

Multiple-Choice versus Open-Ended Questions in Advanced Clinical Neuroanatomy: Using a National Neuroanatomy Assessment to Investigate Variability in Performance Using Different Question Types

Wassim H. Merzougui, ^{1,2*} Matthew A. Myers,^{1,3} Samuel Hall,^{1,4} Ahmad Elmansouri,⁵ Rob Parker,⁶ Alistair D. Robson,⁶ Octavia Kurn,⁶ Rachel Parrott,⁷ Kate Geoghegan,⁸ Charlotte H. Harrison,⁹ Deepika Anbu,⁶ Oliver Dean, ⁶ Scott Border ¹

¹Center for Learning Anatomical Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

²Department of Trauma and Orthopedics, Pilgrim Hospital, Boston, United Kingdom

³Department of Neurosurgery, Salford Royal NHS Foundation Trust, Salford, United Kingdom

⁴Department of Neurosurgery, Wessex Neurological Centre, Southampton, United Kingdom

⁵Department of Medical Education, Brighton and Sussex Medical School, University of Sussex, Brighton, United Kingdom

⁶University of Southampton, School of Medicine, Southampton, United Kingdom

⁷Department of Anatomy, St Andrews University, St Andrews, Scotland

⁸Department of Cardiology, Royal United Hospital, Bath, United Kingdom

⁹Department of Emergency Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

Methods of assessment in anatomy vary across medical schools in the United Kingdom (UK) and beyond; common methods include written, spotter, and oral assessment. However, there is limited research evaluating these methods in regards to student performance and perception. The National Undergraduate Neuroanatomy Competition (NUNC) is held annually for medical students throughout the UK. Prior to 2017, the competition asked open-ended questions (OEQ) in the anatomy spotter examination, and in subsequent years also asked single best answer (SBA) questions. The aim of this study is to assess medical students' performance on, and perception of, SBA and OEQ methods of assessment in a spotter style anatomy examination. Student examination performance was compared between OEQ (2013–2016) and SBA (2017–2020) for overall score and each neuroanatomical subtopic. Additionally, a questionnaire explored students' perceptions of SBAs. A total of 631 students attended the NUNC in the studied period. The average mark was significantly higher in SBAs compared to OEQs (60.6% vs. 43.1%, $P < 0.0001$)—this was true for all neuroanatomical subtopics except the cerebellum. Students felt that they performed better on SBA than OEQs, and diencephalon was felt to be the most difficult neuroanatomical subtopic ($n = 38$, 34.8%). Students perceived SBA questions to be easier than OEQs and performed significantly better on them in a neuroanatomical spotter examination. Further work is needed to ascertain whether this result is replicable throughout anatomy education. *Anat Sci Educ* 0: 1–10. © 2021 The Authors. *Anatomical Sciences Education* published by Wiley Periodicals LLC on behalf of American Association for Anatomy.

Key words: gross anatomy education; neuroanatomy anatomy education; medical education; undergraduate education; methods of assessments; anatomy assessment

*Correspondence to: Dr. Wassim H. Merzougui, Faculty of Medicine, University of Southampton, Tremona Road, Southampton, SO16 6YD, United Kingdom. E-mail: wm1g13@soton.ac.uk

Additional supporting information can be viewed in the online version of this article.

Received 12 June 2020; Revised 10 December 2020; Accepted 5 January 2021.

Published online ## ### ## in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/ase.2053

© 2021 The Authors. *Anatomical Sciences Education* published by Wiley Periodicals LLC on behalf of American Association for Anatomy.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Importance of Anatomy in Undergraduate Medical Education

Anatomy is fundamental within the undergraduate medical education and is strongly associated with clinical competency (Sbayeh et al., 2016). Medical students consider anatomy teaching an essential part of the medical curriculum and deem it vital for future clinical practice (Ali et al., 2015). However, both students themselves and senior NHS clinicians that train them in the workplace report that their anatomical knowledge is insufficient at the point of graduation (Gogalniceanu et al., 2009; Monrouxe et al., 2017; O’Keeffe et al., 2019).

This may be due in part to the globally evolving nature of modern curricula where anatomy has featured less heavily over recent decades (Heylings, 2002; Chapman and Hakeem, 2015; Memon, 2018; Trautman et al., 2019). Considering that assessment is one of the main extrinsic motivators which drive learning, it is important that anatomy assessments are fit for purpose and examine both deep knowledge acquisition and higher order thinking, rather than superficial factual recall (Gogalniceanu et al., 2009). The issue of appropriate assessment strategies is a wider consideration for all disciplines within higher education, and is an area that has been extensively researched.

Research into Higher Education Assessment Practices

Many disciplines will have been dominated historically with the application of both the multiple-choice question (MCQ) examination and the traditional evaluation by written essay style questions (Sambell et al., 1997). In most cases, students prefer the MCQ format and its variants compared to an essay style or short/long structured written type of examination. They reach this conclusion for a whole host of reasons, including perceived difficulty, anxiety, complexity, success expectancy, feeling at ease, and this is true for both male and female students (Zeidner, 1987). However, these conclusions are not reached by all students. There is evidence to suggest that those learners who excel or who are confident in their abilities prefer writing free text answers, where they have an opportunity to fully demonstrate their talents (Birenbaum and Feldman, 1998).

A long-standing viewpoint within this field suggests that MCQ formats or assessments that focus on providing detailed factual answers motivate students more toward using a surface approach, while free text questions tend to encourage a deeper strategy (Entwistle and Entwistle, 1991). This perspective appears to be corroborated by research implying that when students shift from the former to the latter they transfer from a surface approach toward a deep approach (Thomas and Bain, 1984).

Despite this, over time, the MCQ style of question writing has evolved and the current evidence strongly suggests that a well-constructed single best answer (SBA) question can engage problem solving strategies as much as open-ended questions (OEQ) (Hift, 2014). There is no clear position on how students perform in comparison to each question type and this is true across all disciplines within higher education. This has been investigated in a number of subject areas in higher education, often resulting in mixed outcomes (Badger and Thomas, 1991; Bridgeman, 1992; Wolfson et al., 2001; Schuwirth and van der Vleuten, 2004a; Ozuru et al., 2013; Wooton et al., 2014; Brenner et al., 2015; Thompson and O’Loughlin, 2015; Biria and Dehghan, 2016; Adamu et al., 2018; Hauer et al., 2020).

Changes to Modern Day Anatomical Curricula and Assessment Methods

There is currently considerable variation in the way anatomical education is delivered and assessed between medical schools in the United Kingdom (UK) and globally (Turney, 2007). There are multiple types of assessment methods used, which mostly falling into the categories of oral, written, or practical forms.

The repertoire and balance of assessment methods appears to have changed over time, but without any firm or robust conclusions to support one particular approach (Rowland et al., 2011; Papa and Vaccarezza, 2013). Ultimately, the purpose of all curricular assessment is to identify a threshold level of knowledge at which it is appropriate to allow students to progress. (Rowland et al., 2011; Papa and Vaccarezza, 2013). Practical anatomy has been typically assessed via a spotter assessment, also known as the steeplechase examination in some countries (Smith and McManus, 2015; Brenner et al., 2015; Geoghegan et al., 2019; Tirpude et al., 2019). Spotter examinations typically take place in the Dissecting Room and consist of 30–40 stations. Students move around the stations at a predetermined pace (typically 1–1.5 minutes) until all stations are completed. Questions may include various resources including cadaveric specimens, models, imaging, photomicrographs, and diagrams. Anatomy has often been integrated within written papers to assess conceptual and applied aspects of the discipline and that has certainly been the case at Southampton (Smith and McManus, 2015). Typically, questions require long or short structured answers, and are presented as a clinical case.

Evolution of the Spotter Style Examination Approach

This shift in assessment practice is evidence of a gradual evolutionary influence of the spotter style assessment. Historically, candidates have been required to mainly identify structures, however, this has been frequently criticized for only assessing superficial “low level” knowledge based on visual recall. It is now more common to include function-based questions or those that link to clinical understanding. This is considered pedagogically sound due to its capability of testing deeper understanding (Smith and McManus, 2015; Brenner et al., 2015; Choudhury et al., 2016). Furthermore, the standalone spotter style assessment may involve either open-ended questions (OEQ) or a form of multiple-choice question (MCQ) known as single best answers (SBAs) (Sam et al., 2016). However, sometimes the decision on which one to adopt is based on logistical convenience and staffing rather than academic principles (Collins, 2006). Additionally, OEQ marking may involve unwanted variation in scores due to subjective interpretations by teaching staff, and so requires standardized processes (Wass et al., 2001; Mehta et al., 2016). As the emphasis of modern medical curricula has shifted, including notable changes to how practical anatomy is taught and learned, it has become increasingly important to establish evidence-based assessment practices in anatomy, which can ensure safe clinical practice for all future medical graduates (Sugand et al., 2010).

One modern variation of the spotter examination is the Objectively Structured Practical Examination (OSPE). Each station is designed to test a skill, such as relating clinical knowledge to anatomical or histological structures (Yaqinuddin et al., 2013; Choudhury and Freemont, 2017; Tirpude et al., 2019). The integrated anatomy practical paper is another example

(IAPP). The title is intended to acknowledge the inclusion of function and clinical application questions. This format is considered to be an effective and reliable method of assessing anatomical knowledge.

Although various anatomical teaching and assessment methods are being utilized in anatomy, it can often be ethically challenging to undertake assessment-related research within a live medical curriculum, due to the possibility of disadvantaging students through the experimental manipulation of existing procedures. This could easily lead to high stake consequences, such as the possibility of incorrect decisions concerning student progression, or the possibility that students could be held back unfairly.

The Potential Role of Extracurricular Anatomy Assessments

Extracurricular assessments are rare in anatomy and there are seldom opportunities for students to be earn academic awards in niche subject areas (Hall et al., 2020).

Clinical neuroanatomy is one particular specialist field that has been a hot spot for innovative educational interventions because many anatomists struggle to make it understandable and digestible (Chang and Molnár, 2015). Despite this, there are clearly a number of enthusiasts among the medical student population. The National Undergraduate Neuroanatomy Competition (NUNC) is an annual competition held at the University of Southampton run in partnership between medical students, qualified doctors, and academic staff with support from The Anatomical Society (Geoghegan et al., 2019). The aim was to provide an educational and intellectually stimulating event, in which students can enhance their neuroanatomy knowledge, strengthen their portfolios, and help to vertically integrate the subject at all stages of the curriculum. The impact upon Southampton's Centre for Learning Anatomical Sciences has been to increase transparency of the curriculum through a partnership model and to expand the quality of cadaveric resources for undergraduate teaching purposes.

In the current context, the NUNC creates an ideal opportunity to compare performance on SBA and OEQ types using a large cross section self-selecting medical students, representing different institutions, and from all stages of training. Therefore, the current study utilized this event to assess the impact of different question types on neuroanatomy performance in a very traditional spotter test.

We also took the opportunity to ask students about their perceptions of assessment practice at their host institutions. This is important because the learner's assessment experience will determine the way in which the student considers future learning strategies (Ramsden, 1997) and equally, the way in which students think about learning and studying, influences the way in which they tackle assessments.

The aims of this study are to investigate: (1) Medical students' performance in the SBA and OEQ components of the NUNC spotter assessment and (2) The perceptions of medical students toward the use of SBA and OEQ as a method of assessment of anatomical knowledge in the NUNC.

MATERIALS AND METHODS

Examination Structure

The NUNC incorporates two examination papers: a written SBA paper and a practical spotter style examination using brain prosections.

The spotter style examination comprised between 40 and 70 stations, which has increased over the years in line with the evolving nature, and growing attendance, of the NUNC. Each station contained a prosection, or an image, of the brain or spine or spinal cord with two questions per station. At each station, students had 1 minute to answer both parts of the question. All questions were answered under examination conditions and adhere to the standard examination rubric, in line with University of Southampton assessment regulations. For full details see Hall et al. (2016). Delegates were allocated to either a preclinical or clinical category, depending on whether they had started the formal clinical years of medical school (usually third year), and only compared to other delegates in the same category.

In its original format, the NUNC 2013–2016 spotter examinations only comprised OEQ questions. In 2017–2020, the spotter format was changed so that one of the two questions at each station was a SBA question using three distractors. In 2017–2020, the second question at each station remained in the original OEQ format (Table 1). Example questions are demonstrated in Online Appendix 1.

Performance data from SBA questions in 2017, 2018, and 2020 papers (SBA group) were compared to the performance from OEQ in 2013, 2014, 2015, and 2016 (OEQ group). Performance data between SBA questions and OEQ in 2017, 2018, and 2020 papers were further compared to each other. The 2019 data are not available for comparison hence was excluded from the study. At the end of the competition, the scores from the SBA and spotter examination are collated to provide an overall mark for each student.

Quality Assurance and Standard Setting Procedures

The questions for the NUNC were provided by clinicians, members of university faculty, members of the Anatomical Society, and senior medical students on the NUNC committee. All the questions were peer-reviewed by the NUNC committee to ensure quality and consistency. The group of question writers consisted of approximately 10 core members and were reviewed by up to another 20 individuals both within and external to the university. Questions were written based on a number of predefined neuroanatomical subtopics. The number of questions covering each subtopic were discussed and weighted proportionately to the depth and breadth of the subtopic, and also perceived importance in the curriculum. A modified Angoff approach to standard setting was adopted although without the need for determining a passing score (Yim, 2018).

The questions were categorized by neuroanatomical subtopic (cerebrum, diencephalon, vascular, brainstem, spinal cord, cerebellum, and other). The anatomical subtopic "other" typically refers to questions regarding the ventricular system and skull as these questions did not fit into any other category.

Questionnaire

At the end of the 2013–2017 NUNC, students were invited to answer a questionnaire (see Online Appendix 2) which aimed to gauge students' careers intentions and their perceptions of learning neuroanatomy. The items for the questionnaire were chosen with the aim of better understanding students experiences of neuroanatomy at their home institution, and also their

Table 1.

Summary of Participation in the National Undergraduate Neuroanatomy Competition and Examination Information

Year	Number of participants	Number of stations	Number of questions	Total marks available
2013	32	40	80	160
2014	59	42	84	168
2015	91	42	84	168
2016	96	50	100	200
2017	109	42	84	168
2018	126	48	96	192
2020	118	70	137	274
All years	631	334	665	1330

Table illustrates the number of spotter stations and the total marks available in each year, which changed in line with the evolving nature of the National Undergraduate Neuroanatomy Competition. Over the studied period, the total number of questions was 665 equating to 1,330 marks. In 2020, three questions were removed from the examination post hoc review.

perceptions to different assessment styles. Additional information such as student demographics including gender, host institution, and year of study were collected during the online registration process. In 2017, we asked students perceptions toward different assessment methods.

Ethical Approval

This research received ethical approval from the University of Southampton Faculty of Medicine Ethics Committee (Approval ID: 9351).

Statistical Analysis

Performance and feedback responses were compared using descriptive statistics, Mann–Whitney *U* test, Wilcoxon signed-ranked test, or Cohen's *d* as appropriate. Statistical significance was set at $P < 0.05$. Statistical analysis was performed in Prism software, version 7.02 (GraphPad Software Inc., La Jolla CA).

RESULTS

Demographics of Studied Population

Of the seven studied NUNC events, 631 delegates attended; 579 (91.8%) stated their gender and of these 349 (60.3%) were male and 230 (39.7%) female. There were significantly more males in the OEQ group than the SBA group (66.4% vs. 55.6%, $P = 0.01$).

Of the 631 delegates, 336 were in the clinical category and 295 were in the preclinical category. There was no significant difference in the distribution by category between OEQ (57.2% clinical) and SBA (50.1% clinical) groups ($P = 0.09$).

The Spotter Examination

In 2013–2016, 278 delegates completed the spotter examination in the OEQ format, and in 2017–2020 (excluding 2019), 353 delegates completed the new spotter examination format,

with half of the questions as OEQs and the remaining half as SBAs.

The subtopics most heavily represented the cerebrum, vascular system, and brainstem/cranial nerves, whereas spinal cord and cerebellum questions were less common (Table 2). There was no difference in the distribution of questions by subtopic between the two groups (chi-square 2.702, $P = 0.85$).

Examination Performance

Overall, students performed significantly better on the SBA questions than OEQs questions (60.6% vs. 43.1%, $P < 0.0001$) (Table 3). Students perform better in all subtopics with the exception of the cerebellum where they perform significantly worse. Students between 2017 and 2020, who undertook both OEQ and SBA questions within the examination, also performed significantly better on the SBA component of the examination ($P < 0.0001$, Wilcoxon matched pairs signed rank test) (Table 4). Students performed significantly worse on cerebellum OEQ compared to SBA questions ($P < 0.0001$, Wilcoxon matched pairs signed rank test).

Questionnaire

Of the 336 students with a complete data set of 2013–2017, 193 (57%) aspired to enter a neuroscience-based specialty (neurology, neurosurgery, and psychiatry), 48 (14%) aspired to follow a non-neuroscience-based specialty, and 95 (28%) were still undecided.

From the 2017 questionnaire, students agree that they perform better on SBA questions than OEQs with an average rating of 7.47 ± 2.40 (1 = disagree, 10 = agree). They also agree that a timed spotter style assessment is the most effective way of assessing anatomical knowledge (6.54 ± 2.04), although they also consider them to be more stressful due to the time restraints per station (6.23 ± 2.60). From the questionnaire, the subtopic most commonly rated to be the most difficult was

Table 2.

A Summary of the Questions Asked by Subtopic

Subtopic	Question Cohort	
	Open-Ended Questions (2013–2016) n (%)	Single Best Answer Questions (2017–2020) ^a n (%)
Cerebrum	131 (37.6)	61 (38.9)
Diencephalon	42 (12.1)	13 (8.3)
Vascular System	57 (16.4)	24 (15.3)
Brainstem/Cranial Nerves	55 (15.8)	25 (15.9)
Spinal Cord	3 (0.9)	1 (0.6)
Cerebellum	23 (6.6)	11 (7.0)
Other	37 (10.6)	22 (14.0)
Total	348 (100.0)	157 (100.0)

Table shows the distribution of questions by subtopics at the National Undergraduate Neuroanatomy Competitions between the 2013 and 2016 (only cohorts of open-ended questions are shown) and between the 2017 and 2020, excluding 2019, (only single best answer question cohorts are shown).^aFor this period the subtopic distribution for the single best answer questions cohorts in 2017, 2018, and 2020 is only displayed (160 questions).

the diencephalon ($n = 38$, 34.8%), followed by the cerebellum ($n = 24$, 22.0%). Vascular was most commonly rated as the easiest subtopic ($n = 32$, 29.4%).

DISCUSSION

As anatomy assessments evolve to meet the need of larger student numbers and align with revalidated curricula, it

becomes increasingly important for educators to be reliably informed when selecting appropriate assessment approaches (Wass et al., 2001; Chakravarty et al., 2005; Gregory et al., 2009; Smith and Mathias, 2011; Samarasekera et al., 2015). Within modern medical curricula, relying upon the identification of structures alone is no longer considered the most effective way in which to evaluate student's anatomy knowledge (Wass et al., 2001; Samarasekera et al., 2015).

Table 3.

A Summary of Examination Performance by Subtopic

Subtopic	Examination Performance			P-value	Effect Size <i>d</i>
	Cohort (2013–2016) Mean % (\pm SD)	Cohort (2017–2020) ^a Mean % (\pm SD)			
Overall	43.1 (\pm 47.4)	60.6 (\pm 48.9)		<0.0001	–0.36
Cerebrum	41.9 (\pm 47.2)	63.4 (\pm 48.2)		<0.0001	–0.45
Diencephalon	43.9 (\pm 45.8)	67.1 (\pm 47.0)		<0.0001	–0.50
Vascular System	43.2 (\pm 47.3)	50.0 (\pm 49.1)		<0.0001	–0.34
Brainstem/Cranial Nerves	39.7 (\pm 47.4)	58.7 (\pm 49.3)		<0.0001	–0.39
Spinal Cord	52.6 (\pm 49.6)	84.4 (\pm 36.5)		<0.0001	–0.66
Cerebellum	48.1 (\pm 47.6)	37.9 (\pm 48.5)		<0.0001	0.21
Other	47.3 (\pm 48.7)	63.1 (\pm 48.3)		<0.0001	–0.33

Table shows the mean performance (\pm standard deviation) by subtopic for both single best answer (SBA) and open-ended questions (OEQ) at the National Undergraduate Neuroanatomy Competitions.^aExamination performance for the single best answer question cohort include 2017, 2018, and 2020.

Table 4.

Analysis of Open-Ended versus Single Best Answer Questions

Subtopic	Examination Performance		P-value
	Open-Ended Questions Mean % (\pm SD)	Single Best Answer Questions Mean % (\pm SD)	
Overall	44.4 (\pm 19.1)	59.8 (\pm 15.1)	<0.0001
Cerebrum	39.3 (\pm 23.6)	63.3 (\pm 19.2)	<0.0001
Diencephalon	36.5 (\pm 26.6)	66.2 (\pm 27.0)	<0.0001
Vascular System	44.0 (\pm 28.9)	58.8 (\pm 21.1)	<0.0001
Brainstem/Cranial Nerves	48.1 (\pm 26.9)	58.4 (\pm 26.3)	<0.0001
Spinal Cord	41.5 (\pm 32.7)	84.4 (\pm 36.5)	<0.0001
Cerebellum	41.9 (\pm 43.9)	38.5 (\pm 28.5)	0.2214
Other	53.4 (\pm 19.4)	63.6 (\pm 19.0)	<0.0001

A paired analysis of performance data from students completing both single best answer (SBA) and open-ended questions (OEQ) at the National Undergraduate Neuroanatomy Competitions (2017–2020).

However, most anatomists would probably agree that there are certain structures which will always remain on the “need to know” list (Smith and McManus, 2015). It is clear from the core undergraduate syllabi that the amount of knowledge a student should know on graduation has not significantly changed in the UK (Smith et al., 2016). Therefore, robust assessments are required throughout medical school so that students have sufficient anatomical knowledge by the time they reach clinical practice.

Assessment Context

In an age where many anatomists are expressing their concern for the level of knowledge held by graduating doctors, it is important that we understand how changing question structure in this way can impact upon the standards we set for our students. Although anatomy is commonly integrated into many assessment types, most medical schools still host a practical based anatomy examination of some kind held in a dissection room environment (Brenner et al., 2015).

Due to a number of logistical pressures and constraints within anatomy departments, it is becoming common to have components of anatomy assessments that are delivered as SBAs rather than OEQs. This format eases the marking burden but may compromise the academic rigor of assessment strategies. From a cognitive perspective, there is evidence to suggest that answering an SBA question relies more on automatic passive processing, while an OEQs utilizes controlled active processing which is associated with stronger comprehension (Badger and Thomas, 1991; Husain et al., 2012; Ozuru et al., 2013). However, it has also been argued that a carefully crafted, context rich, multiple-choice item also uses cognitive processes that closely parallel OEQ design (Palmer and Devitt, 2007; Hift, 2014). Although both types of question type are commonly used in anatomical science assessments, it has been suggested that the use of SAQs should be considered more suited to those disciplines that require

spontaneous generation of the answer as an essential part of the stimulus (Schuwirth and van der Vleuten, 2004a). This aligns particularly well to bedside or placement style teaching in health science education, where students are expected to be fluent with medical and clinical terminology when engaging professionally with patients.

The UK national context has witnessed a recent General Medical Council drive to reduce the amount of summative assessments in medical curricula and replace them with many more formative opportunities. The academic rigor will of course be influenced by the purpose of the assessment. Formative assessments utilize knowledge testing as a means to promote learning and benchmark progress, while summative papers are there to determine that only students with sufficient levels of knowledge progress within their program of study. The latter requires standard setting procedures which assists in the creation of a pass mark, which is frequently determined by the predicted performance of borderline students within the cohort (Brenner et al., 2015). As a self-selecting national competition there is no expectation of this with the NUNC, other than to follow best practice guidelines with question writing and to ensure anatomical and clinical accuracy. Despite this, standard setting procedures for each practical assessment were undertaken to monitor question level inflation and so that annual results could be meaningfully compared. The number of questions which are considered as part of a core curriculum and those that go beyond it have been quantified annually to maintain consistency (Hall et al., 2017).

Impact of Question Type on Performance

The exploration of this issue has been observed across a number of disciplines in higher education (Badger and Thomas, 1991; Bridgeman, 1992; Wolfson et al., 2001; Ozuru et al., 2013; Wooton et al., 2014; Biria and Dehghan, 2016) and within medical or dental topics (Schuwirth and van der Vleuten,

2004a; Brenner et al., 2015; Thompson and O’Loughlin, 2015; Adamu et al., 2018; Hauer et al., 2020) providing mixed results when student performances were statistically compared or correlated with each other.

The NUNC’s practical element relies entirely on asking participants to identify brain structures, which are non-contextual and mostly not reflective of the full range of questions used in formal anatomy assessments. Despite this, it can be adapted as a tool to compare students’ performance on SBAs and OEQs in a specific subject area without compromising the curriculum. The results of the current investigation demonstrate that students attending the NUNC over its last eight-year duration have performed significantly better on the SBA questions compared to OEQs. When relating these findings within anatomical science our results compare favorably to the observations of Melovitz Vasan et al. (2018), but are not consistent with those of Adamu et al. (2018). Within the field of assessment research more broadly, the picture is somewhat contradictory, with a number of non-significant reports (Bridgeman, 1992; Wolfson et al., 2001). Given this variable history within the published literature, it is worth exploring some of the factors influencing outcomes in some of these settings. The work by Adamu et al. (2018) strongly suggests that performance in each test type is dependent on medical student’s year of study, with more senior students performing better on SBA questions. This theoretical perspective corresponds well with the views of Palmer and Devitt (2007), Schuwirth and van der Vleuten (2004b), and Hift (2014) who indicate that OEQs are not superior to SBAs, especially by the point of exit level summative assessments in medicine. The NUNC is separated into clinical and preclinical categories and so further analysis might well shed some light on this issue. Interestingly, the NUNC is intended for those possessing “high level” neuroanatomical knowledge and there is some evidence to suggest these individuals (with high levels of subject-specific knowledge) are more positively correlated to a stronger SBA performance (Ozuru et al., 2013). The current findings appear to support this hypothesis and highlight the strong possibility of statistical interactions beyond the main effect of the outcome.

Perceptions versus Performance

The existing results show that medical students perform differently when an alternative test method is used for the same topic, and to this end, if used as part of a summative curricular process, may lead to a misleading impression of a student’s knowledge in that area.

Unsurprisingly, the majority of medical students predicted that they would perform better on SBA questions than OEQs in our bespoke survey and this fits well with the existing evidence (Zeidner, 1987). This finding also resonates well with work exploring self-assessment ability of medical students in practical anatomy assessments. Published studies have demonstrated a negative correlation between the students predicted grade and actual grade (Sawdon and Finn, 2014; Hall et al., 2016). This questions the ability of students to accurately self-assess their own performance in pressured environments, and in a vocational subject such as medicine, being aware of knowledge limitations and competency is important for patient safety.

Student attitudes are interesting because they may influence psychological sensitivity toward each of these approaches when used as part of a summative curricular assessment. The current body of evidence indicates that students alter

their learning and revision strategy depending on the nature of type of examination they anticipate undertaking (Simkin and Kuechler, 2005; Stanger-Hall, 2012). These studies further suggest that the variation of response formats will have an impact in anticipatory learning. Understanding why this occurs might be intuitive to some degree; when a student does not know the answer to an OEQ question, it becomes challenging for them to self-generate one because it is reliant on active rather than passive cognitive processing (Ozuru et al., 2013). However, with the same level of uncertainty a student may be able to select the most correct option (utilizing calculated reasoning) from the SBA, which is known as the “cueing effect” (Stiggins and Chappuis, 2005). Over time, students can learn and develop the ability to differentiate between genuine relevant answers and “distractors” due to their common characteristics—something often referred to as “test wiseness” (Cahill and Leonard, 1999). In educational terms, this is considered to be strategic learning approach and contributes to a variance in score that is attributable to the test method and not ability (Bridgeman, 1992; Sam et al., 2016). Despite this, it has also been reported that higher achieving students who engage with each question on a cognitive level are sometimes known to overthink questions (Hubbard et al., 2017) and this may be why previous evidence suggests that they prefer OEQs (Birenbaum and Feldman, 1998). Given the self-selecting nature of the NUNC, this is certainly something that could have affected this study. One approach of mitigating against issues with the reliability are to increase the overall number of SBA questions (Sam et al., 2018).

Although it is well known that poorly constructed SBA questions can lead to alphabetical pattern recognition or correct answers being identified through other characteristics such as being longer or more detailed than their distractors (Evans, 1984), writing good quality SBA questions is now considered somewhat esoteric. The skills and expertise to undertake the task effectively are acquired through training and experience (Evans, 1984; Case and Swanson, 2002). It is not uncommon for administrative teams to provide detailed post assessment analysis of student performance using common performance metrics such as “facility” and most common selected wrong answer to inform educators on how to improve upon their bank of SBA questions.

Neuroanatomy Subtopics

The subtopic and question difficulty distribution between 2013 and 2016 and that of the SBA in 2017–2020 were not significantly different. From a discipline-specific perspective, the subtopic frequently rated as the most difficult was the diencephalon, yet, it generated one of the highest mean performance scores in the NUNC. Conversely, the topic vascular was rated as the easiest but produced one of the lower mean performance scores. There is an existing argument which implies that when a subject area is considered to be challenging it stimulates an increase in time devoted to it, which subsequently impacts upon the depth of knowledge (Thompson and O’Loughlin, 2015; Melovitz Vasan et al., 2018). In this particular study, the SBAs were compared to those from the OEQs (between two modules) and concluded that the improved performance in the OEQs had resulted from a change in learning approach. This would suggest that by altering the assessment type can influence student approaches to learning, which has been historically supported by the literature (Birria and Dehghan, 2016).

Student Views on Modern Anatomy Assessments

The general consensus from participants in our study (and therefore from those participating in the NUNC) was that the timed spotter style assessment was the most effective way of assessing anatomical knowledge. Understandably, students voiced that this type of assessment method is more stressful than others in the curriculum, for the reason that in most cases each station is individually timed, and therefore, they are unable to revisit questions in the same way that they could for either a written MCQ or long structured essay style examinations. There have been some examples of institutions transitioning the traditional anatomy “steeplechase” style assessment to online alternatives. In terms of student perceptions, the removal of the need to physically move around the laboratory and being seated instead reduced anxiety levels, (Inuwa et al., 2011). The current evidence supports the view that there is no detrimental effect in outcomes with a shift in assessment modality of this nature, so educators could safely consider this option without fear of jeopardizing academic rigor (Meyer et al., 2016).

Despite the encouraging work on the efficacy of online assessment approaches the NUNC is one of only a handful of UK-based opportunities where students can demonstrate their motivation and knowledge (above and beyond a standard curriculum syllabus) in a specific field outside of a formal curriculum context. It is also an opportunity to network and physically interact with other enthusiasts and professionals at varying levels of expertise, which is what ultimately makes the NUNC such a valuable learning experience. The combination of physical attendance and individual feedback (which is emailed to all students after each competition) is considered invaluable even in the absence of winning prizes.

Limitations of the Study

It is important to acknowledge that there is clearly a selection bias in this study since students with a special interest in neuroscience would compete and devote more extracurricular time to win this competition. There were significantly more males the OEQ group than the SBA group (66.4% vs. 55.6%, $P = 0.01$). While this may reflect an increasing female attendance over the years which makes the current NUNC more in line with the female dominated medical student population, the early data may be less generalizable to the rest of the medical student population. It may also be less generalizable to other anatomy subject areas—the specialist pathway of neurosurgery tends to be male dominated in the UK. It is also possible that question type could have been influenced by the mean performance between the different subtopics. For example, mean performance in all OEQ cerebrum questions were compared to all SBA cerebrum questions. Like-for-like questions were not compared in this study. Therefore, the suggestion that question difficulty is accounting for some of the differences reported cannot be ruled out when interpreting these results. We recognize the assessments are designed to identify a winner, which does not reflect the purpose of assessments in the medical curriculum, which may impact on the generalizability of the data. However, given the limitations restricting changes to an assessment within the curriculum, we feel the present study uses a reasonable and comparable alternative to assess approaches to assessment within a similar population of students.

CONCLUSIONS

This study provides further contributory evidence suggesting that a change in assessment modality can result in variable outcomes when used as part of a bespoke extracurricular assessment in clinically oriented neuroanatomy (using identify only questions). However, it also endorses the view that performance of students in assessments of this nature can be extremely sensitive to other factors such as familiarity with the test scenario, stage of training, and level of topic-specific knowledge.

It adds some further credence to the view that the use of OEQs enables students to demonstrate deeper understanding, and through the process of generating cognitive errors, can positively impact upon learning, particularly at the earlier stages of training. Therefore, it is suggested that, where possible, medical schools include this question type within assessment formats at some stage during their programs. However, it is also acknowledged that well-written OEQ have the ability to test higher cognitive functions but may be more suited to those with existing levels of integrated knowledge, such as senior students.

Given that clinical practice requires the spontaneous recall of anatomical knowledge in professional conversations, it might be more suitable to use OEQs as the preferred assessment type over SBAs (at least when requesting students to identify essential anatomical structures) despite the fact that such context free questions as these can only ever determine superficial knowledge, which is not representative of the deeper levels of understanding required for clinical competency.

ACKNOWLEDGMENTS

The authors would like to acknowledge and thank Dr David Walker, Laboratory Manager, Centre for Learning Anatomical Sciences and Miss Ellen Adams, Deputy Laboratory Manager, Centre for Learning Anatomical Sciences, for their assistance and technical support in running assessments for NUNC between 2013 and 2020. The authors have no conflicts of interest to declare.

NOTES ON CONTRIBUTORS

WASSIM H. MERZOUGUI, B.M.B.S., is a junior doctor in the Department of Trauma and Orthopedics at Pilgrim Hospital in Boston, United Kingdom. He is an alumnus of the University of Southampton and is involved in organizing the National Undergraduate Neuroanatomy Competition and is interested in leadership, management, and medical education.

MATTHEW A. MYERS, B.M.B.S. (Hons.), M.Med.Sci. (Hons.), is a junior doctor at Salford Royal NHS Foundation Trust, Greater Manchester, United Kingdom. He is involved in organizing the National Undergraduate Neuroanatomy Competition and has been involved in teaching neuroanatomy to undergraduate medical students. In addition to clinical practice, he has a research interest in medical education and the neurosciences. He led the competition in 2018 and has been involved in all aspects of NUNC organization, delivery, and research for over five years.

SAMUEL HALL., B.M., M.Med.Sci., M.R.C.S., is a neurosurgical registrar and honorary teaching fellow in anatomy at Southampton General Hospital in Southampton United Kingdom. His educational and research interests focus on how students learn neuroanatomy, the use of peer assisted learning

and exploring how and why neurophobia manifests itself within undergraduate medical students. He is a cofounder of the NUNC and is a principal investigator of all NUNC-related projects.

AHMAD ELMANSOURI, B.M.B.S., is an anatomy demonstrator at the Department of Medical Education at Brighton and Sussex Medical School in Brighton, United Kingdom. His research interest is in medical education, specifically in the impact of near-peer teaching and technology-enhanced learning. He is the founder of The Wessex Finals Revision Weekend and the Coordinator and presenter for the Healthcare Leadership Academy Podcast. He is a senior member of the NUNC committee and has led on data collection for all strands of research related to NUNC events.

ROB PARKER, is a fifth-year medical student at the University of Southampton School of Medicine, Southampton, United Kingdom. His research interest is in medical education and immunopsychiatry. He led on anatomical dissections and devised questions for the spotter element of NUNC over a period of four years.

ALISTAIR D. ROBSON, M.Med.Sci., is a fifth-year medical student at the University of Southampton School of Medicine, Southampton, United Kingdom. He has a keen interest in technology-enhanced learning resources in neuroanatomy, a theme he developed for the completion of his integrated Masters project. He coordinated the 2020 National Undergraduate Neuroanatomy Competition, and has been involved in data collection and in contributing to all NUNC-related research themes since 2017.

OCTAVIA KURN, B.Med.Sci., is a fourth-year medical student at the University of Southampton School of Medicine, Southampton, United Kingdom. She has led the peer assisted teaching program in anatomy and is a core member of the NUNC committee. She has been involved in data collection and in contributing to all National Undergraduate Neuroanatomy Competition related research since 2017.

RACHEL PARROTT, is an assistant prosector in the Department of Anatomy at St Andrews University, St Andrews, Scotland. She provides dissection training to technical staff and demonstrators as amended by the Human Tissue Authority Act 2006 (Scotland). She has a research interest in education and provides technical support to both academic and technical staff, first-, second-, and third-year medical students in the dissecting room, and perform various technical duties. She is a senior committee member of the National Undergraduate Neuroanatomy Competition, overseeing the organization and running of the laboratory-based activities of the competition.

KATE GEOGHEGAN, B.M.B.S., is a junior doctor working in the Department of Cardiology at Royal United Hospital in Bath, United Kingdom. Her research interest is in medical education, neurology, and global health. She is a core member of the University of Southampton's peer-led neuroanatomy teaching team and coordinated the 2019 National Undergraduate Neuroanatomy Competition.

CHARLOTTE H. HARRISON, B.M.B.S., M.Med.Sci., is a junior doctor working in the Department of Emergency Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom. She is an alumnus of the University of Southampton and has been involved in the organization of neuroanatomy near-peer teaching at the University of Southampton School of Medicine. Her research interests range from medical education to neuropathology. She is the lead for assessment at each NUNC, coordinating and overseeing the

entire marking team and process, while ensuring quality control and correct data entry procedures.

DEEPIKA ANBU, M.Med.Sci., is a fourth-year medical student at the University of Southampton School of Medicine, Southampton, United Kingdom. She has a keen interest in medical education and conducted her masters' study on near-peer teaching and technology-enhanced learning resources in neuroanatomy. She is a core member of the peer-led neuroanatomy teaching team and coordinated the 2020 National Undergraduate Neuroanatomy Competition. She has been involved in data collection and in contributing to all NUNC-related research.

OLIVER DEAN, B.Med.Sci., is a fourth-year medical student at the University of Southampton School of Medicine, Southampton, United Kingdom. Alongside undertaking his academic studies, he teaches anatomy to second-year medical students and is on the committee of the National Undergraduate Neuroanatomy Competition. His research interest is in medical education. He was responsible for the MCQ element of the NUNC, leading a team of question writers, and compiling the MCQ paper.

SCOTT BORDER, B.Sc. (Hons) Ph.D., S.F.A.H.E., F.A.S., N.T.F., is a principal teaching fellow in anatomy in the Centre for Learning Anatomical Sciences at the University of Southampton, Southampton, United Kingdom and Chair of the Education Committee for the Anatomical Society in United Kingdom. His pedagogical focus is on the scholarship of neuroanatomy and head and neck anatomy, working with students as partners in anatomical education, and his research interest is focused on how learning technologies can enhance the student experience and improve knowledge gain. He is a cofounder of NUNC and senior author of all NUNC-related research work, including managing all contributions to the current project.

LITERATURE CITED

- Adamu LH, Yusuf AH, Taura MG. 2018. Differences in multiple choice and short answer questions performance among medical and dental trainees as reflected in their anatomy course. *J Contemp Med Educ* 6:7-16.
- Ali A, Khan ZN, Konczalik W, Coughlin P, El Sayed S. 2015. The perception of anatomy teaching among UK medical students. *Bull Roy Coll Surg Engl* 97:397-400.
- Badger E, Thomas B. 1991. Open-ended questions in reading. *Pract Assess Res Eval* 3:4.
- Birenbaum M, Feldman RA. 1998. Relationships between learning patterns and attitudes towards two assessment formats. *Educ Res* 40:90-98.
- Biria RE, Dehghan MA. 2016. The relationship between self-regulated learning and test method format: The case of multiple choice and open-ended questions in L2 reading. *J Glob Res Educ Soc Sci* 7:160-167.
- Brenner E, Chirculescu AR, Reblet C, Smith C. 2015. Assessment in anatomy. *Eur J Anat* 19:105-124.
- Bridgeman B. 1992. A comparison of quantitative questions in open-ended and multiple-choice formats. *J Educ Meas* 29:253-271.
- Cahill DR, Leonard RJ. 1999. Missteps and masquerade in American medical academe: Clinical anatomists call for action. *Clin Anat* 12:220-222.
- Case SM, Swanson DB. 2002. *Constructing Written Test Questions for the Basic and Clinical Sciences*. 3rd Ed. Philadelphia, PA: National Board of Medical Examiners (NBME). 180 p.
- Chakravarty M, Latif NA, Abu-Hijleh MF, Osman M, Dharap AS, Ganguly PK. 2005. Assessment of anatomy in a problem-based medical curriculum. *Clin Anat* 18:131-136.
- Chang BS, Molnár Z. 2015. Practical neuroanatomy teaching in the 21st century. *Ann Neurol* 77:911-916.
- Chapman SJ, Hakeem AR. 2015. Undergraduate anatomy teaching: Evaluations and conclusions. *Acad Med* 90:399.
- Choudhury B, Freemont A. 2017. Assessment of anatomical knowledge: Approaches taken by higher education institutions. *Clin Anat* 30:290-299.
- Choudhury B, Gouldsborough I, Shaw FL. 2016. The intelligent anatomy spotter: A new approach to incorporate higher levels of Bloom's taxonomy. *Anat Sci Educ* 9:440-445.

- Collins J. 2006. Education techniques for lifelong learning: Writing multiple-choice questions for continuing medical education activities and self-assessment modules. *Radiographics* 26:543–551.
- Entwistle NJ, Entwistle A. 1991. Contrasting forms of understanding for degree examinations: The student experience and its implications. *High Educ* 22:205–227.
- Evans W. 1984. Test wiseness: An examination of cue-using strategies. *J Exp Educ* 52:141–144.
- Geoghegan K, Payne DR, Myers MA, Hall S, Elmansouri A, Parton WJC, Harrison CH, Stephens J, Parker R, Rae S, Merzougui W, Nagy E, Venkatesh P, Parrott R, Border S. 2019. The national undergraduate neuroanatomy competition: Lessons learned from partnering with students to innovate undergraduate neuroanatomy education. *Neuroscientist* 25:271–280.
- Gogalniceanu PF, O'Connor EF, Raftery A. 2009. Undergraduate anatomy teaching in the UK. *Bull Roy Coll Surg Engl* 91:102–106.
- Gregory JK, Lachman N, Camp CL, Chen LP, Pawlina W. 2009. Restructuring a basic science course for core competencies: An example from anatomy teaching. *Med Teach* 31:855–861.
- Hall S, Myers M, Stephens J, Seaby E, Lowry A, Parton W, Gesteria Andrade M, Arofrin V, McElligott S, Harrison C, Parrott R, Border S. 2017. Are medical students achieving learning outcomes in neuroanatomy? Mapping performance in the national undergraduate neuroanatomy competition with a published neuroanatomy syllabus. *J Anat* 230:364.
- Hall S, Stephens JR, Myers MA, Elmansouri A, Geoghegan K, Harrison CH, Nagy E, Anbu D, Parton WJ, Payne DR, Seaby E. 2020. The career impact of the national undergraduate neuroanatomy competition. *World Neurosurg* 133:e535–e539.
- Hall SR, Stephens JR, Seaby EG, Andrade MG, Lowry AF, Parton WJ, Smith CF, Border S. 2016. Can medical students accurately predict their learning? A study comparing perceived and actual performance in neuroanatomy. *Anat Sci Educ* 9:488–495.
- Hauer KE, Boscardin C, Brenner JM, van Schaik SM, Papp KK. 2020. Twelve tips for assessing medical knowledge with open-ended questions: Designing constructed response examinations in medical education. *Med Teach* 42:880–885.
- Heylings DJ. 2002. Anatomy 1999-2000: The curriculum, who teaches it and how? *Med Educ* 36:702–710.
- Hift RJ. 2014. Should essays and other “open-ended”-type questions retain a place in written summative assessment in clinical medicine? *BMC Med Educ* 14:249.
- Hubbard JK, Potts MA, Couch BA. 2017. How question types reveal student thinking: An experimental comparison of multiple-true-false and free-response formats. *CBE Life Sci Educ* 16:26.
- Husain H, Bais B, Hussain A, Samad SA. 2012. How to construct open ended questions. *Procedia Soc Behav Sci* 60:456–462.
- Inuwah IM, Taranikanti V, Al-Rawahy M, Habbal O. 2011. Perceptions and attitudes of medical students towards two methods of assessing practical anatomy knowledge. *Sultan Qaboos Univ Med J* 11:383–390.
- Mehta B, Bhandari B, Sharma P, Kaur R. 2016. Short answer open-ended versus multiple-choice questions: A comparison of objectivity. *Ann Nat Acad Med Sci (India)* 52:173–182.
- Melovitz Vasan CA, DeFouw DO, Holland BK, Vasan NS. 2018. Analysis of testing with multiple choice versus open-ended questions: Outcome-based observations in an anatomy course. *Anat Sci Educ* 11:254–261.
- Memon I. 2018. Cadaver dissection is obsolete in medical training! A misinterpreted notion. *Med Princ Pract* 27:201–210.
- Meyer AJ, Innes SI, Stomski NJ, Armson AJ. 2016. Student performance on practical gross anatomy examinations is not affected by assessment modality. *Anat Sci Educ* 9:111–120.
- Monrouxe LV, Grundy L, Mann M, John Z, Panagoulas E, Bullock A, Mattick K. 2017. How prepared are UK medical graduates for practice? A rapid review of the literature 2009–2014. *BMJ Open* 7:e013656.
- O'Keeffe GW, Davy S, Barry DS. 2019. Radiologist's views on anatomical knowledge amongst junior doctors and the teaching of anatomy in medical curricula. *Ann Anat* 223:70–76.
- Ozuru Y, Briner S, Kurby CA, McNamara DS. 2013. Comparing comprehension measured by multiple-choice and open-ended questions. *Can J Exp Psychol* 67:215–227.
- Palmer EJ, Devitt PG. 2007. Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple-choice questions? Research paper. *BMC Med Educ* 7:49.
- Papa V, Vaccarezza M. 2013. Teaching anatomy in the XXI century: New aspects and pitfalls. *Sci World J* 2013:310348.
- Ramsden P. 1997. The context of learning in academic departments. *Ex Learn* 2:198–216.
- Rowland S, Ahmed K, Davies DC, Ashrafian H, Patel V, Darzi A, Paraskeva PA, Athanasiou T. 2011. Assessment of anatomical knowledge for clinical practice: Perceptions of clinicians and students. *Surg Radiol Anat* 33:263–269.
- Sam AH, Field SM, Collares CF, van der Vleuten CP, Wass VJ, Melville C, Harris J, Meeran K. 2018. Very-short-answer questions: Reliability, discrimination and acceptability. *Med Educ* 52:447–455.
- Sam AH, Hameed S, Harris J, Meeran K. 2016. Validity of very short answer versus single best answer questions for undergraduate assessment. *BMC Med Educ* 16:266.
- Samarasekera DD, Gopalakrishnakone P, Gwee MC. 2015. Assessing anatomy as a basic medical science. In: Chan LK, Pawlina W (Editors). *Teaching Anatomy: A Practical Guide*. 1st Ed. New York, NY: Springer International Publishing. p 279–289.
- Sambell K, Brown S, McDowell L. 1997. "But is it fair?": An exploratory study of student perceptions of the consequential validity of assessment. *Stud Eval* 23:349–371.
- Sawdon M, Finn G. 2014. The “unskilled and unaware” effect is linear in a real-world setting. *J Anat* 224:279–285.
- Sbayeh A, Qaedi Choo MA, Quane KA, Finucane P, McGrath D, O'Flynn S, O'Mahony SM, O'Tuathaigh CM. 2016. Relevance of anatomy to medical education and clinical practice: Perspectives of medical students, clinicians, and educators. *Perspect Med Educ* 5:338–346.
- Schuwirth LW, van der Vleuten CP. 2004a. Changing education, changing assessment, changing research? *Med Educ* 38:805–812.
- Schuwirth LW, van der Vleuten CP. 2004b. Different written assessment methods: What can be said about their strengths and weaknesses? *Med Educ* 38:974–979.
- Simkin MG, Kuechler WL. 2005. Multiple-choice tests and student understanding: What is the connection? *Decis Sci J Innovat Educ* 3:73–98.
- Smith CF, Finn GM, Stewart J, Atkinson MA, Davies DC, Dyball R, Morris J, Ockelford C, Parkin I, Standing S, Whiten S. 2016. The Anatomical Society core regional anatomy syllabus for undergraduate medicine. *J Anat* 228:15–23.
- Smith CF, Mathias HS. 2011. What impact does anatomy education have on clinical practice? *Clin Anat* 24:113–119.
- Smith CF, McManus B. 2015. The integrated anatomy practical paper: A robust assessment method for anatomy education today. *Anat Sci Educ* 8:63–73.
- Stanger-Hall KF. 2012. Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE Life Sci Educ* 11:294–306.
- Stiggins R, Chappuis J. 2005. Using student-involved classroom assessment to close achievement gaps. *Theor Into Pract* 44:11–18.
- Sugand K, Abrahams P, Khurana A. 2010. The anatomy of anatomy: A review for its modernization. *Anat Sci Educ* 3:83–93.
- Thomas PR, Bain JD. 1984. Contextual dependence of learning approaches: The effects of assessment. *Hum Learn* 3:227–240.
- Thompson AR, O'Loughlin VD. 2015. The Blooming anatomy tool (BAT): A discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. *Anat Sci Educ* 8:493–501.
- Tirpude AP, Gaikwad M, Tirpude PA, Jain M, Bora S. 2019. Retrospective analysis of prevalent anatomy spotter's examination: An educational audit. *Korean J Med Educ* 31:115–124.
- Trautman J, McAndrew D, Craig SJ. 2019. Anatomy teaching stuck in time? A 10-year follow-up of anatomy education in Australian and New Zealand medical schools. *Aust J Educ* 63:340–350.
- Turney BW. 2007. Anatomy in a modern medical curriculum. *Ann Roy Coll Surg Engl* 89:104–107.
- Wass V, Van der Vleuten C, Shatzer J, Jones R. 2001. Assessment of clinical competence. *Lancet* 357:945–949.
- Wolfson PJ, Veloski JJ, Robeson MR, Maxwell KS. 2001. Administration of open-ended test questions by computer in a clerkship final examination. *Acad Med* 76:835–839.
- Wooten MM, Cool AM, Prather EE, Tanner KD. 2014. Comparison of performance on multiple-choice questions and open-ended questions in an introductory anatomy laboratory. *Phys Rev ST Phys Educ Res* 10:020103.
- Yaqinuddin A, Zafar M, Ikram MF, Ganguly P. 2013. What is an objective structured practical examination in anatomy? *Anat Sci Educ* 6:125–133.
- Yim M. 2018. Comparison of results between modified-Angoff and bookmark methods for estimating cut score of the Korean medical licensing examination. *Korean J Med Educ* 30:347–357.
- Zeidner M. 1987. Essay versus multiple-choice type classroom exams: The student's perspective. *J Educ Res* 80:352–358.