

9-2013

Equivalence of students' scores on timed and untimed anatomy practical examinations.

Guiyan Zhang
Thomas Jefferson University

Bruce Fenderson
Thomas Jefferson University

Richard R Schmidt
Thomas Jefferson University

J J. Veloski
Thomas Jefferson University

Follow this and additional works at: <https://jdc.jefferson.edu/pacbfp>

 Part of the [Anatomy Commons](#), and the [Education Commons](#)

[Let us know how access to this document benefits you](#)

Recommended Citation

Zhang, Guiyan; Fenderson, Bruce; Schmidt, Richard R; and Veloski, J J., "Equivalence of students' scores on timed and untimed anatomy practical examinations." (2013). *Department of Pathology, Anatomy, and Cell Biology Faculty Papers*. Paper 174.

<https://jdc.jefferson.edu/pacbfp/174>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Pathology, Anatomy, and Cell Biology Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.

ASE-12-0133

Research Report

Equivalence of Students' Scores on Timed and Untimed Anatomy Practical Examinations

Guiyun Zhang^{1,*}, Bruce A. Fenderson¹, Richard R. Schmidt¹, J. Jon Veloski²

¹Department of Pathology, Anatomy and Cell Biology, Jefferson Medical College, Thomas Jefferson University Philadelphia, Pennsylvania

²Center for Research in Medical Education and Health Care, Jefferson Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania

Abbreviated title: Anatomy Practical Examinations

*Correspondence to: Dr. Guiyun Zhang, Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, 1020 Locust Street, Room 263 Jefferson Alumni Hall, Philadelphia, PA 19107. USA. E-mail: guiyun.zhang@jefferson.edu

Phone: 215-503-7019; Fax: 215-923-3808

Abstract

Untimed examinations are popular with students because there is a perception that first impressions may be incorrect, and that difficult questions require more time for reflection. In this report, we tested the hypothesis that timed anatomy practical examinations are inherently more difficult than untimed examinations. Students in the Doctor of Physical Therapy program at Thomas Jefferson University were assessed on their understanding of anatomic relationships using multiple-choice questions. For the class of 2012 (n = 46), students were allowed to circulate freely among 40 testing stations during the 40-minute testing session. For the class of 2013 (n = 46), students were required to move sequentially through the 40 testing stations (one minute per item). Students in both years were given three practical examinations covering the back/upper limb, lower limb, and trunk. An identical set of questions was used for both groups of students (untimed and timed examinations). Our results indicate that there is no significant difference between student performance on untimed and timed examinations (final percent scores of 87.3 and 88.9, respectively). This result also held true for students in the top and bottom 20th percentiles of the class. Moreover, time limits did not lead to errors on even the most difficult, higher-order questions (i.e., items with p-values < 0.70). Thus, limiting time at testing stations during an anatomy practical examination does not adversely affect student performance.

Key words: gross anatomy education, anatomy assessment, practical examination, Doctor of Physical Therapy

Introduction

Despite a reduction in time devoted to cadaver dissection among U.S. medical schools (Heylings, 2002; Gartner, 2003; Drake et al., 2009; Rizzolo et al., 2010), gross anatomy continues to serve as a foundation for the preclinical basic science curriculum (Granger, 2004; Prideaux, 2005; Böckers et al., 2010). During dissection, students confront mortality and begin to appreciate the generosity of the donor's gift for the benefit of their medical education. Cadaver dissection also promotes teamwork and professionalism (Escobar-Poni and Poni 2006; Böckers et al., 2010). Above all, cadaver dissection provides medical students and allied health professionals with a unique opportunity to gain a deep understanding of human form and function via a hands-on learning experience.

Various approaches have been developed to assess students' understanding of anatomic structures and their complex relationships (Chakravarty et al., 2005; Chan and Ganguly, 2008; Hofer et al., 2011; Shaibah and van der Vleuten, 2012). For example, computer-based (virtual) anatomy examinations have been introduced at some medical schools (Krippendorf et al., 2008; Inuwa et al., 2011, 2012). At other institutions, where the basic medical sciences are integrated with clinical medicine, objective structured practical examinations (OSPEs) have been introduced to assess students' comprehensive understanding of structures, functions, and related clinicopathologic findings (Schoeman and Chandratilake, 2012; Yaqinuddin et al., 2012).

Jefferson Medical College, of Thomas Jefferson University follows a traditional format for gross anatomy instruction based on overview lectures and cadaver dissection. Our laboratory practical

examinations follow the so-called “steeplechase” format (Chirculescu et al., 2007), in which students encounter a series of faculty-generated questions based on prosected specimens, bones, and images. Students are asked to name (or interpret) key anatomic structures or features that are identified by arrows, paint, colored pins, and/or loops of red string. The traditional approach for assessing student knowledge during practical examinations has involved free-response questions. Although free-response (open-ended) questions have greater face validity than multiple-choice questions (Fenderson et al., 1997), hand-written examinations in this format are more difficult to score accurately (Krippendorf et al., 2008; Shaibah and van der Vleuten, 2012). For example, students may write “superior laryngeal nerve” instead of “internal laryngeal nerve”. Such minor differences in terminology lead to variation in grading due to different points of view among the faculty that introduces errors in students’ scores.

To address issues of reliability and practicality, alternative formats for scoring practical examinations have been reported (Krippendorf et al., 2008; Inuwa et al., 2011, 2012). A recent study by Shaibah and van der Vleuten (2012) confirmed the validity of multiple-choice questions as an alternative to free-response questions in anatomy. Previous studies indicated that mean scores are higher on multiple-choice questions, compared to free response questions (Ward, 1982; Veloski et al., 1993). However, multiple-choice questions appear to assess the same set of knowledge and abilities as free-response questions. Moreover, students’ rank order and score distributions are not affected by the use of multiple-choice questions on anatomy examinations (Shaibah and van der Vleuten, 2012). We have also been using multiple-choice questions for anatomy practical examinations at Jefferson Medical College. The combination of multiple-choice questions and optical mark read answer sheets has improved our assessment process,

enabling us to reduce (or eliminate) human coding errors. Machine-graded, examinations also facilitate the generation of valuable psychometric and statistical data regarding overall student performance.

During a practical examination at Jefferson, anatomy students are typically required to stay at a particular testing station for one minute before moving to the next station in sequence. We believe that this system provides sufficient time for the prepared student to examine a specimen and answer a relevant question. However, we realize that examinations are inherently stressful for many students (Ng et al., 2003a; 2003b; Singh et al., 2012), and that time limitations add additional stress that may undermine student confidence and/or performance. Students frequently complain that they need more time to answer difficult test questions, arguing that the “one size fits all” approach is fatally flawed. To resolve this issue, we analyzed the effects of time limits on student performance. Use of the multiple-choice testing format enabled us to collect data on student performance from two similar groups of students, who received a set of either untimed or timed anatomy practical examinations.

Methods

Doctor of Physical Therapy Students at Jefferson

The Doctor of Physical Therapy (DPT) training program at Jefferson consists of an intensive 3-year, full-time curriculum. During their first year, DPT students are required to complete a lecture and laboratory dissection course in human gross anatomy. In this study, we followed the performance of students in the class of 2012 (n=46) and class of 2013 (n=46) as they received either untimed or timed practical examinations. The Institutional Review Board at Thomas Jefferson University reviewed our research proposal and approved the collection of aggregate student data for this study. The admission requirements for the two classes of Jefferson students were similar and data indicate that both groups were comparable in terms of their demographic and academic background (Table 1).

Anatomy Course Offered to Doctor of Physical Therapy Students

The human anatomy course for Doctor of Physical Therapy students at Jefferson consists of 36 hours of theoretical lecture and 78 hours of laboratory dissection. Students are required to pass both components of the course. The lecture portion of the course is divided into 4 blocks: 1) back and upper limb, 2) lower limb, 3) trunk, and 4) head and neck. At the end of each block, a multiple-choice written examination is administered to evaluate the students' understanding of the lecture-based theoretical material. The advanced anatomy laboratory course is divided into 3 blocks: 1) back and upper limb, 2) lower limb, and 3) trunk.

Practical Examinations

Practical examinations are given at the end of each dissection block. Percentage weights for these examinations are: 35%, 35%, and 30%, respectively. Scores on examinations then determine the final course grade. For each examination, the faculty generated 40 multiple-choice (single best answer) questions that focused on osteology (15% of questions), cadaver items (65% of questions), and diagnostic images (20% of questions). Specific anatomical structures or features were tagged on prosected cadavers using colored pins and red strings. Approximately 15-20% of the questions were higher-order questions that required students to integrate their understanding of anatomic structure/function, or address an important clinical correlation.

Experimental Design

Physical therapy students in the class of 2012 and 2013 were given an identical set of multiple-choice anatomy questions (40 questions per examination). The class of 2012 was allowed to move freely from station-to-station at their own pace. They were at liberty to revisit stations to confirm or change their answer. They had 40 minutes to complete the examination. By contrast, the class of 2013 was assigned to a station number before they entered the room and they began the examination at their assigned station. Six “rest stations” were introduced so as to accommodate 46 students in one testing session. Students in the class of 2013 spent one minute at each testing station and they moved sequentially. They were not allowed to revisit testing stations.

Data Analysis

The performance of students on untimed or timed anatomy practical examinations was analyzed and compared using independent t-tests. We compared student scores on block examinations, as well as their overall final course grades. We also examined the performance of students in the top 20% and bottom 20% of their class. To address the issue of item difficulty (one size fits all criticism), we compared student performance on 10 difficult questions with proportions correct (p-values) below 0.70. These questions had been constructed as higher-order test questions that required a deeper understanding of human gross anatomy. Calculations were performed using Stata Statistical Software: Release 11.0., Stat Corp., College Station, TX.

Results

We followed the performance of PT anatomy students in the class of 2012 (n = 46) and class of 2013 (n = 46) as they received either untimed or timed practical examinations. The median test reliability (Coefficient Alpha) for three block examinations administered to the class of 2012 was 0.56. The median reliability for the same three block examinations administered to the class of 2013 was 0.60. These results indicate that the reliability of these examination scores was sufficient to support the analysis of differences between the two groups of students.

Comparison of Untimed and Timed Groups

Table 2 shows the comparison of the overall final course grades between the untimed (class of 2012) and timed (class of 2013) groups. The results indicate that there is no significant difference between student performance on untimed and timed examinations. Final percent numerical grades were 87.3 and 88.9, respectively.

Comparison of Student Performance on Individual Block Examinations

Table 3 summarizes student performance on the three block examinations that covered: back and upper limb, lower limb, and trunk. Differences in student performance between the untimed and timed examination groups were not apparent on examinations that covered the upper and lower limbs (Table 3). However, the average score of the timed group on the exam of the trunk was significantly ($p < 0.005$) higher than the untimed group.

Effect of Time Limits on Performance of Marginal Students

Limits on time for problem solving during a practical examination could affect the performance of weaker (marginal) students, who are generally believed to be more vulnerable to the deleterious effects of heightened emotional stress. It seemed reasonable to us that this group of students would preferentially benefit by having more time to answer questions (untimed protocol). Contrary to our expectations, average scores of marginal students in the bottom 20th percentile of the class did not differ significantly between timed (80.9%) and untimed (80.0%) groups. This lack of difference between groups was also noted for students in the top 20th percentile of the class: 95.0% (timed examination) and 94.0% (untimed examination).

Comparison of Student Performance on Most Difficult Questions

To further explore the effect of time constraints on problem solving during anatomy practical examinations, we analyzed student performance on difficult items with p-values <0.70. We reasoned that students taking untimed examinations would score higher on difficult questions because they had more time for problem solving. Mean scores for the 10 most difficult items were compared and the results are shown in Table 4. Differences between groups taking untimed and timed examinations were not statistically significant: 68% (timed group) versus 63% (untimed group). These data suggest that time limits do not affect problem solving, even when students consider complicated, higher-order test questions. Although differences between untimed and timed groups were noted for many of the difficult questions (Table 4), there was no consistent pattern to these differences. Indeed, on 4 of the 10 most difficult test questions, students given timed examinations scored higher than those receiving untimed examinations.

Discussion

Anatomy is an important rite of passage for medical students. Students are confronted with the face of mortality (Escobar-Poni and Poni, 2006; Sugand et al., 2010) and are required to rapidly learn a new language of anatomy and medicine (Educational Affairs Committee, American Association of Clinical Anatomists, 1996; Hanwell et al., 2007). In a traditional curriculum, where anatomy is offered at the beginning of the academic year, heightened emotional and academic stress can be challenging for many students. Previous studies have demonstrated that students possess higher levels of stress prior to examinations, as monitored by significantly increased levels of salivary cortisol (Ng et al., 2003a; 2003b; Singh et al., 2012). A study on stress associated with anatomy practical examinations demonstrated that the concentration of salivary cortisol was at its highest level prior to the first examination (Lester et al., 2010). Empathic concern and mentoring at this point in the medical curriculum are essential. At the same time, faculty must provide reasonable assessments of student knowledge that are objective and valid.

This study was designed to address a recurring criticism from medical students and allied health professionals that timed examinations are inherently stressful and unfair. Informal polling of our students indicates an overwhelming preference for untimed practical examinations. Our results do not support the students' contention that timed examinations are more difficult than untimed examinations. Rather, our results demonstrate that there is no significant difference between student performance on untimed and timed practical examinations (Table 2). This result also holds true for students in the top as well as the bottom 20th percentiles of the class.

We were surprised that untimed examinations did not benefit students in the bottom 20th percentile of the class. We believed that this group would benefit by having more time to consider anatomic relationships and formulate answers to questions. This was not the case: restricting time at testing stations during an anatomy practical examination did not adversely affect student performance.

To clarify the effect of time on student performance we analyzed the 10 most difficult questions on our examinations (i.e., items with p-values <0.70). These questions were judged by the faculty to be either second-order questions or questions that probed sophisticated concepts. We believed that students taking untimed examinations would score higher on difficult questions (more time for reflection and problem solving). Our results (Table 4) do show differences between untimed and timed groups, but in different directions (i.e., there is no pattern). Indeed, on 4 of the 10 most difficult test questions, students given timed examinations scored higher than those receiving untimed examinations (Table 4). This may be explained by variations in learning between the classes. Overall, our results demonstrate that placing reasonable time limits on problem solving in the cadaver laboratory does not lead to errors, nor does it reduce student performance on even the most difficult questions (Table 4).

Differences in student performance between the untimed and timed examination groups were apparent on the block examination that covered the trunk (Table 3). The average grade of the timed group (Class 2013) on the exam of the trunk was statistically higher than the untimed group (Class 2012). This was not observed for block examinations that covered the upper and

lower limbs (Table 3). We do not believe that this difference was related to the issue of timing. Our explanation for this finding is that students in the class of 2013 had a more profound understanding and knowledge of the trunk. Indeed, students in the class 2012 informed us that they did not believe the anatomy of the trunk was a major focus for physical therapy, and so they preferred to concentrate their study on the upper and lower limbs.

In summary, our data demonstrate that placing time limits on problem solving during an anatomy practical examination does not adversely affect student performance. Based on these data, we prefer the use of timed examinations at Jefferson Medical College. Having students move every minute, from one station to the next, allows us to run our practical examinations smoothly and efficiently. This is particularly important when administering an examination to hundreds of students at the same time. Timed examinations also eliminate student crowding around stations that present difficult questions, making timed examinations easier to proctor.

NOTES ON CONTRIBUTORS

GUIYUN ZHANG, M.D., Ph.D., is an assistant professor in the Department of Pathology, Anatomy, and Cell Biology at Thomas Jefferson University, Philadelphia, Pennsylvania. She is Director for the Advanced Human Anatomy Dissection course for Physical Therapy students, and also teaches anatomy, histology, and neuroscience to the first year medical students and allied health professionals.

BRUCE A. FENDERSON, Ph.D., is a professor in the Department of Pathology, Anatomy, and Cell Biology at Thomas Jefferson University, Philadelphia, Pennsylvania. He is Director of the Foundations of Pathology course for second year medical students. He also teaches anatomy and histology to first year medical students and allied health professionals.

RICHARD R. SCHMIDT, Ph.D. is a professor and Vice Chair in the Department of Pathology, Anatomy, and Cell Biology at Thomas Jefferson University, Philadelphia, Pennsylvania. He is Director of the Human Form and Development (gross anatomy) course for first year medical students. He also teaches histology and neuroscience to medical students and allied health professionals.

J. JON VELOSKI, M.S., is Director of Medical Education Research in the Center for Research in Medical Education and Health Care at Jefferson Medical College of Thomas Jefferson University, Philadelphia, Pennsylvania.

LITERATURE CITED

Böckers A, Jerg-Bretzke L, Lamp C, Brinkmann A, Traue HC, Böckers TM. 2010. The gross anatomy course: An analysis of its importance. *Anat Sci Educ* 3:3–11.

Chan LK, Ganguly PK. 2008. Evaluation of small-group teaching in human gross anatomy in a Caribbean medical school. *Anat Sci Educ* 1:19-22.

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.

Chirculescu AR, Chirculescu M, Morris JF. 2007. Anatomical teaching for medical students from the perspective of European Union enlargement. *Eur J Anat* 11:S63-S66.

Drake RL, McBride JM, Lachman N, Pawlina W. 2009. Medical education in the anatomical sciences: The winds of change continue to blow. *Anat Sci Educ* 2:253–259.

Educational Affairs Committee, American Association of Clinical Anatomists. 1996. A clinical anatomy curriculum for the medical students of the 21st century: gross anatomy. *Clin Anat* 9:71-99

Escobar-Poni B, Poni ES. 2006. The role of gross anatomy in promoting professionalism: A neglected opportunity. *Clin Anat* 19:461–467.

Fenderson BA, Damjanov I, Robeson MR, Veloski JJ, Rubin E. 1997. The virtues of extended matching and uncued tests as alternatives to multiple choice questions. *Human Path* 28:526–532.

Gartner LP. 2003. Anatomical sciences in the allopathic medical school curriculum in the United States between 1967-2001. *Clin Anat* 16:434–439.

Granger NA. 2004. Dissection laboratory is vital to medical gross anatomy education. *Anat Rec* 281B:6–8.

Hanwell SM, Atkinson M, Davies DC, Dyball R, Morris J, Ockleford C, Parkin I, Standring S, Whiten S, Wilton J. 2007. A core syllabus in anatomy for medical students-Adding common sense to need to know. *Eur J Anat* 11:3-18.

Heylings DJ. 2002. Anatomy 1999-2000: The curriculum, who teaches it and how? *Med Educ* 36:702–710.

Hofer RE, Nikolaus OB, Pawlina W. 2011. Using checklist in a gross anatomy laboratory improves learning outcomes and dissection quality. *Anat Sci Educ* 4:249-255.

Inuwa 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.

Inuwa IM, Taranikanti V, Al-Rawahy M, Habbal O. 2012. Anatomy practical examinations: How does student performance on computerized evaluation compare with the traditional format? *Anat Sci Educ* 5:27–32.

Krippendorf BB, Bolender DL, Kolesari GL. 2008. Computerized grading of anatomy laboratory practical examinations. *Anat Sci Educ* 1:220–223.

Lester SR, Brown JR, Aycock JE, Grubbs SL, Johnson RB. 2010. Use of saliva for assessment of stress and its effect on the immune system prior to gross anatomy practical examinations. *Anat Sci Educ* 3:160–167.

Ng V, Koh D, Mok BY, Chia SE, Lim LP. 2003a. Salivary biomarkers associated with academic assessment stress among dental undergraduates. *J Dent Educ* 67:1091-1094

Ng V, Koh D, Chia SE. 2003b. Examination stress, salivary cortisol, and academic performance. *Psychol Rep* 93:1133-1134

Prideaux D. 2005. Think global, act regional: Promoting change in medical education. *Med Educ* 39:756–757.

Rizzolo LJ, Rando WC, O'Brien MK, Haims AH, Abrahams JJ, Stewart WB. 2010. Design, Implementation, and Evaluation of an Innovative Anatomy Course. *Anat Sci Educ* 3:109-120.

Schoeman S, Chandratilake M. 2012. The anatomy competence score: a new marker for anatomical ability. *Anat Sci Educ* 5:33-40.

Shaibah HS, van der Vleuten CP. 2012. The validity of multiple choice practical examinations as an alternative to traditional free response examination formats in gross anatomy. *Anat Sci Educ* (in press; doi: 10.1002/ase.1325).

Singh R, Goyal M, Tiwari S, Ghildiyal A, Nattu SM, Das S. 2012. Effect of examination stress on mood, performance and cortisol levels in medical students. *Indian J Physiol Pharmacol* 56:48-55.

Sugand K, Abrahams P, Khurana A. 2010. The anatomy of anatomy: A review for its modernization. *Anat Sci Educ* 3:83–93.

Veloski JJ, Rabinowitz HK, Robeson MR. 1993. A solution to the cueing effects of multiple choice questions: The Un-Q format. *Med Educ* 27:371-375.

Ward WC. 1982. A comparison of free-response and multiple-choice forms of verbal aptitude tests. *Appl Psychol Meas* 6:1-11.

Yaqinuddin A, Zafar M, Ikram MF, Ganguly P. 2012. What is an objective structured practical examination in anatomy? *Anat Sci Educ* (in press; doi: 10.1002/ase.1305).