



Assessment of spatial anatomical knowledge with a 'three-dimensional multiple choice test' (3D-MC)

Sebastian Schubert, Kai P. Schnabel & Andreas Winkelmann

To cite this article: Sebastian Schubert, Kai P. Schnabel & Andreas Winkelmann (2009) Assessment of spatial anatomical knowledge with a 'three-dimensional multiple choice test' (3D-MC), *Medical Teacher*, 31:1, e13-e17, DOI: [10.1080/01421590802334325](https://doi.org/10.1080/01421590802334325)

To link to this article: <https://doi.org/10.1080/01421590802334325>



Published online: 03 Jul 2009.



Submit your article to this journal [↗](#)



Article views: 483



Citing articles: 5 View citing articles [↗](#)

WEB PAPER

Assessment of spatial anatomical knowledge with a ‘three-dimensional multiple choice test’ (3D-MC)

SEBASTIAN SCHUBERT, KAI P. SCHNABEL & ANDREAS WINKELMANN

Charité Universitätsmedizin Berlin, Berlin, Germany

Abstract

Background: Text only multiple choice questions (MCQs) are often inadequate to assess anatomical and histological knowledge and may encourage students to memorize abstract textbook knowledge. An alternative are the “spotters” or “tag tests” well-known in North American and British anatomy. However, the psychometric properties of this assessment have only been reported in one study for a format using short answer questions.

Aims: To describe the implementation and feasibility of a multiple choice “tag test” (3D-MC) using prosected specimens, histological slides, models and radiographs; to report the psychometric properties and students’ acceptance of the 3D-MC; to compare it with a traditional multiple choice format.

Results: The administration of the 3D-MC did not pose any major problems. The 3D-MC was significantly easier (mean scores 75% vs. 64%) than traditional MCQs. The estimated correlation (corrected for attenuation) between the two MCQ formats was $r=0.814$. Reliability for the 3D-MC was .665 for 30 items. Student acceptance was very high.

Conclusions: The 3D-MC is a feasible, reliable and well-accepted test of anatomical knowledge. Further research should investigate if the higher cost as compared to MCQs using photographs is justified by the assessment of different knowledge and abilities as compared to MCQs using photographs.

Introduction

Well-constructed multiple choice questions (MCQs) are objective, reliable, valid and efficient assessments of cognitive knowledge (Downing 2002; Schuwirth & van der Vleuten 2003, 2004; Haladyna 2004). Nevertheless, MCQs using only written text as stimulus are often inadequate to assess three-dimensional spatial knowledge in anatomy and histology, as the stimulus format largely determines what an item assesses (Schuwirth & van der Vleuten 2004). If used as the sole assessment method, they may also encourage students to memorize abstract textbook knowledge instead of gaining a three-dimensional spatial understanding in the dissecting room (Garg et al. 2001). On the other hand, oral exams (vivas) using prosected specimens or histological slides, which are still a widespread form of assessment in many countries, usually have a lower reliability per hour of testing time (Wass et al. 2001), require more resources and are therefore much less efficient than MCQs. To combine the advantages of MCQs with those of exams using three-dimensional prosected specimens and anatomical models, we introduced a new assessment method into our Reformed Medical Curriculum (RMC). We called this new assessment ‘3D-MC’ because it combines the MCQ format with 3D-objects (specimens, models, bones etc.) as stimulus material. The 3D-MC is a single best answer MCQ version of the

Practice points

- A multiple choice tag test (3D-MC) is a feasible, reliable and well-accepted test of anatomical knowledge.
- Further research should investigate if the higher cost as compared to MCQs using photographs is justified by the assessment of different knowledge and abilities as compared to MCQs using photographs.

‘spotter’ or ‘tag test’ well-known in North American and British anatomy (Peel 1998; Heylings 2002). To our knowledge, the psychometric properties of this kind of assessment method have only been reported in one study for a format using short answer questions (Adamczyk et al. 2007).

Aims of this paper

- (1) To describe the implementation of the 3D-MC.
- (2) To ascertain the feasibility of the 3D-MC as a summative test.
- (3) To report descriptive statistics and correlations with traditional MCQs.
- (4) To report the students’ acceptance of the new assessment.

Correspondence: Sebastian Schubert, Assessment Division and Reformed Medical Curriculum, Working Group, Deanery of Student Affairs, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. Tel: +4930450576037; fax: +4930450576952; email: sebastian.schubert@charite.de

Background

The RMC is a problem-based, parallel-track curriculum, which was introduced in 1999 (Ortwein et al. 2004). After a two-week orientation unit it begins with a five week 'locomotion' module. This module teaches basic and clinical sciences of the locomotor system, focussing on the lower limb. Anatomy is one of the major topics of this module. Gross anatomy of the lower limb is taught in introductory seminars and in three practical courses, which use prosected specimens, skeletons and models for demonstration and discussion. This is supported by courses in living anatomy taught by clinicians. Hands-on dissection by students, however, is restricted to a two-hour session, which only suffices to give students a general idea of the difficulties in accessing certain anatomical structures and of the production of anatomical specimens. Histology of connective tissues, muscles and nerves is taught in seminars and two practicals. Attendance in all of these courses is voluntary but usually close to 100%.

At the end of the 14-week semester, all learning objectives of the semester, including those of the module 'locomotion', are assessed with MCQs and an Objective Structured Clinical Examination (OSCE). Prior to 2005, the OSCE included one station with anatomical models or living anatomy and one with radiological anatomy. For practical reasons, prosected specimens could not be used in the OSCE. This meant that most anatomical learning objectives were assessed with MCQs only.

As stated above, it was our impression that three-dimensional anatomical knowledge was not adequately assessed this way. Even though the true 'object' of medicine is, of course, the living organism, all anatomical knowledge cannot be taught and tested on the living. Therefore anatomical specimens often come as close as possible to the authentic medical context.

Methods

Test instrument

Test development was based on the learning objectives of the module 'locomotion' in the first semester. Twelve of the forty-two learning objectives were selected to be assessed with 3D-MCQs. A test blueprint specified the number of questions per learning objective. They cover the macroscopic, microscopic and radiological anatomy of the lower limb. Based on these learning objectives, 31 3D-MCQs covering bones, muscles, ligaments, nerves and blood vessels in all regions of the lower limb were produced by two anatomists who teach in the RMC. The 3D-MCQs were written according to published item writing guidelines (Case & Swanson 2002; Haladyna 2004). They were all 'first order' questions, i.e. they asked for identification of structures, and not for function of the tagged structure or similar 'secondary' information. Eight questions related to prosected specimens, eight to histological slides, nine to models or bones and six to radiographs. The questions were reviewed by two other anatomists and an expert in MCQ-writing using photographs of the specimens.

Three sample questions are included in Appendix A. All questions were printed as a test sheet on which the students directly marked their answers. To compare the 3D-MCQs with multiple choice questions using only written text (Text-Only-MCQs), we also prepared 31 Text-Only-MCQs covering the same content as the 3D-MCQs.

Test administration

On the day of the examination the specimens, models, bones and radiographs were arranged in a circuit of 31 'stations' in a dissecting room of the Center for Anatomy (Figure 1). They were labelled with a single tag for questions of the type 'Which structure is tagged in this specimen' or with multiple numbered tags for questions of the type 'Which tag marks structure X'. The histological slides were fixed on microscopes and required either the identification of the tissue (e.g. skeletal muscle) or the identification of a specific structure identified by a pointer (e.g. a nerve in a specimen of skeletal muscle). Radiographs were presented on light boxes.

The test was administered in two groups of 31 and 30 students, with the second following the first immediately. The students were allocated to the groups in alphabetical order of their last name. After a detailed explanation of the testing procedure each student received a copy of the test sheet indicating his or her starting position in the circuit. All students in a group began the test simultaneously. They were given one minute to mark their answer on the test sheet before a bell prompted them to rotate to the next station. The students were instructed not to touch any of the specimens in order to ensure the same conditions for all students. They were allowed to focus the microscopes if necessary. Two anatomists were present to check the microscopes and macroscopic specimens during the test. The students handed in their test sheets immediately after the last station. They answered the Text-Only-MCQs directly after the 3D-MCQs as part of the final



Figure 1. View of the dissecting room, Center for Anatomy, during the time of the exam. Tables with dissected specimens and light boxes for radiographs are in the bays on the right, microscopes and anatomical models on the tables on the left.

assessment for all the learning objectives of that semester. This assessment contained a total of 155 MCQs.

Evaluation instrument

Students' opinions about the 3D-MC were collected immediately after the test using a questionnaire. The questionnaire contained three statements for each of the four question categories, i.e. macroscopic specimens, histological slides, bones and models, radiographs. Students were asked to indicate on a five point Likert-type scale whether they believed the 3D-MCQs to be more meaningful to assess anatomical knowledge than classical MCQs, to be easier than classical MCQs and if one minute per station was sufficient.

Sample

The 3D-MC was part of the summative end of semester examination for the first-semester medical students in the winter semester 2004/2005. Sixty-one students participated in the test. All students completed the test.

Statistics

All students were included in the analysis. One question had to be excluded from analysis due to two correct answers. The answers on the test sheets and evaluation forms were manually entered into SPSS. The answers were recoded to '1' for a correct answer and '0' for an incorrect answer. A total score was calculated for each student. Means, standard deviations and Cronbach's Alpha reliability coefficients were computed for each scale.

Results

Feasibility

Despite being an unfamiliar testing format for the students, the administration of the 3D-MC did not pose any major problems. No student missed an item and the minute-by-minute rotation of 31 students did not produce distracting levels of noise or disturbance. The human resources required for this exam are difficult to estimate precisely and obviously depend on the availability of prosected specimens, models, radiographs etc. In our case, where no previous exam item collections were available, it took two experienced teachers about one working

day each to design the different stations (this excludes the time of the reviewers) and another half working day each to arrange them directly before the exam. Three teachers were present during the exam itself (about 1.5 h).

Descriptive statistics

The score distribution for both tests equals a normal distribution. The average test score for the 3D-MC was 22.4 points (75%, SD 3.6 points, range 11–28 points). The average test score for the Text-Only-MC was 19.1 points (64%, SD 4.5, points range 5–28 points). This difference is statistically significant ($t_{60}=6.8$, $p<0.001$). There was no statistically significant difference between the two groups of students in the 3D-MC. Using our usual passing rate of 60%, seven students would have failed the 3D-MC and 14 students would have failed the Text-Only-MC, if students would have had to pass the two tests separately and not as part of a larger assessment. Table 1 shows the constitution of the test score by question type.

There were no significant differences between the mean scores of the question types in the 3D-MC. In the Text-Only-MC students scored relatively poorly on the histological questions compared to the prosected specimens ($t_{60}=8.7$, $p<0.001$), models or bones ($t_{60}=-7.0$, $p<0.001$), radiographs ($t_{60}=-8.4$, $p<0.001$) and the histological slides in the 3D-MC ($t_{60}=12.8$, $p<0.001$). The mean corrected item total correlation was 0.221 for the 3D-MC and 0.279 for the Text-Only-MC. The total scale reliability (Cronbach's alpha) was 0.665 for the 3D-MC and 0.760 for the Text-Only-MC. Using the Spearman Brown Prediction Formula, 61 items would be needed in the 3D-MC and 38 in the Text-Only-MC to reach an $\alpha>0.80$. The estimated corrected (for attenuation) correlation between the 3D-MC and the Text-Only-MC was $r=0.814$ ($p<0.001$).

Acceptance

The results of the evaluation questions, presented in Table 2, show that a large majority of the students considered the 3D-MCQs a better and easier assessment of anatomical knowledge than the Text-Only-MCQs. A minority of between 10.9% and 21.4% did not find the 3D-MCQs easier than the Text-Only-MCQs. Most students found one minute per question sufficient. Almost 20% found one minute too short for the histological slides. On the other hand more than 30%

Table 1. Constitution of test score by question type.

Question types	N	3D-MC		Text-only-MC	
		Mean scores	Standard deviations	Mean scores	Standard deviations
Prosected specimens	8	73%	20%	70%	19%
Histological slides	7	78%	17%	49%	21%
Models or bones	9	73%	18%	67%	19%
Radiographs	6	74%	16%	68%	19%
Total	30	75%	12%	64%	15%

Table 2. Results of evaluation questions.

	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
3D-MC more meaningful than text-only-MC					
Prosected specimens	56.1%	31.6%	10.5%	1.8%	0.0%
Histological slides	47.4%	28.1%	21.1%	1.8%	1.8%
Models or bones	51.8%	32.1%	14.3%	1.8%	0.0%
Radiographs	58.9%	28.6%	10.7%	1.8%	0.0%
3D-MC easier to answer than text-only-MC					
Prosected specimens	28.6%	30.4%	23.2%	14.3%	3.6%
Histological slides	28.6%	28.6%	21.4%	19.6%	1.8%
Models or bones	41.8%	27.3%	20.0%	7.3%	3.6%
Radiographs	25.5%	21.8%	40.0%	9.1%	3.6%
3D-MC should be regular part of assessment					
	47.2%	34.0%	13.2%	3.8%	1.9%
Time (one minute) per question					
	Much too long	Too long	Just right	Too short	Much too short
Prosected specimens	3.5%	14.0%	75.4%	5.3%	1.8%
Histological slides	3.5%	14.0%	63.2%	14.0%	5.3%
Models or bones	1.8%	14.3%	71.4%	10.7%	1.8%
Radiographs	5.4%	26.8%	60.7%	5.4%	1.8%

found one minute too long for radiographs. Eighty percent of the students indicated that a 3D-MC should be a regular part of the end of semester exams.

Discussion

The 3D-MC, a multiple choice 'spotter' examination for summative assessment of anatomical knowledge, has been successfully introduced into the first year of our reformed medical curriculum. It is a feasible and well-accepted alternative to more traditional written or oral examinations. An advantage of the 3D-MC over Text-Only-MCQs is the increased construct and consequential validity (Downing 2002), i.e. it assesses three-dimensional spatial knowledge instead of abstract textbook knowledge and it encourages students to gain a three-dimensional spatial understanding in the dissecting room. An advantage of the 3D-MC over oral exams using prosected specimens, histological slides, models, bones and radiographs is the higher reliability per hour of testing time and the efficiency of the highly structured assessment and rating procedure of a multiple choice test. Our first exam in the new format achieved an acceptable level of total scale reliability (Cronbach's alpha) at 0.67. Approximately one hour of testing time would be needed to reach a Cronbach's Alpha of 0.8, which is usually regarded as minimum for summative examinations (Downing 2004).

Despite the fact that we tried to write Text-Only-MCQs that covered exactly the same knowledge as the corresponding 3D-MCQs, the two tests varied significantly in difficulty. This is congruent with the fact that the students found the 3D-MC questions easier to answer and may be an indication that assessing three-dimensional anatomical knowledge with Text-Only MCQs adds difficulty unrelated to content. The difference in difficulty may also, at least in theory, be due to the different administration modes. In the 3D-MC students had exactly one minute to answer each question and could not go back to review the questions as in the Text-Only-MC. However, the option of changing initial answers for which students previously had doubts during a multiple choice test usually brings about better overall test results (Fischer et al. 2005),

which is in contrast to our results. Further research is needed to clarify this issue. For reasons of anonymity, we were not able to analyse whether the students who did find the 3D-MC more difficult also had lower mean scores. The high corrected correlation of 0.814 indicates that our new assessment method largely measures the same or very closely related knowledge as the traditional multiple choice test. It therefore challenges, at least to some degree, our initial assumption that text-only MCQs are often inadequate to assess three-dimensional spatial knowledge in anatomy and histology. There may be an even higher correlation between 3D-MCQs and Text-Only-MCQs using photographs, especially for the histological slides and radiographs. Further research should investigate the different abilities involved in answering 3D-MCQs, MCQs using photographs instead of real specimens and text only MCQs.

Although we do not have any 'hard data' to prove this assumption, the introduction of the 3D-MC had an apparent formative effect in that students now regularly use the individual study times in our dissecting rooms, while previously, when they were assessed by text-only MCQs only, there was little incentive for the reformed track students to use this study tool of the Anatomy Department. One disadvantage with this format is the time and effort of re-testing students who failed or missed the test, as the same number of stations will usually have to be set up for a small number of students. This format may also have practical limitations with large numbers of students. If large numbers of students have to be tested in consecutive groups, every other group will need different questions to insure that communication between groups does not threaten the validity of the test.

Conclusion

The 3D-MC is a feasible, reliable and well-accepted test of anatomical knowledge. Further research should investigate if the higher cost of the 3D-MC as compared to MCQs simply using high-quality photographs is justified by the assessment of different knowledge and abilities (e.g. spatial aspects of anatomical knowledge).

Acknowledgements

We would like to thank PD Dr. Pia Welker, Prof. Gottfried Bogusch, Dieter Lange, René Lange and all the students who participated in the study.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Notes on contributors

Mr. SCHUBERT is a physician, Assessment Division, Dean's Office for Student Affairs, Charité Universitätsmedizin Berlin, Berlin, Germany.

Dr. SCHNABEL is a physician and Master of Medical Education, Reformed Medical Curriculum Working Group, Dean's Office for Student Affairs, Charité Universitätsmedizin Berlin, Berlin, Germany.

Dr. WINKELMANN is senior lecturer, Institute of Cell Biology and Neurobiology, Center for Anatomy, Charité Universitätsmedizin Berlin, Berlin, Germany.

References

Adamczyk C, Huenges B, Müller-Gerbl M, Putz R. 2007. Tag test as a particular kind of examination in the dissection course, GMS Zeitschrift für Medizinische Ausbildung, 24, Doc152. Available at <http://www.egms.de/en/journals/zma/2007-24/zma000446.shtml>.

- Case SM, Swanson DB. 2002. Constructing written test questions for the basic and clinical sciences Philadelphia, PA: National Board of Medical Examiners
- Downing SM. 2002. Assessment of knowledge with written test forms. In: G.R. Norman, CPM van der Vleuten, D.I. Newble, editors. International Handbook of Research in Medical Education. Dordrecht: Kluwer Academic Publishers. pp 647–672.
- Downing SM. 2003. Validity: On meaningful interpretation of assessment data. Med Educ 37:830–837.
- Downing SM. 2004. Reliability: On the reproducibility of assessment data. Med Educ 38:1006–1012.
- Fischer MR, Herrmann S, Kopp V. 2005. Answering multiple-choice questions in high-stakes medical examinations. Med Educ 39:890–894.
- Garg AX, Norman G, Sperotable L. 2001. How medical students learn spatial anatomy. Lancet 357:363–364.
- Haladyna TM. 2004. Developing and validating multiple-choice test items. Mahwah, NJ: Lawrence Erlbaum Associates.
- Heylings DJ. 2002. Anatomy 1999–2000: The curriculum, who teaches it and how? Med Educ 36:702–710.
- Ortwein H, Mühlhngaus I, Schnabel KP, Terzioglu P, Wilke A, Scheffner D, Burger W. 2004. Medical education in Berlin – Reformed curriculum and communication skills training. J Kansai Med Univ 56:172–181.
- Peel S. 1998. An innovative problem-solving assessment for groups of first-year medical undergraduates–Think Tanks. Med Educ 32:35–39.
- Schuwirth LW, van der Vleuten CP. 2003. ABC of learning and teaching in medicine: written assessment. Br Med J 326:643–645.
- Schuwirth LW, van der Vleuten CP. 2004. Different written assessment methods: What can be said about their strengths and weaknesses? Med Educ 38:974–979.

Appendix A



Which structure is marked in this specimen?

- (A) Femoral artery
- (B) Profunda femoris artery
- (C) Obturator artery
- (D) Superior gluteal artery
- (E) Inferior gluteal artery



Which ligament is marked with the orange dot in this model?

- (A) Deltoid ligament
- (B) Anterior tibiofibular ligament
- (C) Anterior talofibular ligament
- (D) Calcaneofibular ligament
- (E) Posterior talofibular ligament



Which of the arrows marks the anterior superior iliac spine?

- (A)
- (B)
- (C)
- (D)
- (E)