Annals of Anatomy 209 (2017) 93–100

Contents lists available at ScienceDirect

Annals of Anatomy

journal homepage: www.elsevier.com/locate/aanat

Do they know too little? An inter-institutional study on the anatomical knowledge of upper-year medical students based on multiple choice questions of a progress test☆

Irene Brunk[a,*], Stefan Schauber[b], Waltraud Georg[c]

[a] Institut für Integrative Neuroanatomie, Charité-Universitätsmedizin Berlin, Germany
[b] Faculty of Educational Sciences/Faculty of Medicine, University of Oslo, Norway
[c] Helios Kliniken GmbH, Germany

ARTICLE INFO

Article history:
Received 24 February 2016
Received in revised form 3 September 2016
Accepted 16 September 2016

Keywords:
Anatomy
Undergraduate medical education
Knowledge
Outcome
Surgery
Progress testing
Standard setting

ABSTRACT

The depth of medical students' knowledge of human anatomy is often controversially discussed. In particular, members of surgical disciplines raise concerns regarding deficits in the factual anatomical and topographical knowledge of upper-year students. The question often raised is whether or not medical students have sufficient anatomical and topographical knowledge when they graduate from medical school. Indeed, this question is highly relevant for curricular planners. Therefore, we have addressed it by evaluating the performance of students in the 5th and 6th years of their studies on anatomical multiple choice questions from the Berlin Progress Test Medicine performed at 10 German university medical schools. Results were compared to a reference based on a standard setting (modified Angoff-procedure). The reference was established independently by 5 panels of anatomists at different universities across Germany. As the ratings were independent of university affiliation, teaching-experience or training of the anatomists, an overall cut off score could be calculated which corresponded to 60.4% correct answers for the question set used in this study.

In the progress test, on average only 29.9% of the students' answers were correct, reflecting that the performance was significantly below the expected standard. On the basis of the test results it remained unclear whether acquisition or retention of anatomical information was insufficient. Further evaluation by item characteristics revealed that the students had major difficulty in applying their theoretical knowledge to practical problems in the context of a clinical setting. Thus, our results reveal deficits in the anatomical knowledge of medical students in their final years. Therefore medical curricula should not only focus on enhancing the acquisition and retention of core anatomical knowledge, but aim at improving skills applying this in a clinical setting.

1. Introduction

Anatomy is a central component of preclinical medical education and serves as an essential basis for the understanding of the human body, enabling doctors to perform proper physical examination, derive structural diagnosis and apply therapeutic procedures to patients (Mylopoulos and Woods, 2014; Rikers et al., 2005b; Woods, 2007; Woods et al., 2007a). Nevertheless, the anatomical knowledge of medical students has been controversially discussed in recent years (Bergman et al., 2008, 2011; Prince et al., 2005). In particular, members and associations of surgical disciplines pointed toward deficits in the topographical knowledge of upper-year medical students and young doctors (Chirurgie, 2009; Waterston and Stewart, 2005).

Clinicians from different specialities have repeatedly expressed their impression that anatomical knowledge of medical students is below a minimum level and may even endanger patient safety (Waterston and Stewart, 2005). Some authors propose links between changes in anatomy teaching, the perceived decline in anatomy knowledge of students and young physicians and the increase in reported medical malpractice (Older, 2004; Turney, 2007). Deficits could be attributed to various factors: (1) A major factor is how teaching anatomy has changed over the last decades: While changes such as the vertical integration of subjects into the curriculum and an interdisciplinary approach are positively seen and believed to promote retention of knowledge (Bergman et al., 2011), the broad introduction of problem based learning (PBL)

☆ This paper belongs to the special issue Progress in Medical Education.

* Corresponding author.
E-mail address: irene.brunk@charite.de (I. Brunk).

http://dx.doi.org/10.1016/j.aanat.2016.09.004
0940-9602/

raised concerns regarding the acquisition of anatomical/basic science knowledge (Bergman et al., 2014; Cahill et al., 2000; Williams and Lau, 2004). However, a comparison of students from non-PBL *versus* PBL-curricula revealed no difference in their anatomical knowledge (Prince et al., 2003). In fact, students perceived their knowledge as deficient independent of the type of curriculum they were in (Prince et al., 2003). (2) Some new curriculum dropped dissection classes from their program, despite the fact that the use of human cadavers for teaching anatomy has been found to have a positive impact on the acquisition of topographical as well as general anatomical knowledge (Biasutto et al., 2006; Saltarelli et al., 2014; Winkelmann, 2007). (3) Another factor might be the fact that anatomy is increasingly taught by non-medically trained staff and the student–staff ratios have severely increased (Pryde and Black, 2005). There is, however, currently no published study investigating whether the field of qualification of teachers has any impact on the anatomical knowledge of students. Intriguingly, complaints about the declining anatomical knowledge of students and graduates were already raised more than forty years ago (Sinclair, 1975). Thus, concerns expressed in the recent years may be independent of the curricular changes occurring in the last couple of decades.

Nevertheless, this ongoing discussion on the appropriateness of graduates' familiarity with anatomy flags up a potential discrepancy between the importance of anatomical knowledge for clinical practice (Rikers et al., 2005a; Woods et al., 2007a, 2007b, 2006) and the outcome of current medical education (Older, 2004). Therefore, it is important to know whether the *perceived* deficits can indeed be verified. In the context of the worldwide movement for reforming medical curricula taking place in the last couple of decades, this question has particularly relevance: curricular organisers require information about the acquisition of subject-specific content in different curricula in order to counteract potential deficits. Such information can then form the basis to integrate the various medical disciplines, including anatomy, in such a way that students are best prepared for their future clinical practice. However, before these practical issues can be addressed, obtaining an assessment of student proficiency is imperative.

Anatomical knowledge, as other medical subjects, is typically measured by various assessment tools such as multiple choice exams, oral exams or objectively structured practical examinations (Schoeman and Chandratilake, 2012). One established way to assess factual knowledge and its clinical application are multiple choice items (Wass et al., 2001). Compared to other procedures, one clear benefit of multiple choice items is that it is easier to administer them in a standardised way across different faculties and/or curricula. In addition, MCQs provide high reliability and objectiveness. One testing format that utilises the virtues of multiple choice items is progress testing. The main aim of this procedure is to follow students learning trajectories over the course of their studies. Progress tests depict the development – hence '*progress*' – of knowledge over the course of (medical) education. To this aim, students are repeatedly tested with different items while the overall distribution of content (topics/subjects) remains constant. The items are intended to assess knowledge relevant for a doctor's "first-day-in-practice" and thus are comparable to the demands of a graduation exam. In the case of the Progress Test Medicine (PTM) of the *Charité Universitätsmedizin Berlin* each question is reviewed twice by a multidisciplinary review team with regard to this aspect. Further details on progress tests can be found in Wrigley et al. (2012).

Typically, progress tests are organised as inter-institutional or even international collaborations (Freeman et al., 2010; Tio et al., 2016). Such cooperative efforts provide an optimal basis for comparisons of large cohorts of student levels of medical knowledge across institutions. In Germany, the PTM is set up as an inter-institutional cooperation. The test is conducted biannually (once every semester) and consists of 200 multiple choice questions which are constructed as single best answer items (Nouns and Georg, 2010). Students from their 1st to the 6th academic years take part in the test. A total number of about 180,000 participants from 17 medical schools in Germany and Austria have taken the test since its introduction in 1999. As the PTM is a formative assessment tool, students do not prepare extensively. Consequently, results can be assumed to be unbiased by test-preparation efforts and thus reflect students' readily retrievable knowledge. Previously, results from the PTM have been used for curricular comparisons and benchmarking the performance of students (in different subjects) from different backgrounds (home medical school, types of curricula) (Nouns et al., 2012; Schauber et al., 2015; Tio et al., 2016; Verhoeven et al., 1998). Three specific features of the PTM make this an attractive assessment tool for our question delineated above. First, the PTM is synchronously administered across many medical schools in Germany. Secondly, each PTM contains between 15 and 20 questions testing anatomy knowledge, including gross anatomy, neuroanatomy and histology/cell biology. Thirdly, students participate at all stages of their studies. Hence, information about medical student performance in their 5th and 6th academic years can be extracted from the PTM and used for assessing their knowledge in human anatomy across several medical faculties in Germany.

In order to approach the question of whether or not student performances are satisfactory, an objective, reliable and valid standard is crucial. While an assessment tool such as the PTM provides data on *actual* performance levels, these empirical results need to be compared to what teachers, lecturers, or experts (*i.e.* content matter experts) *expect* from graduates. If such expectancies are obtained in a systematic and objective manner, they can serve as a reference – a *standard* – to which students' actual performances can be compared. Put briefly, such standards set by content matter experts define how many students *should* correctly respond to a specific question. The procedures for obtaining such a standard for a whole set of questions (*i.e.* a test) are referred to as "standard setting procedures" in the educational assessment literature. When making claims about performance or ability deficits, standards are crucial as they link the tested content to the expected competence levels (Bandaranayake, 2008; Ben-David, 2000).

Standard-setting procedures have already been used for establishing references for progress test results (Verhoeven et al., 2002) as well as for the evaluation of anatomy knowledge (Prince et al., 2005). One of the best-known methods, the Angoff method, is suitable for setting criterion referenced standards for multiple choice examinations (Angoff, 1971; Bandaranayake, 2008). According to this method, for each question, a number of judges have to estimate the percentage of a group of minimally-competent candidates who are at the borderline of pass and fail (*i.e.* a 'borderline'-examinee or minimally-competent student) that would respond correctly. For instance, the judges have to answer a question such as: "How often would you expect a group of students on the verge of passing/failing this exam to answer this specific item correctly?" However, research on this procedure has found that estimating the responses of a group of "borderline"-examinees is difficult for unexperienced judges (Norcini, 1994). Additionally, the estimation of the percentage of correct responses of a group is a problem, as even experienced judges tend to choose values between 40% and 60%. Thus, Impara and Plake (1997) proposed a modified Angoff-procedure, also referred to as the he 'Yes/No-Method'. In this variant, judges have to imagine *one* 'borderline'-candidate and to decide for each question, if he/she would give a correct answer or not (Chinn and Hertz, 2002). The reference standard is then calculated as average number of questions a 'borderline'-examinee is expected to answer correctly by the judges.

In summary, to investigate the question as to whether a deficit in medical students' anatomical and topographical knowledge can be found empirically, two sources of information are needed: First,