explaining factors. In the fourth year it is again a combination, but now the biomedical and behavioural subscores. In the Ro-1 and -2 groups we see a single effect of the clinical sciences subscore, while in the Ro-3 group the variation in the Clinical Reasoning Test scores is explained by a joint

Table 1 Reliabilities, true and observed correlations between Clinical Reasoning Group and Progress Test scores ($N = 195$)

	Clinical Reasoning	Group	Progress test total	Subscore biomedical	Subscore clinical	Subscore behavioural
Clinical Reasoning	*0·91*	0·90	0·93	0·92	1	0·59
Group	0·86	–	0·85	0.81	0·96	0·47
Progress test total	0·85	0·81	*0·91*	1	1	0·73
Subscore Biomedical	0·77	0·71	0·91	*0·77*	1	0·61
Subscore Clinical	0·90	0·89	0·93	0·83	*0·86*	0·57
Subscore Behavioural	0·43	0·36	0·53	0·41	0·40	*0·58*

Reliabilities in diagonal entries (italics), observed correlations in lower triangle and true correlations in upper triangle.

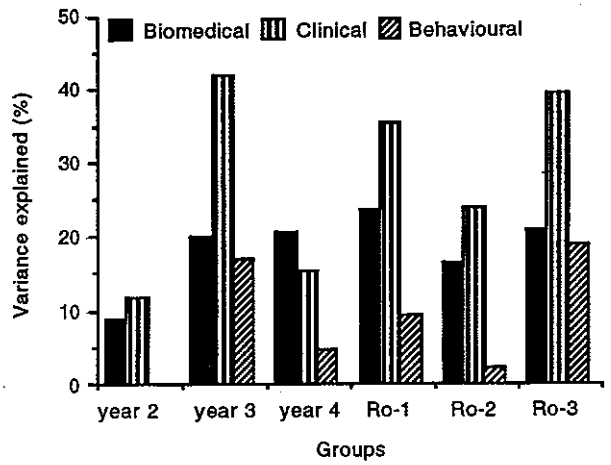

Figure 3 Variance explained (%) in the Clinical Reasoning scores by the biomedical, clinical and behavioural sciences scores on the progress test.

effect of the clinical sciences and behavioural sciences subscores.

The main aim of the present study was to explore the validity of the Progress Test. Therefore a stepwise regression analysis of all the data using the Total Progress Test score and group as predictors for clinical reasoning was carried out (Table 3). It shows that both factors have unique, almost equal contributions, together explaining 80·9% of the variance. This may indicate that the Progress Test and the Clinical Reasoning Test have less in common than is suggested by the high overall correlation. The finding of the two equal components may cast doubt on the validity of the Progress Test. However, further analysis amends this conclusion. In this analysis again the three subscores instead of the total Progress Test score were used (Table 4). A major part of the variance in the Clinical Reasoning Test outcomes (95·6%) can thus be explained using three factors: clinical sciences subscore; group; and again, despite the low correlations, the behavioural sciences subscore. The basic sciences subscore partial correlation was not higher than 0·062.

This final analysis conveys the impression that the Clinical Reasoning Test draws heavily on the clinical sciences component in the Progress Test. However, the mere fact of going through the curriculum and experiencing medical practice in the clinical rotations itself makes a unique, but relatively small, contribution to the variance in clinical reasoning. A still smaller contribution comes from behavioural sciences knowledge. Hence we may conclude that knowledge of the basic sciences and of the behavioural sciences seem to contribute differently to clinical reasoning. Biomedical knowledge is probably integrated with clinical knowledge as was described by Schmidt & Boshuizen (1993). These authors assert that without the integration of basic sciences knowledge in clinical science knowledge, the

Table 2 Step-wise regression analysis for the separate groups with the Clinical Reasoning Test score as dependent variable and the three Progress Test subscores as independent

Group	Independent variable(s)*	*F*	df	*P*	Adjusted R^2
Second year	Biomedical Science	119·179	1·38	< 0·001	0·742
Third year	Clinical Science	26·086	1·39	< 0·01	0·895
	Behavioural Science	7·644	1·39	< 0·01	
Fourth year	Biomedical Science	20·607	1·29	< 0·01	0·952
	Behavioural Science	20·298	1·29	< 0·01	
Ro-1	Clinical Science	997·115	1·39	< 0·001	0·961
Ro-2	Clinical Science	551·262	1·19	< 0·001	0·965
Ro-3	Clinical Science	34·065	1·23	< 0·01	0·974
	Behavioural Science	6·312	1·23	< 0·05	

*Only significant contributions are reported.

Table 3 Step-wise regression analysis for the total group with the Clinical Reasoning Test score as dependent variable and the Progress Test total score and groups as independent variables

Group	Independent variable(s)*	*F*	df	*P*	Adjusted R^2
Total	Group	79·0499	1·192	< 0·001	0·8092
	PT (total score)	77·4584	1·192	< 0·001	

*Only significant contributions are reported.

Group	Independent variable(s)*	F	df	P	Adjusted R^2
Total	Clinical Science	123·907	1·191	< 0·001	0·956
	Group	18·178	1·191	< 0·01	
	Behaviour Science	5·539	1·191	< 0·025	

*Only significant contributions are reported.

Table 4 Step-wise regression analysis for the total group with the Clinical Reasoning Test score as dependent variable and the three Progress Test subscores and groups as independent variables

basic science knowledge cannot be applied flexibly in clinical settings. The same was expected for the behavioural sciences knowledge. The results, however, suggest that behavioural sciences knowledge is less integrated, and can hence play a role of its own. The latter finding is in line with the findings by Hobus *et al.* (1987) who found that behavioural sciences knowledge was not yet integrated with the clinical knowledge of doctors who had recently graduated, but was integrated with the clinical knowledge of more experienced doctors. Later research by Boshuizen *et al.* (1995) suggests that sixth-year students may have behavioural science knowledge, but do not apply it in clinical reasoning. The family doctors in that study appeared to have integrated this kind of knowledge with clinical knowledge. Behavioural science knowledge may take more time to prove its relevance in clinical settings.

The educational conclusions that can be drawn from this study depend largely on faculty policy. The outcomes of this study suggest that the Progress Test is a very valuable instrument for monitoring the students' advancement. Despite the format and the kinds of questions asked, it correlates very well with the Clinical Reasoning Test scores. This especially applies to the clinical sciences subscore of the Progress Test, suggesting that the true–false format does not access merely pure factual knowledge, but also addresses clinical reasoning. Depending on the school's aim and policy, the kind of questions asked might however be reconsidered. Provided the medical school accepts the criterion used in this study, i.e. the Clinical Reasoning Test, it might also re-evaluate the structure of the Progress Test. It might reduce the allotment of biomedical science items in the test in favour of the clinical science items, and hence reduce their influence on the total progress test score. Assigning different weights on these areas, perhaps even differing per year (group), might serve the same purpose.

ACKNOWLEDGEMENTS

The authors wish to thank Ron Hoogeboom, who provided the Progress Test score sets. Hélène Hermans, who constructed the Clinical Reasoning Test, Eline Tiems and Ingrid Hertogs, who analysed the data, and Dr Penny Hansen, who replaced our Dutch constructions by real English. Part of this paper was presented at the Annual Meeting of AERA, San Francisco, April 1995.

REFERENCES

Boshuizen H P A, Schmidt H G, Custers E J F M & Van de Wiel M W (1995) Knowledge development and restructuring in the domain of medicine: The role of theory and practice. *Learning and Instruction* 5, 269–89.

Hobus P P M, Schmidt H G, Boshuizen H P A & Patel V L (1987) Contextual factors in the activation of first hypotheses: expert-novice differences. *Medical Education* 21, 471–6.

Norcini J J, Swanson D B, Gross L J, Shea J A & Webster G D (1985) Reliability, validity and efficiency of multiple choice question and patient management problem item formats in the assessment of physician competence. *Medical Education* 19, 238–47.

Norman G R (1991) Can an examination predict competence? The role of recertification in maintenance of competence. *Annals of the Royal College of Physicians and Surgeons of Canada* 24, 121–4.

Schmidt H G & Boshuizen H P A (1993) On acquiring expertise in medicine. *Educational Psychology Review* 5, 1–17.

Schmidt H G, Machiels-Bongaerts M, Hermans H *et al.* (1996). The development of diagnostic competence: a comparison between a problem-based, an integrated and a conventional medical curriculum. *Academic Medicine* 71, 658–64.

Schuwirth L W T, Vleuten C P M van der & Donkers H H L M (1996) A closer look at cueing effects in multiple-choice questions. *Medical Education* 30, 44–9.

Semb G B & Ellis J A (1994) Knowledge taught in school: What is remembered? *Review of Educational Research* 64, 253–86.

Verwijnen G M, Imbos T J, Snellen H A M *et al.* (1982) The evaluation system of the medical school of Maastricht. *Assessment and evaluation in higher education* 3, 225–44.

Received 20 *February* 1996; *editorial comments to authors* 6 *September* 1996; *accepted for publication* 24 *October* 1996

APPENDIX 1

Example of a vignette in the Clinical Reasoning Test.

A 65-year-old woman visits her family doctor. She enters your office with red eyes suggesting that she has been crying. She tells you that she worries a lot because she has been losing so much weight. After you have calmed her down, she tells you in a rush of words that she has lost 25 pounds, although she eats well. She is very worried about this state of affairs, sleeps poorly and is restless and agitated. She does not take any drugs. Her family history displays nothing unusual. Upon physical examination you find a sick, restless woman with a sweaty, warm skin. Her thyroid gland is diffusely enlarged. Blood pressure 150/89; pulse rate 104/min. irregular and unequal. Her legs show pitting edema. Her heart is enlarged and you hear a murmur suggesting mitral valve insufficiency. Lab data: T4 300 nmol/l, T3 nmol/1, TSH 0·05 mU/l ECG: atrial fibrillation accompanied by a high ventricular frequency.