Do students have sufficient knowledge of clinical anatomy?

Katinka J A H Prince,[1] Albert J A A Scherpbier,[2] Henk van Mameren,[3] Jan Drukker[3] & Cees P M van der Vleuten[4]

INTRODUCTION Comparisons of anatomy knowledge levels of students from various curricula show either no differences or small differences to the detriment of innovative schools. To pass judgement on the general level of students' anatomy knowledge, we need an absolute standard. The purpose of this study was to compare students' levels of anatomy knowledge as measured by a case-based anatomy test with standards set by different groups of experts.

METHODS A modified Angoff procedure was used to establish an absolute standard against which the students' results could be evaluated. Four panels of 9 anatomists, 7 clinicians, 9 recent graduates and 9 Year 4 students, respectively, judged 107 items of an anatomy test. The students' results on these items were compared with the standards obtained by the panels.

RESULTS If the standard established by the panel of Year 4 students was used, 64% of the students would fail the test. The standards established by the anatomists, clinicians and recent graduates would yield failure rates of 42%, 58% and 26%, respectively.

CONCLUSION According to the panels' standards, many students did not know enough about anatomy. The high expectations that the Year 4 students appeared to have of their peers may contribute to students' uncertainty about their level of anatomy knowledge.

KEYWORDS education, medical, undergraduate/ *standards; anatomy/ *education; educational measurement; students, medical/ *standards; curriculum; Netherlands.

*Medical Education 2005; **39**: 326–332*
doi:10.1111/j.1365-2929.2005.02096.x

[1]Skillslab, Faculty of Medicine, University of Maastricht, Maastricht, The Netherlands
[2]Institute for Medical Education, University of Maastricht, Maastricht, The Netherlands
[3]Department of Anatomy/Embryology, University of Maastricht, Maastricht, The Netherlands
[4]Department of Educational Development and Research, University of Maastricht, Maastricht, The Netherlands

Correspondence: Katinka J A H Prince, Skillslab, University of Maastricht, PO Box 616, 6200 MD Maastricht, The Netherlands.
Tel: 00 31 43 388 1978; Fax 00 31 43 388 4127;
E-mail: K.Prince@sk.unimaas.nl

INTRODUCTION

In every medical curriculum educators have to strike an optimal balance between basic sciences and clinical sciences. Innovative educational approaches, such as problem-based learning (PBL), are aimed at integrating basic and clinical sciences. However, concerns have been raised about the level of basic science knowledge attained by students in innovative curricula. Studies on the outcomes of PBL have shown contradictory results, sometimes indicating that PBL schools equip their students with less basic science knowledge than do more traditional medical schools.[1–4]

In general, the integrated PBL approach has seemed to be associated with uncertainty among students about their basic sciences knowledge as well as alleged deficiencies in this knowledge, particularly in clinical anatomy.[5,6] A comparative study among students at Dutch medical schools did not support these earlier findings.[7] In other words, PBL students felt equally confident (or uncertain) about their level of anatomy knowledge as students from other curricula. The same study showed relatively small differences between the different schools in the students' actual clinical anatomy knowledge.[7] The results of the PBL students were

Overview

What is already known on this subject

There is either no or only a small difference between students from PBL and non-PBL schools in levels of anatomy knowledge. More information is needed to make an absolute judgement.

What this study adds

The anatomy knowledge of many students falls short of the standards set by different groups of judges. The high expectations regarding anatomy knowledge reflected by the standards of the Year 4 students might be among the causes of their uncertainty regarding their level of anatomy knowledge.

Suggestions for future research

Further research should aim to clarify whether students' anatomy knowledge is indeed under par, or whether judges overestimate the amount of knowledge students should have acquired.

comparable to those of their colleagues at other medical schools.

In other words, the evidence from comparisons of anatomy knowledge acquired by students in innovative and other curricula is ambiguous. This ambiguity may be explained by the considerable variation in curricular designs and implementations, due to the fact that the PBL label is attached to a variety of heterogeneous educational activities. Other developments can also impact on students' anatomy knowledge, such as the reduction in the available time in the curriculum and the shortage of suitable staff to teach anatomy.[8] Whatever causes these differences may have, it still remains to be ascertained whether medical students across the curricular spectrum attain a level of anatomy knowledge that is considered adequate.

This study sought to answer the question: do medical students in the Netherlands have sufficient knowledge of anatomy at the end of Year 4 (i.e. at the end of the pre-clinical curriculum)? Having examined the differences in knowledge between students from different curricula in a previous study, we now wanted to measure students' anatomy knowledge against an absolute benchmark.[7] Although there is no gold standard in common use, there are many acceptable methods for generating an absolute standard. A popular method for standard setting is the modified Angoff procedure, which is supported by research.[9,10] The modified Angoff procedure involves the estimation of the performance of borderline examinees by a panel of judges. The judges must be knowledgeable in the content area of the examination.

We expected different panels of judges with different levels of expertise to reach different outcomes, which would be in line with the literature.[11] As content experts who are the most distant from students in terms of expertise, anatomists may have the highest expectations of students, which is why we expected the anatomists to set a relatively high standard. We assumed that clinicians would give a lower estimation because they would expect students to continue to learn about anatomy during their clerkships. Because students are content experts as consumers of the curriculum, we expected them to be more familiar with the candidates than the teachers. Verhoeven *et al.* showed that recently graduated students showed more agreement and produced more reliable Angoff estimates than a panel of staff item writers.[11] We expected recently graduated doctors to make the most realistic estimation. The estimations of Year 4 students were expected to be somewhat higher, due to their uncertainty about their own knowledge level.

In order to answer our research question, we determined whether students have sufficient knowledge of clinical anatomy by comparing the anatomy test results of a national sample of Year 4 medical students with the absolute standards set by 4 panels of different judges, consisting of anatomists, clinicians, recent graduates and Year 4 students, respectively.

METHODS

Materials

We used the results obtained by a sample of students from the 8 medical schools in the Netherlands on the same computerised, case-based, clinical anatomy test.[7] Clinical anatomy is defined as anatomy that is indispensable for a good understanding of the medical physical examination, modern imaging techniques, diagnosis and many invasive and

non-invasive procedures. The test consisted of 138 items, linked to 16 patient cases. The items comprised open questions, multiple-choice questions and true/false questions, without a do-not-know option. The clinically contextualised items required students not only to recall names of structures, but also to apply their knowledge to clinical problems. Examples of (parts of) patient cases with accompanying items are shown in the Appendix.

Because of time constraints, we used 13 cases from this test, with an accompanying 107 items. The cases were related to the domains of ENT, family medicine, neurology, surgery, orthopaedic surgery, internal medicine, cardiology, ophthalmology and radiology.

Student results

The test was administered to a sample of all students in Year 4 of the 6-year Dutch undergraduate medical curriculum shortly before the beginning of their clerkships. A total of 348 students answered all 13 cases. The overall test score was calculated as the percentage correct answers. Because the number of questions varied across cases, scores were calculated per case and then averaged.

Angoff procedure

A modified Angoff procedure was used to establish an absolute passing score for the anatomy test.[10] Prior to the standard setting procedure, the judges received a letter describing the purpose of the study and a paper copy of the anatomy test (the actual test was computerised).

Judges

We assembled 4 panels of judges for the Angoff procedure:

1 anatomists;
2 clinicians;
3 graduates, and
4 Year 4 medical students.

We aimed to have 6–10 judges per panel.[10,12]

The anatomy departments at the 8 medical schools in the Netherlands were each asked to supply judges. Nine anatomists from 7 medical schools participated. The panel of clinicians was recruited from staff at the Maastricht Academic Hospital. We invited 1 person from each of 9 specialties. Seven clinicians, representing all the domains included in the test, participated. For the graduate panel we invited recent graduates and Year 6 students at Maastricht Medical School who were close to graduation. The group of 'recent graduates' consisted of 4 graduates and 5 students who were in their final weeks before graduation. The student panel consisted of 9 Year 4 students from Maastricht Medical School.

Each panel of judges met for 1 3-hour session with the same 2 moderators (KP and AS). The judges were asked to imagine a borderline student at the end of Year 4 and estimate for each test item, taking account of item content and difficulty, the probability that such a borderline student would know/give the correct answer. Because the items did not have a do-not-know option, the judges were instructed to apply a correction for guessing. The judges were given the correct answers to the test items. After a discussion of the items and on the basis of empirical data (the actual percentages of correct answers given by the students who took the test [*P*-values]), the judges were able to revise their estimates. This procedure was repeated until the judges had given Angoff estimates for all the items.

Statistical analysis

Each judge gave 2 estimates for each test item, 1 before and 1 after the item had been discussed by the panel. The 2 estimates were used to calculate 2 separate scores. The individual judges' Angoff scores were established by averaging each judge's estimations across items per case, and then averaging these estimations across all 13 cases. The mean Angoff score per panel was arrived at by averaging the mean scores of the individual judges. Compared with the initial scores of the judges, the estimates after the panel discussion were slightly more lenient and the confidence intervals smaller. Only these scores were used for further analysis and reported in this paper.

In order to estimate the variability of these scores, generalisability was used following the procedures as described by Verhoeven *et al.*[12] A root mean square error (RMSE) was established and used to estimate a confidence interval around the mean Angoff score for each panel. The scores of the 4 panels were compared.

Because a high correlation between the Angoff estimate and the actual item score can be regarded as an indicator of validity, the Pearson correlation was calculated for the mean of the panel estimates per item and the *P*-values (percentage correct).

Table 1 Angoff scores, RMSE, confidence intervals after group discussion of the 4 panels, including the number and percentage of students that would fail given the Angoff score and the Pearson correlation of each panel based on 13 cases/107 items

Panel	Panel size (*n*)	Angoff score	RMSE	95% CI	*N* failed *n*	(%)	Correlation Angoff *P*-value
Anatomists	9	50.2	1.03	48.2–52.2	146	(42.0%)	0.80
Clinicians	7	54.3	0.76	52.8–55.8	200	(57.5%)	0.83
Graduates	9	46.9	0.88	45.2–48.6	91	(26.1%)	0.84
Year 4 students	9	56.0	0.85	54.4–57.7	223	(64.1%)	0.91

Student scores were compared with the various Angoff scores and the pass/fail ratio was calculated for each Angoff score.

RESULTS

The mean Angoff score (the mean of all judges' estimates across 13 cases after the panel discussion) was 51.7%. Table 1 presents descriptive statistics of the judges' Angoff scores.

The use of the variability index RMSE allows estimation of confidence intervals around the Angoff mean and inference of statistical significance between differences. The results show some significant differences between the standards set by the different panels. Both the anatomists and the graduates set the standard significantly lower than the Year 4 students and the clinicians.

All panels obtained correlations of 0.8 or more between item difficulty (*P*-value) and the item Angoff estimates.

The mean score across 13 cases obtained by the students sitting the test ($n = 348$) was 53.20. Figure 1 shows the distribution of the students' scores (in grey) and the 4 panels' Angoff scores after discussion (black lines).

The area to the left of the Angoff scores in Fig. 1 represents the number of students failing the test if that score were to be used as the pass/fail cut-off point. The failure rates associated with the Angoff scores of the graduates, anatomists, clinicians and Year 4 students were 26.1%, 42.0%, 57.5% and 64.1%, respectively.

DISCUSSION

This study was performed to set an absolute standard for an anatomy test to ascertain whether or not medical students have an adequate knowledge of clinical anatomy.

The various Angoff estimates yielded percentages ranging from 26% to 64% of students failing to

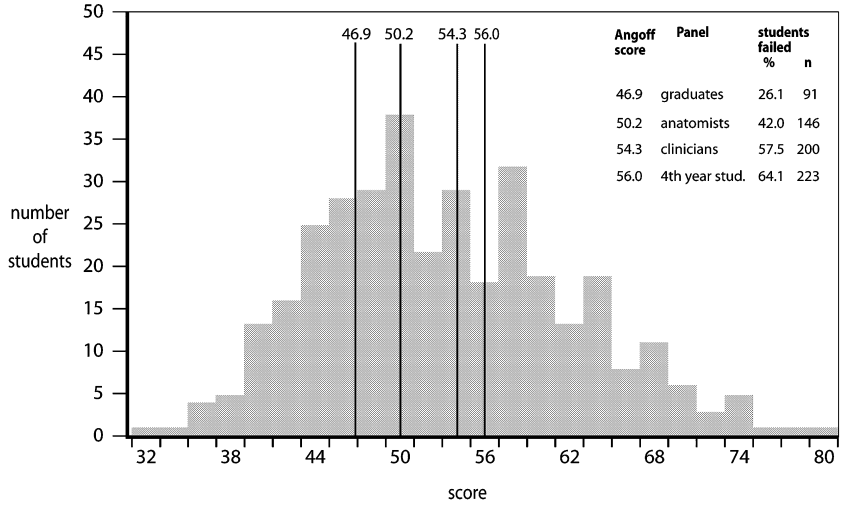

Figure 1 The distribution of the students' test results (shaded) and the Angoff scores of the 4 panels of judges.

demonstrate the required knowledge level. In short, every standard that emerged from the Angoff procedure would result in a substantial number of students failing to perform to standard, although the number of students failing varied considerably depending on the type of judges in the Angoff panel.

Although the absolute differences between the panels' Angoff scores were small, the different scores led to substantial differences in failure rates. This strong impact on the pass/fail ratio is caused by the fact that the Angoff scores of the 4 panels were all in the centre of the distribution of the students' test results. As a consequence small differences in standards led to dramatic differences in outcomes.

There are several possible explanations for the diversity of the Angoff scores. We expected the anatomists to set a high standard based on their reported dissatisfaction with students' anatomy knowledge. However, the results showed that the anatomists set the standard significantly lower than the clinicians and the Year 4 students. Possibly, the anatomists' more extensive contacts with students in the pre-clinical years afforded them a better insight into the students' actual knowledge levels. However, this does not necessarily imply that they were satisfied with the students' knowledge levels. If we had asked them to indicate students' desired rather than actual knowledge levels, we might well have found considerably higher estimates. Clinicians, on the other hand, have little contact with pre-clinical students and thus may be unable to give a good estimate of what students actually know. As expected, the graduates set the lowest standard. A possible explanation is that the judges in the panel of graduates had recently or were about to finish their clerkships and thus had a good idea of the level of knowledge required in practice. Secondly, the graduates were aware of how much they had learned about anatomy in the clerkships and therefore were likely to expect less of Year 4 students.[13] One might expect Year 4 students to have an accurate picture of their peers' knowledge. However, this assumption was not confirmed by the results: Year 4 students set the highest standard, resulting in a 64% failure rate. Apparently, they overestimated the achievements of their peers, which might offer an explanation for their uncertainty about their own knowledge.

Apart from the difficulties that arise when we want to appraise students' anatomy knowledge, these results also show that there is no consensus on what students need to know. This alone can be a cause of uncertainty among students. Clear guidelines as to the required anatomy knowledge should be developed and agreed on by anatomists and clinicians, with students and graduates contributing to the process.

The differences between the Angoff scores also show the limitations of this Angoff method as well as the importance of the selection of the judges. The procedure may be flawed due to judges' insufficient knowledge about students, particularly borderline students. Variable outcomes of Angoff procedures are not surprising in themselves,[14] but the considerable magnitude of the differences we found was rather unexpected.

In this study, the judges were given feedback in the form of the actual test results. The participants said that it was difficult for them to estimate the level of the borderline student while receiving feedback in the form of the average student's test results. This may have led to higher estimates than they would have given otherwise. However, the fact that all 4 panels were exposed to the same feedback makes it unlikely that the higher estimates accounted for all the differences between the groups.

All panels were well aware of item difficulty, as can be concluded from the high Pearson correlations. The more experienced panels (the anatomists because they taught the students and the graduates because they had recently finished their clerkships and knew what knowledge was required) set lower standards than the less experienced panels (the clinicians, who had hardly any contact with Year 4 students and the Year 4 students, who did not yet know what would be expected of them in clerkships). The standards set by the more experienced judges – the anatomists and the recent graduates – would appear to be the most plausible and justifiable. If these standards were to be applied, between a quarter and a half of the students in this study would fail the test. This applies equally to PBL and non-PBL students, who had similar test results, as has been shown.[7] This can be seen as an indication that students' anatomy knowledge is insufficient. However, the possibility of overestimation cannot be ruled out, particularly as the literature shows that students' performance tends to be overestimated.[11]

CONCLUSION

The results of this study show a sizeable discrepancy between students' performance on an anatomy test and the standards set by experts. This suggests that there is cause for concern regarding students' level of

anatomy knowledge. It is possible that similar research in other topics, such as physiology or pharmacology, would yield similar results. In other words, the problem may well be that it is not students' anatomy knowledge that is deficient, but that there is a more general problem in identifying the content of medical curricula for different disciplines and establishing how much knowledge students should acquire. The significant differences between the standards set by staff (clinicians and anatomists) in our study support this assumption. Further research will be needed to clarify whether or not students' knowledge is deficient and, if so, how it can be improved.

Contributors: all authors contributed to the design of the study. KP and AS collected the data. KP and CvdV analysed the data. KP wrote the first draft of the paper. All authors contributed to the interpretation of the results and the final version of the manuscript.
Acknowledgements: the authors thank Ron Hoogenboom for his support with the statistical analysis and Mereke Gorsira for critically reading and correcting the English manuscript.
Funding: none.
Conflicts of interest: none.
Ethical approval: not required.

REFERENCES

1 Schmidt HG, Dauphinee WD, Patel VL. Comparing the effects of problem-based and conventional curricula in an international sample. *J Med Educ* 1987;**62**:305–15.
2 Albanese MA, Mitchell S. Problem-based learning: a review of literature on its outcomes and implementation issues. *Acad Med* 1993;**68**:52–81.
3 Vernon DT, Blake RL. Does problem-based learning work? A meta-analysis of evaluative research. *Acad Med* 1993;**68**:550–63.
4 Verhoeven BH, Verwijnen GM, Scherpbier AJJA, Holdrinet RSG, Oeseburg B, Bulte JA, van der Vleuten CPM. An analysis of progress test results of PBL and non-PBL students. *Med Teach* 1998;**20**:310–6.
5 Prince KJAH, Van de Wiel MWJ, Scherpbier AJJA, van der Vleuten CPM, Boshuizen HPA. A qualitative analysis of the transition from theory to practice in undergraduate training in a PBL-medical school. *Adv Health Sci Educ Theory Pract* 2000;**5**:105–16.
6 van Mameren H, Leiner T, Wanders A, van der Vleuten CPM. Anatomie/embryologieonderwijs aan de Universiteit Maastricht; het oordeel van studenten geneeskunde. [Students' evaluation of anatomy/embryology teaching at Maastricht University.] *Bull Med Onderwijs* 1997;**16**:91–100.
7 Prince KJAH, van Mameren H, Hylkema N, Drukker J, Scherpbier AJJA, van der Vleuten CPM. Does problem-based learning lead to deficiencies in basic science knowledge? An empirical case on anatomy. *Med Educ* 2003;**37**:15–21.
8 Monkhouse WS. Anatomy and the medical school curriculum. *Lancet* 1992;**340**:834–5.
9 Angoff WH. Scales, norms and equivalent scores. In: Thorndike RL, ed. *Educational Measurement.* Washington DC: American Council on Education 1971;508–600.
10 Norcini JJ, Shea JA. The credibility and comparability of standards. *Appl Measurement Educ* 1997;**10**:39–59.
11 Verhoeven BH, Verwijnen GM, Muijtjens AMM, Scherpbier AJJA, van der Vleuten CPM. Panel expertise for an Angoff standard setting procedure in progress testing: item writers compared to recently graduated students. *Med Educ* 2002;**36**:860–7.
12 Verhoeven BH, Van der Steeg AFW, Scherpbier AJJA, Muijtjens AMM, Verwijnen GM, van der Vleuten CPM. Reliability and credibility of an Angoff standard setting procedure in progress testing using recent graduates as judges. *Med Educ* 1999;**33**:832–7.
13 Imbos T, Drukker J, van Mameren H, Verwijnen M. The growth in knowledge of anatomy in a problem-based curriculum. In: Schmidt HG, De Volder ML, eds. *Tutorials in Problem-Based Learning.* Maastricht/Assen: Van Gorcum 1984;106–16.
14 Cizek GJ. *Setting Performance Standards: Concepts, Methods and Perspectives.* Mahwah, New Jersey: Erlbaum 2001.

Received 7 April 2004; editorial comments to authors 11 May 2004; accepted for publication 25 August 2004

APPENDIX

Examples of test items

Patient case description 1

The general practitioner (GP) examines the eye movements of Mrs Ommen. He asks her to look in all directions. Her right eye does not move to the right when she looks in that direction. The GP suspects a paralysis of a specific extra-ocular muscle.

Item Which muscle is involved?

Answer Lateral rectus

Patient case description 2

Paula Woods, 17 years of age, enters the emergency department hopping on 1 leg. Two hours earlier she sprained her right ankle during a badminton match.

She felt something crack at the outside of her foot. Her ankle swelled immediately. She was able to stand on it, although this hurt a lot. She did not cool the ankle.

The doctor in the emergency department suspects an inversion trauma.

Item The talus forms a synovial joint with the tibia and fibula (= talocrural joint), the calcaneus (= subtalar joint) and the calcaneus and navicular bone (= talocalcaneonavicular joint)

Inversion of the foot takes place:

(a) only in the talocrural joint
(b) only in the subtalar joint
(c) in both the talocrural joint and the subtalar joint
(d) in both the subtalar and the talocalcaneonavicular joints

Answer (d)