ASE-16-0125.R1

Research report

Climbing Bloom's Taxonomy Pyramid:

Lessons from a Graduate Histology Course

Nikki B. Zaidi[1], Charles Hwang[2], Sara Scott[2], Stefanie Stallard[2], Joel Purkiss[1,3,4], Michael Hortsch[3,5]*

[1]Office of Medical Student Education, University of Michigan Medical School, Ann Arbor, Michigan

[2]Universty of Michigan Medical School, Ann Arbor, Michigan

[3]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, Michigan

[4]Office of the Curriculum, School of Medicine, Baylor College for Medicine, Houston, Texas

[5]Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, Michigan

Running title: Bloom's Taxonomy Histology Tool

*Correspondence to: Dr. Michael Hortsch, Department of Cell and Developmental Biology, University of Michigan Medical School, 109 Zina Pitcher Place, Ann Arbor, MI 48109, USA. E-mail: hortsch@umich.edu

ABSTRACT

Bloom's taxonomy was adopted to create a subject-specific scoring tool for histology multiple-choice questions (MCQs). This Bloom's Taxonomy Histology Tool (BTHT) was used to analyze teacher- and student-generated quiz and examination questions from a graduate level histology course. Multiple-choice questions using histological images were generally assigned a higher BTHT level than simple text questions. The type of microscopy technique (light or electron microscopy) used for these image-based questions did not result in any significant differences in their Bloom's taxonomy scores. The BTHT levels for teacher-generated MCQs correlated positively with higher discrimination indices and inversely with the percent of students answering these questions correctly (difficulty index), suggesting that higher-level Bloom's taxonomy questions differentiate well between higher- and lower-performing students. When examining BTHT scores for MCQs that were written by students in a Multiple-Choice Item Development Assignment (MCIDA) there was no significant correlation between these scores and the students' ability to answer teacher-generated MCQs. This suggests that the ability to answer histology MCQs relies on a different skill set than the aptitude to construct higher-level Bloom's taxonomy questions. However, students significantly improved their average BTHT scores from the midterm to the final MCIDA task, which indicates that practice, experience and feedback increased their MCQ writing proficiency.

Key words: histology education, medical education, graduate education, assessment, microscopic anatomy, Bloom's taxonomy, multiple choice questions

INTRODUCTION

Bloom's Taxonomy is widely used in educational research to stratify learning activities into different cognitive levels (Miller et al., 1991; Kim et al., 2012; Thompson and O'Loughlin, 2015; Morton and Colbert-Getz, 2017). It categorizes cognitive activities into six hierarchical levels that range from basic recall to higher educational objectives such as application and synthesis (Bloom, 1956). Bloom's Taxonomy has been adopted as a valuable tool for examining students' learning and to classify examination questions based on the cognitive levels and skills the questions are attempting to assess. Over time, the original version has evolved and modified versions have been published (Anderson et al., 2001; Krathwohl, 2002). However, even these modified versions of Bloom's taxonomy are often too general to serve as useful tools for specific subject areas. Therefore, educational researchers have created specialized adaptations of Bloom's taxonomy for assessing student performance and rating educational tasks within specific fields, such as the biomedical sciences (Su et al., 2005; Plack et al., 2007; Crowe et al., 2008; Phillips et al., 2013; Thompson and O'Loughlin, 2015).

As medical education continues to evolve, it is important to evaluate the effectiveness of new didactic strategies and learning methods by assessing student learning. A common method of assessment in medical education is the use of multiple-choice questions (MCQ) in examinations (Case and Swanson, 2002; Haladyna et al., 2002). Although there are challenges associated with MCQ assessments, it is commonly accepted that MCQs can be used to test a variety of Bloom's taxonomy performance levels (Aiken, 1982; Morrison and Free, 2001; Brady, 2005; Palmer and Devitt, 2007; Clifton and Schriner, 2010; Tiemeier et al., 2011). A wealth of information is available to aid with the writing of efficient and fair MCQs (Case and Swanson, 2002; Haladyna at al., 2002; McCoubrie, 2002), especially for use in medical examinations (Downing, 2005; Golda, 2011). Ideally, MCQs are written to assess higher-order thinking skills. However, achieving this goal can be difficult (Bissell and Lemons, 2006). Nevertheless, there is general agreement that higher-level examination questions foster a deeper understanding of the material by the learner (Winne, 1979; Burns, 2010; Jensen et al., 2014).

Another approach that is used to elicit critical thinking by students has been described by Fellenz and is now known as multiple-choice item development assessment (MCIDA) (Fellenz, 2004). Instead of answering teacher-generated MCQs, students are asked to generate their own MCQs from the material they encountered in prior didactic sessions. Students not only have to create new questions and provide a correct answer, but they must also justify the questions and answers they have created. This requires students not only to recall learned facts, but also to use them in new and creative ways, which itself represents a higher-level cognitive activity.

In the CDB450/550 histology course at the University of Michigan, both of the above techniques were utilized to assess students' learning. Both undergraduate and graduate students enrolled in this course were asked to answer teacher-generated MCQs. In addition, graduate-level students were also asked to complete a MCIDA task at two different time points of the course. There is limited research that compares the effectiveness and the relationship between students' ability to answer traditional teacher-generated MCQs with students' ability to create MCQs in an MCIDA task (Foos, 1989; Belanich et al., 2004).

Being a subject with a central visual component, histology or microanatomy presents its own distinct challenge when creating, answering, and evaluating MCQs. Therefore, based on a previously published Blooming Anatomy Tool (BAT) (Thompson and O'Loughlin, 2015), a unique Bloom's taxonomy-based rubric - a Bloom's Taxonomy Histology Tool (BTHT) - was created for the purpose of evaluating histology MCQs. Together with other evaluation parameters, this new BTHT resource will help educators teaching histology to assess the didactic level of histology MCQs and to formulate more challenging examination questions that go beyond a simple recall task. It can also serve as a research resource to better understand the relationship between the ability of students to answer histology MCQs versus to create them. To test this hypothesized relationship, teacher- and student-generated MCQs from a graduate-level histology course at the University of Michigan were analyzed and questions were categorized according to their Bloom's level by assigning a BTHT score. These scores were examined in terms of how they correlate with students' course performance. Specifically, students' ability to answer teacher-generated MCQs was compared with students' aptitude to generate high Bloom's taxonomy level questions.

MATERIALS AND METHODS

Structure of the "Through the Looking Glass – From Stem Cells to Tissues and Organs" Histology Course

The CCDB450/550 course entitled "Through the Looking Glass – From Stem Cells to Tissues and Organs" is a graduate-level histology class at the University of Michigan in Ann Arbor, MI, that is offered once a year during the Winter term to undergraduate students in junior or senior standing and to graduate students at any level. The course is modeled after the first-year medical school histology component and consists of 25 two-hour lectures and two review sessions covering the histology of all basic tissues, major human organs and organ systems (UMMS, 2016). After the first one-hour lecture, which introduces a topic/organ/organ system, the virtual slides on the course website are introduced to the class in another 30 to 40-minute lecture-style presentation. Subsequently, all students are expected to study the virtual slides on the course's website (UMMS, 2016) on their own time. Students also had access to several types of supplementary learning material that are described by Holaday et al. (2013). The data analyzed in this manuscript cover the years 2011 to 2014. Over this time period the overall syllabus, the course content, student evaluation and grading policy, and the principal faculty instructors teaching in the course remained largely unchanged.

Examination of Students' Histology Knowledge in the CDB450/550 Course

Undergraduate students who enrolled at the CDB450 level were graded solely based on their performance in six short online MCQ quizzes and two longer online MCQ examinations (one midterm examination and one final examination), which resulted in approximately 180 assessment questions. These questions evaluate students' knowledge and understanding of the course material, as well as their skill of recognizing histological structures. The quizzes and examinations were timed (90 to 120 seconds per questions) and open-book with the exclusion of Internet use. Graduate students and a small number of undergraduate students enrolled at the CDB550 course level were required to take the same quizzes and examinations as CDB450 students and had an additional assignment of creating five MCQs covering the first half of the course and a second set of five MCQs covering the second half of the course. Grading of these student-generated MCQs was guided by the following set of rules: (1) No two submitted questions may be derived from the same lecture topic; (2) All questions must have only one undisputable correct answer; (3) Four of the five questions must be based on images of the student's choosing; (4) The sources of all images must be acknowledged; (5) Only one question may be a simple identification problem; (6) Only one question may have a true/false format, and (7) All questions must include a short justification for the correct answer. Students received no further training or instructions in writing MCQs other than the feedback they received for their five submitted midterm MCQs, which explained why they might not have received full credit for their questions.

For course grades, a strategy based on the University of Michigan Medical School was adopted. A student performance under 75% was considered a failing performance. The University of Michigan Rackham Graduate School considers any grade of C+ and below as failing. Borderlines between other letter grades were adjusted from year to year, but never differed by more than 2% during the four-year period covered by this study.

Student Demographics

The sample for this study included 51 students enrolled at the undergraduate level (CDB450) and 71 students enrolled at the graduate level (CDB550) during the 2011 to 2014 academic years. Of the undergraduate students, 33 were female and 18 were male, whereas 32 of the graduate students were female and 39 were male. All students included in this study completed all evaluations and the entire course. The majority of undergraduates enrolled were either pre-medical or pre-dental students. Graduate students were usually enrolled in biomedical Master or Ph.D. programs, specifically biomedical engineering; physiology; oral health sciences; molecular, cellular and developmental biology; environmental health sciences; epidemiology and others.

Statistical Analysis of Data

All student- and teacher-generated questions were independently analyzed and scored by three second-year medical students, who had successfully completed the first year histology component of the University of Michigan Medical School curriculum. We conducted a retrospective analysis of how the BTHT tool performed by examining the patterns and associations in student performance on MCQs across levels of BTHT scores. All statistical analyses were conducted using SPSS statistical package, version 22 (IBM Corp., Armonk, NY). To examine associations among raters' scores for both student-generated MCQs and teacher-generated MCQs, the inter-rater reliability for BTHT scores was determined using Cohen's Kappa (Cohen, 1960; Stemler, 2004; McHugh, 2012). To examine graduate and undergraduate students' performance on teacher-generated MCQs and how graduate students performed on the midterm compared to the final MCIDA task, independent-samples *t*-tests were performed. Pearson Correlation Coefficient R was used to examine whether raters' BTHT scores for student-generated MCQs correlated with students' examination scores for answering teacher-generated MCQs.

The project received an Institutional Review Board (IRB) exemption from the University of Michigan medical IRB panel (application number HUM00091932).

RESULTS

Generation of a Bloom's Taxonomy Tool for Histology Multiple-Choice Questions

Based on a previously published Blooming Anatomy Tool (BAT) (Thompson and O'Loughlin, 2015), a Bloom's taxonomy-type scoring system was developed to differentiate among different cognitive levels of histology MCQs (Table 1). This tool was developed with feedback from the participating medical student raters (C.H., S.S., and S.S.), who previously had completed the histology component of the M1 year before participating in this retrospective study. After several rounds of modifications, a five-level scoring rubric was judged by all raters to be most practical for allowing a reproducible and well-defined discrimination between different levels of histology MCQs. Level 1 questions only require a simple recall performance, whereas level 5 questions force students to remember and critically judge multiple facts in order to decide and predict a possible outcome of a complex, often clinical scenario. All higher-level BTHT questions typically involve a multi-step solution process. Table 2 displays a series of example MCQs that represent the five levels of the BTHT resource, including short justifications for their assigned BTHT scores.

Subsequently, the BTHT, as outlined in Tables 1 and 2, was used to evaluate 180 teacher-generated MCQs and 710 student-generated MCQs. The student-generated MCQs were submitted as part of two required MCIDA tasks by students participating in the graduate CDB550 course level at the University of Michigan. Table 3 displays an analysis of inter-rater reliability of BTHT scores. For both groups of questions, the Cohen's Kappa between all three scorers is significant at a $P < 0.01$ level. A comparison of Cohen's Kappa inter-rater reliability scores (Table 3) indicates that raters' BTHT grades display a moderate level of agreement for student-generated MCQs and a substantial level of agreement for teacher-generated MCQs (Landis and Koch, 1977).

Analysis of Teacher-Generated Histology Multiple-Choice Questions

Both undergraduate and graduate students had to answer all 180 teacher-generated MCQs, which were divided into six smaller quizzes and two larger midterm and final examinations. The 51 undergraduate students scored a cumulative mean of 83.46% for all quizzes and examinations, whereas the 71 graduate students scored a cumulative mean of 88.96% (Table 4). This difference between the two means was found to be highly significant with a medium effect size (Table 4). A paired-samples t-test of these data was conducted to compare course grades in the first half (including the midterm examination) and the second half of the course for both graduate students and undergraduate students. For graduate students, there was a significant decline (2.37%) in the scores for the first half of course compared to the second half of course; $t(70) = 2.980$, $P = 0.004$. Likewise, for undergraduate students, there was also a significant drop in the scores (3.52%) for the first half of course compared to the second half of course; $t(50) = 3.168$ $P = 0.003$.

Overall, the three raters assigned the 180 teacher-generated questions an average BTHT score of 2.16 with a ±SD of 0.12. A subsequent analysis of image-based questions versus text-only questions revealed that image-based questions had a higher mean BTHT score (N = 145, M = 2.43 ±0.56) than text-only questions (N = 35, M = 1.04 ±0.13). An independent *t*-test demonstrated this difference to be significant ($P < 0.001$) with a large effect size (Cohen's $d = 3.42$). A further analysis differentiating between different types of images, specifically light micrographs, electron micrographs, and graphic representations of histological structures, did not indicate a statistically significant difference in BTHT scores for these three image-type groups (not shown).

Since the quality of an MCQ is often judged by its discrimination and its difficulty index (Kelley, 1939; Moussa et al., 1991; Meshkani and Hossein Abadie, 2005; Clifton and Schriner, 2010), the BTHT scores for all teacher-generated MCQs were correlated with these two measures as derived from students' results in the course quizzes and examinations. This analysis uncovered a small, but statistically significant (r = 0.25; P = 0.001) correlation between the average raters' BTHT scores and the discrimination index. Moreover, a small, inverse correlation was also found between the average raters' BTHT scores for all teacher-generated questions and their difficulty indices (r = -0.22; P = 0.003).

Analysis of Student-Generated Histology Multiple-Choice Questions

A total of 710 student-generated MCQs were analyzed using the BTHT resource. Each student who registered at the CDB550 course level in the years 2011 to 2014 (n = 71) was required to submit five newly written MCQs at the time of the midterm examination and an additional five MCQs after the final examination. The overall average BTHT scores for the 10 MCQs submitted by each student ranged from 2.07 to 3.33. There was an increase in raters' BTHT scores for student-generated MCQs submitted at the midterm examination (average midterm BTHT score of 2.68 ±0.30) when compared to those submitted at the final examination (average BTHT score 2.87 ±0.37). This difference was statistically highly significant (P < 0.000; t = -4.30; df = 70).

To address the question whether students' ability to write higher-level Bloom's MCQs correlated with their ability to answer teacher-generated MCQs, the average BTHT scores for all 71 sets of student-generated MCQs were correlated with students' cumulative quiz and examination results. This analysis did not indicate any statistically significant association between the students' ability to answer teacher-generated MCQs and students' ability to create high-level Bloom's score MCQs (r = -0.08; P = 0.507).

DISCUSSION

The new BTHT will help histology educators evaluate the cognitive levels associated with MCQs in their histology examinations and aid them in constructing new higher-level questions. This tool can also help to elucidate how students learn and which cognitive abilities are important for both writing and solving MCQs. The analysis that is presented in this study suggests that the experience of the person(s) generating the questions might sometimes influence and occasionally limit the effectiveness of a Bloom's taxonomy-style tool. The raters, who evaluated MCQs submitted by the students enrolled at the CDB550 course level, reported that student-generated questions were sometimes overly verbose, more ambiguous, less focused, contained more unnecessary distractors, and often made suboptimal use of the images linked to the questions. In comparison, the raters found that the teacher-generated questions were easier to score, which is evidenced by the higher correlation coefficient values (Table 3). It should be noted that due to the grading strategy applied to this course, the teacher-generated questions had lower overall BTHT scores when compared to the student-generated questions. Nevertheless, this finding is consistent with other studies that looked at the influence of MCQ writer experience, training and feedback on various aspects of MCQ item quality (Jozefowicz et al., 2002;

Naeem et al., 2012; Sadaf et al., 2012; Meyari and Beiglarkhani, 2013; Webb et al., 2015).

Use of the Bloom's Taxonomy Histology Tool for the Analysis of Histology Multiple-Choice Questions

Different parameters are being used in evaluating the effectiveness of MCQs. Specifically, discrimination and difficulty indices are common measures to determine whether examination questions discriminate between high- and low-performing students (Kelley, 1939; Moussa et al., 1991; Meshkani and Hossein Abadie, 2005; Clifton and Schriner, 2010). However, these two parameters represent different aspects of a test question's efficacy and only exhibit a moderate, non-linear correlation with each other (Sim and Rasiah, 2006; Mitra et al., 2009; Karelia et al., 2013). Neither the discrimination nor the difficulty index provides information about the cognitive requirements involved in solving an examination question (Kibble and Johnson, 2011). This makes them incomplete and moderately useful measures of test item quality (Pyrczak, 1973; Notebaert, 2017). A well-written test question will discriminate between high- and low-performing students based on the learners' mastery of the material and their ability to apply it to new situations. In this context, the BTHT provides a valuable additional quantifier for the quality of histology MCQs, thereby extending the usual measures derived from a standard item analysis.

Histology has an important visual component and the analysis and interpretation of micrographic images are major challenges for many students

(Loo et al., 1995; Harris et al., 2001; Kumar et al., 2006; Mione et al., 2016). By definition, images almost automatically move MCQs beyond the lowest cognitive level as defined by the BTHT (Table 2). The new BTHT resource places an emphasis on the importance of histology images when evaluating learning success. In creating the BTHT resource and using it for MCQ analysis, the researchers assumed that the images utilized for examination questions had not been used during previous didactic sessions and therefore represented novel material to the learner. Otherwise, an examination question might be reduced to a simple image recall task, which would be categorized as a low level Bloom's cognitive activity. Therefore, image recall was not considered in the BTHT grading scheme. For these reasons, reusing images should be avoided in histology examinations that are designed to test actual histology knowledge and relevant analytical and synthetic abilities of students.

Skills Needed to Solve Histology Questions versus Skills that Support the Creation of High-Level Histology Questions

Because the new BTHT was not available at the time when students took the CDB450/550 course in the years 2011 to 2014, the student-generated MCQs were not scored using this new grading resource. Student-generated MCQs were graded according to a set of rules defined in the course syllabus and summarized in this paper's Material and Methods section. However, several of these rules encouraged and rewarded the writing of higher-level BTHT questions (e.g., inclusion of images, requirement for multiple-step questions instead of simple identification etc.). Although student-generated MCQs were not scored according to their BTHT level, the analysis of midterm versus final student-submitted questions indicates a clear improvement in the BTHT quality of the student-generated questions. This suggests that the feedback provided to the students, as well as the practice and experience gathered from constructing the first set of questions was helpful in developing the skills necessary to write higher-level BTHT MCQs. Part of this improvement may also be attributed to students developing a level of familiarity with histology as the course progressed. Many students require some time to become comfortable with histology, especially if it is a new and unfamiliar subject to them, and as a result, they are initially challenged (Hortsch and Mangrulkar, 2015).

The BTHT analysis of student-generated MCQs demonstrated no correlation with the same students' ability to answer teacher-generated questions. The actual act of writing MCQs is itself a higher-level Bloom's task and requires a detailed knowledge of the material usually well beyond a simple recall ability. In contrast, answering MCQs often only requires lower- to middle-Bloom's level activities. Some of the skills needed to do well in both tasks most certainly overlap, such as a general mastery of the course material. However, it appears that being good at answering MCQs does not always translate into being a good MCQ writer. In contrast, Foss (1989) reported that students who were assigned to write multiple-choice or essay questions in an introductory psychology class outperformed non-writers on the regular course tests. Although this observation may be partially explained by the additional exposure to the course material for question writers, it nevertheless suggests that MCIDA tasks are helpful in elevating students' proficiency with the course material to higher levels and in fostering higher-order thinking skills. This conclusion is also supported by two more recent studies (Belanich et al., 2004; Bottomley and Denny, 2011). This study's finding that students' ability to answer teacher-generated MCQs does not correlate with their ability to generate higher-level MCQs warrants further investigation. It does not exclude that students who are adept at writing higher-level BTHT MCQs outperform classmates in answering higher BTHT-level, teacher-generated questions. The overall level of teacher-generated questions in this analysis is in the low to mid-level BTHT range (2.16). Another variable that might contribute to the difference in the ability of solving versus creating MCQs are time restrictions, which students face during classroom examinations. Assuming that students started the MCIDA task well before the submission deadline, the MCIDA task had no such constraint. Also, when writing new MCQs, students were able to choose topics they felt comfortable in tackling. In contrast, when answering examination questions, the course director decides about the content and students have no influence on the topics addressed by these questions. Additional research is needed to identify specific parameters, abilities, and skills that are involved in writing versus solving MCQ histology problems and to test for more specific correlations and interdependencies between these activities.

Limitations of the Study

Because a few undergraduate students registered for the course at the graduate level, the reported difference between graduate and undergraduate students in answering teacher-generated MCQs may be an overestimation (Table 4). These subscribers to the CDB550 course version are usually more academically advanced undergraduate students. In addition, considering the findings reported by Foss (1989) that suggest writing test questions enhances a student's ability to answer examination questions, the activity of the CDB550 students writing MCQs for the midterm and the final examination might have elevated their performance over time on the quizzes and the final examination. This may have also resulted in the smaller decrease in average graduate student examination scores for the second half of the course when the histology of more complex organ systems was taught.

Although the proposed BTHT provides a useful resource for evaluating histology MCQs, the limitations of this tool should be noted. The experience of the question writer will influence the fidelity of BTHT scores. Other scoring mechanisms can also provide additional and complementary information about the quality and effectiveness of the question asked and the intellectual demands required to solve it.

CONCLUSIONS

This study presents a new, subject-specific rating tool for histology MCQs that is rooted in Bloom's taxonomy. The BTHT and the results reported will allow educators and educational researchers to reproducibly grade histology MCQs according to their cognitive level and to create more challenging examination problems. Although the ability of solving MCQs is not correlated with the ability to write high-level MCQs, feedback, experience and practice appear to foster the creation of more challenging histology MCQs. In addition, the incorporation of images that are new to the learner is often an effective method of elevating histology MCQs to higher Bloom's taxonomy levels. The BTHT complements standard parameters of analyzing MCQ item quality, such as differentiation and difficulty indices, and may help educators to better understand the cognitive processes that are involved in answering and in writing high-level MCQs for histology.

ACKNOWLEDGEMENTS

The authors report no conflicts of interest and they alone are responsible for the content and writing of the paper. The authors would like to acknowledge the support of Ms. Jill Miller and the entire staff in the UMMS Evaluation and Assessment office and thank Ms. Sarah Hortsch for her diligent proofreading of the manuscript.

NOTES ON CONTRIBUTORS

NIKKI BIBLER ZAIDI, Ph.D., is Associate Director of Evaluation and Assessment in the Office of Medical Student Education at the University of Michigan Medical School in Ann Arbor, Michigan. She has worked in various roles within medical education for nearly ten years. Her primary research interests include developing novel assessment and evaluation tools and processes, as well as examining the reliability and validity of measurement scores.

CHARLES HWANG, B.S., is a third-year medical student at the University of Michigan Medical School. He is interested in the introduction of technology into classrooms and the development of learning tools geared towards improving learning efficiency. Other interests include the elucidation of inflammatory pathways in human pathology, particularly in regards to heterotopic ossification and other sequelae of burn injury.

SARA SCOTT, B.S., is a third-year medical student at the University of Michigan Medical School. She is interested in primary care and improving medical student education.

STEFANIE STALLARD, B.A., is a third-year medical student at the University of Michigan Medical School. She spends much of her time advocating for her classmates, both in regards to academics and the learning environment.

Research interests include deciphering how glioblastoma multiforme (GBM) evades the immune response and mechanisms to bolster the immune system's ability to combat GBM.

JOEL PURKISS, Ph.D., is an assistant professor in the Department of Internal Medicine and Assistant Dean for Evaluation, Assessment and Education Research in the Office of the Curriculum, Baylor College of Medicine in Houston, Texas. Previously he was Director of Evaluation and Assessment in the Office of Medical Student Education at the University of Michigan Medical School and a Research Investigator in the Department of Learning Health Sciences. His research interests are in medical education curriculum evaluation and improvement, as well as in the prediction of medical education performance outcomes.

MICHAEL HORTSCH, Ph.D., is an associate professor in the Departments of Cell and Developmental Biology and of Learning Health Sciences at the University of Michigan Medical School in Ann Arbor, Michigan. Since 1991 he has taught medical and dental histology at the University of Michigan. He is a recipient of the 2012 Kaiser Permanente Award for Excellence in Pre-Clinical Teaching from the University of Michigan Medical School and the 2013 University of Michigan Provost's Teaching Innovation Prize. He is interested in the development of novel electronic teaching tools and how these new resources impact students' learning.

LITERATURE CITED

Aiken LR. 1982. Writing multiple-choice items to measure higher-order educational-objectives. Educ Psychol Meas 42:803–806.

Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, Wittrock MC. 2001. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. 1st Ed. New York City, NY: Longman. 336 p.

Belanich J, Wisher RA, Orvis KL. 2004. A question-collaboration approach to web-based learning. Am J Dist Educ 18:169–185.

Bissell AN, Lemons PP. 2006. A new method for assessing critical thinking in the classroom. BioScience 56:66–72.

Bloom BS (Editor). 1956. Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. 1st Ed. New York, NY: David McKay Co. 201 p.

Bottomley S, Denny P. 2011. A participatory learning approach to biochemistry using student authored and evaluated multiple-choice questions. Biochem Mol Biol Educ 39:352–361.

Brady AM. 2005. Assessment of learning with multiple-choice questions. Nurse Educ Pract 5:238–242.

Burns ER. 2010. "Anatomizing" reversed: Use of examination questions that foster use of higher order learning skills by students. Anat Sci Educ 3:330–334.

Case SM, Swanson DB. 2002. Constructing Written Test Questions for the Basic and Clinical Sciences. 3rd Ed. Philadelphia, PA: National Board of Medical Examiners. 180 p. URL: http://www.nbme.org/pdf/itemwriting_2003/2003iwgwhole.pdf [accessed 3 January 2017].

Cohen J. 1960. A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46.

Clifton SL, Schriner CL. 2010. Assessing the quality of multiple-choice test items. Nurse Educ 35:12–16.

Crowe A, Dirks C, Wenderoth MP. 2008. Biology in bloom: Implementing Bloom's taxonomy to enhance student learning in biology. CBE Life Sci Educ 7:368–381.

Downing SM. 2005. The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. Adv Health Sci Educ Theory Pract 10:133–143.

Fellenz MR. 2004. Using assessment to support higher level learning: The multiple choice item development assignment. Assess Eval High Educ 29:703–719.

Foos PW. 1989. Effects of student-written questions on student test-performance. Teach Psychol 16:77–78.

Golda SD. 2011. A case study of multiple-choice testing in anatomical sciences. Anat Sci Educ 4:44–48.

Harris T, Leaven T, Heidger P, Kreiter C, Duncan J, Dick F. 2001. Comparison of a virtual microscope laboratory to a regular microscope laboratory for teaching histology. Anat Rec 265:10–14.

Haladyna TM, Downing SM, Rodriguez MC. 2002. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ 15:309–334.

Holaday L, Selvig, D, Pukiss J, Hortsch M. 2013. Preference of interactive electronic versus traditional learning resources by University of Michigan medical students during the first year histology component. Med Sci Educ 23:607–619.

Hortsch M, Mangrulkar RS. 2015. When students struggle with gross anatomy and histology: A strategy for monitoring, reviewing, and promoting student academic success in an integrated preclinical medical curriculum. Anat Sci Educ 8:478–483.

Jensen JL, McDaniel MA, Woodard SM, Kummer TA. 2014. Teaching to the test…or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. Educ Psychol Rev 26:307–329.

Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. 2002. The quality of in-house medical school examinations. Acad Med 77:156–161.

Karelia BN, Pillai AM, Vegada BN. 2013. The levels of difficulty and discrimination indices and relationship between them in four-response type multiple choice questions of pharmacology summative tests of Year II M.B.B.S students. Int e-J Sci Med Educ 7:41–46.

Kelley TL. 1939. The selection of upper and lower groups for validation of test items. J Educ Psychol 30:17–24.

Kibble JD, Johnson T. 2011. Are faculty predictions or item taxonomies useful for estimating the outcome of multiple-choice examinations? Adv Physiol Educ 35:396–401.

Kim MK, Patel RA, Uchizono JA, Beck L. 2012. Incorporation of Bloom's taxonomy into multiple-choice examination questions for a pharmacotherapeutics course. Am J Pharm Educ 76:114.

Krathwohl DR. 2002. A revision of Bloom's taxonomy: An overview. Theory Pract 41:212–218.

Kumar RK, Freeman B, Velan GM, De Permentier PJ. 2006. Integrating histology and histopathology teaching in practical classes using virtual slides. Anat Rec 289B:128–133.

Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159–174.

Loo SK, Freeman B, Moses D, Kofod M. 1995. Fabric of life: The design of a system for computer-assisted-instruction in histology. Med Teach 17:269–276.

McCoubrie P. 2004. Improving the fairness of multiple-choice questions: A literature review. Med Teach 26:709–712.

McHugh ML. 2012. Interrater reliability: The kappa statistic. Biochem Med (Zagreb) 22:276–282.

Meshkani Z, Hossein Abadie F. 2005. Multivariate analysis of factors influencing reliability of teacher made tests. J Med Educ 6:149–152.

Meyari A, Beiglarkhani M. 2013. Improvement of design of multiple choice questions in annual residency exams by giving feedback. Strides Dev Med Educ 10:109–118.

Miller DA, Sadler JZ, Mohl PC, Melchiode GA. 1991. The cognitive context of examinations in psychiatry using Blooms taxonomy. Med Educ 25:480–484.

Mione S, Valcke M, Cornelissen M. 2016. Remote histology learning from static versus dynamic microscopic images. Anat Sci Educ 9:222–230.

Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. 2009. The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. Int e-J Sci Med Educ 3:2–7.

Morrison S, Free KW. 2001. Writing multiple-choice test items that promote and measure critical thinking. J Nurs Educ 40:17–24.

Morton DA, Colbert-Getz JM. 2017. Measuring the impact of the flipped anatomy classroom: The importance of categorizing an assessment by Bloom's taxonomy. Anat Sci Educ (in press; doi: 10.1002/ase.1635).

Moussa MA, Ouda BA, Nemeth A. 1991. Analysis of multiple-choice items. Comput Meth Programs Biomed 34:283–289.

Naeem N, van der Vleuten C, Alfaris EA. 2012. Faculty development on item writing substantially improves item quality. Adv Health Sci Educ Theory Pract 17:369–376.

Notebaert AJ. 2017. The effect of images on item statistics in multiple choice anatomy examinations. Anat Sci Educ 10:68–78.

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.

Phillips AW, Smith SG, Straus CM. 2013. Driving deeper learning by assessment: An adaptation of the revised Bloom's taxonomy for medical imaging in gross anatomy. Acad Radiol 20:784–789.

Plack MM, Driscoll M, Marquez M, Cuppernull L, Maring J, Greenberg L. 2007. Assessing reflective writing on a pediatric clerkship by using a modified Bloom's Taxonomy. Ambul Pediatr 7:285–291.

Pyrczak F. 1973. Validity of the discrimination index as a measure of item quality. J Educ Meas 10:227–231.

Sadaf S, Khan S, Ali SK. 2012. Tips for developing a valid and reliable bank of multiple choice questions (MCQs). Educ Health (Abingdon) 25:195–197.

Sim SM, Rasiah RI. 2006. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. Ann Acad Med Singapore 35:67–71.

Stemler SE. 2004. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Practical Assess Res Eval 9:1–11.

Su WM, Osisek PJ, Starnes B. 2005. Using the revised Bloom's taxonomy in the clinical laboratory: Thinking skills involved in diagnostic reasoning. Nurse Educ 30:117–122.

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.

Tiemeier AM, Stacy ZA, Burke JM. 2011. Using multiple choice questions written at various Bloom's taxonomy levels to evaluate student performance across a therapeutics sequence. Innovat Pharm 2:41.

UMMS. 2016. University of Michigan Medical School. Michigan Histology and Virtual Microscopy Learning Resources: Looking Glass Schedule. University of Michigan Medical School, Ann Arbor, MI. URL: http://histology.sites.uofmhosting.net/looking-glass-schedule [accessed 3 January 2017].

Webb EM, Phuong JS, Naeger DM. 2015. Does educator training or experience affect the quality of multiple-choice questions? Acad Radiol 22:1317–1322.

Winne PH. 1979. Experiments relating teachers' use of higher cognitive questions to student-achievement. Rev Educ Res 49:13–49.

Table 1. Bloom's Taxonomy Histology Tool (BTHT)

Bloom's Taxonomy Histology Tool Score:	1	2	3	4	5
Key skills assessed:	Recall	Explain, identify	Apply, connect	Analyze, classify	Predict, judge, critique, decide
Types of histological information assessed:	Basic definitions, facts, and terms.	Basic understanding of architectural organization of histological features and concepts (connective tissue, muscle tissue, neural tissue, etc.). Interpretation and organization of organs or cell types from novel images confined to **single** cell type/structure.	Visual identification in new situations by applying acquired knowledge. Additional functional or structural knowledge about the cell/tissue is also required.	Visual identification and analysis of **comprehensive** additional knowledge. Connection between structure and function confined to **single** cell type/structure.	Interactions **between different** cell types/tissues to predict relationships; judge and critique knowledge of multiple cell types/tissues at same time in new situations. Potential to use clinical judgment to make decisions.
Characteristics of multiple-choice questions:	Only requires recall. Students may memorize answer without understanding the process. Knowing the "what", but not understanding the "why".	Requires recall and comprehension of facts. Image questions asking to identify a structure/cell type without requiring a full understanding of the relationship of all parts. The process of identification requires student to evaluate internal or external contextual clues without requiring knowledge of functional aspects.	Two-step questions that require image-based identification as well as the application of knowledge (e.g., identify structure and know function/ purpose).	Students must call upon multiple independent facts and properly join them together. May be required to correctly analyze accuracy of multiple statements in order to elucidate the correct answer (e.g., generally answer choices with "I & II" or "I & II & III"). Also evaluate all options/ understand all steps and can't rely on simple recall.	Use information in a **new** context with the possibility for a clinical judgment. Students are required to go through multiple steps and apply those connections to a situation, e.g., predicting an outcome or diagnosis or critiquing a suggested plan.
Equivalent level of Bloom's taxonomy:	Knowledge	Comprehension	Application	Analysis	Synthesis/Evaluate

Table 2. Example Multiple-Choice Questions for Bloom's Taxonomy Histology Tool Levels

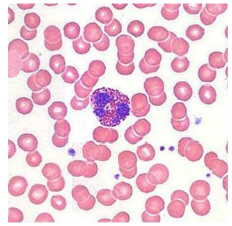
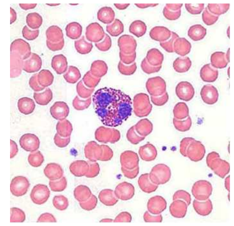
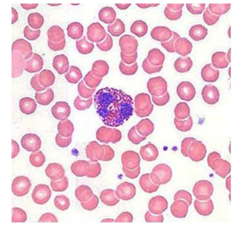
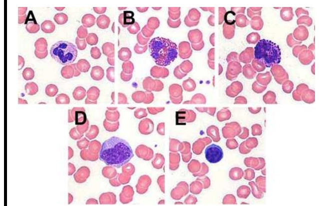
Bloom's Taxonomy Histology Tool Score:	**1**	**2**	**3**	**4**	**5**
Sample multiple-choice questions:	The major function of an eosinophil cell is _________? A. Phagocytosis B. Secretion of antibodies C. Mediation of allergic/inflammatory reactions D. Anti-bacterial Correct answer: C. Identify a function of an eosinophil cell.	 The leukocyte depicted in the image is a ___________? A. Lymphocyte B. Monocyte C. Eosinophil D. Neutrophil Correct answer: C. Recognize the red granules as typical for an eosinophil.	 The leukocyte depicted in the image … A. releases its specific granules in a hypersensitivity reaction, which can lead to anaphylactic shock. B. produces antibodies. C. functions primarily to combat bacterial infections. D. mediates inflammatory/ allergic reactions. Correct answer: D. Identify the cell as an eosinophil and one of its functions.	 Which of the following functions is/are associated with the depicted leukocyte? I. Release its specific granules in a hypersensitivity reaction, which can lead to anaphylactic shock. II. Anti-parasitic activities. III. Production of antibodies. IV. Primarily combats bacterial infections. V. Mediation of inflammatory/ allergic reactions. A. I and III B. II and V C. II and IV D. I and V E. III and IV F. Only II Correct answer: B. The cell is an eosinophil, which has both anti-parasitic and inflammatory /allergic functions.	 A patient complains of fatigue and occasional shortness of breath. A blood sample is taken from which it is determined that the erythrocyte and platelet counts are **NORMAL**. Differential counts of the leukocyte types shown are as follows: Panel A: 55%; Panel B: 15%; Panel C: 1%; Panel D: 8%, Panel E: 21%. Based on this information, what is likely the cause of the patient's symptoms? A. Anemia B. Asthma/respiratory allergies C. Lymphoid leukemia with metastasis to the lungs D. Pneumococcal pneumonia (bacterial infection of the lungs) Correct answer: B. The count for eosinophil cells is too high (normally 1-5%) indicating an ongoing allergic reaction. Identify the different cell types, know their normal abundance in a peripheral blood count, identify the abnormal cell concentration, know the function of the identified cell type and correlate it with the pathological symptoms shown by the patient.
Justification for scoring the example question:	Requires only basic knowledge of eosinophil function.	Students must be able to visually identify an eosinophil in a new image.	Student identifies the histological slide and is prompted to recall a functional detail of the organ/cell. Two independent steps are required. Students must correctly identify the cell as an eosinophil and then also correctly identify a function of eosinophil cells.	Combo options. Student identifies the tissue/cell and then must individually evaluate several possible functions that are associated with this cell.	Students must be able to recognize five types of leukocytes in addition to knowing their normal abundance and function of each type. Students must also bridge the clinical manifestations of histological scenarios. Multiple steps are required.

Table 3. Inter-Rater Reliability for Bloom's Taxonomy Histology Tool Scores

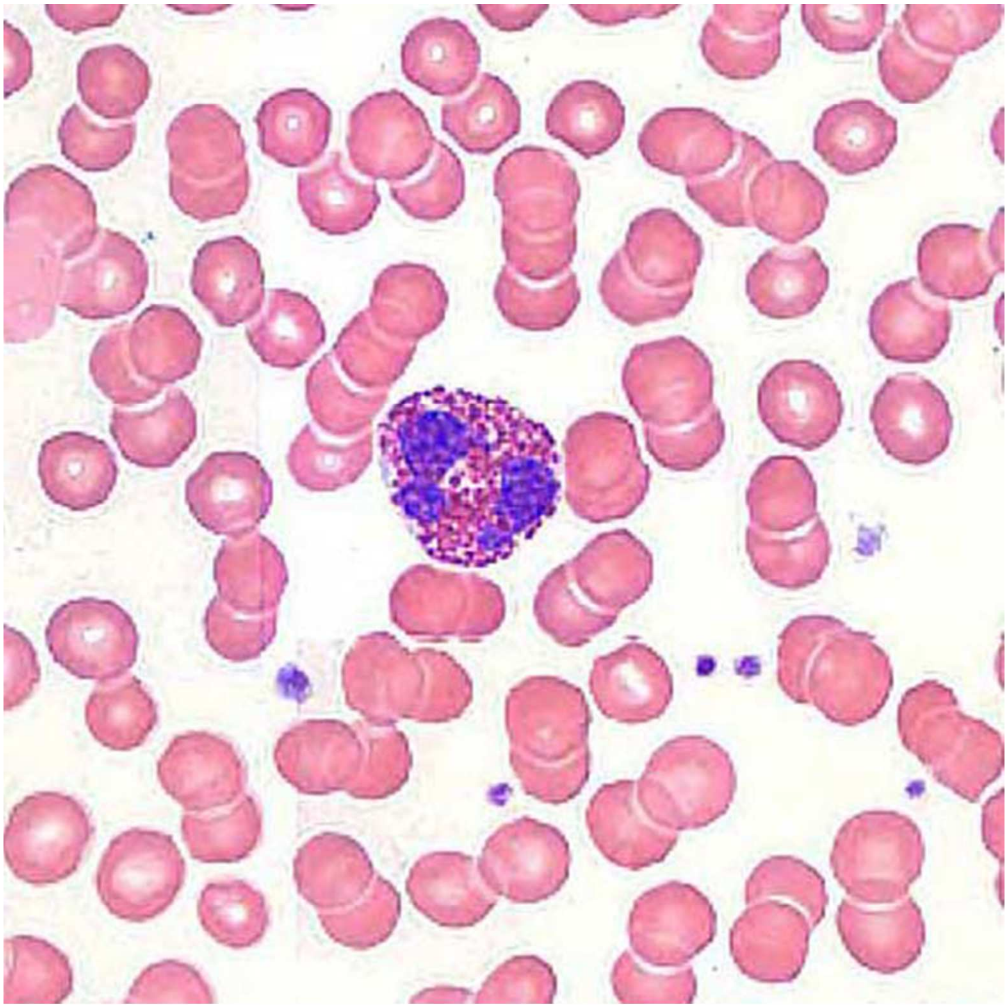
Cohen's Kappa Between Raters' Scores for Student-Generated Multiple Choice Questions (N = 710)			
	Rater 1	**Rater 2**	**Rater 3**
Rater 1	-	0.583[a]	0.583[a]
Rater 2		-	0.452[a]
Rater 3			-
Cohen's Kappa Between Raters' Scores for Teacher-Generated Multiple Choice Questions (N = 180)			
	Rater 1	**Rater 2**	**Rater 3**
Rater 1	-	0.764[a]	0.897[a]
Rater 2		-	0.763[a]
Rater 3			-

[a]Significant at the 0.01 level

Table 4. Difference in Performance of Answering Teacher-Generated Multiple Choice Questions between Undergraduate and Graduate Students

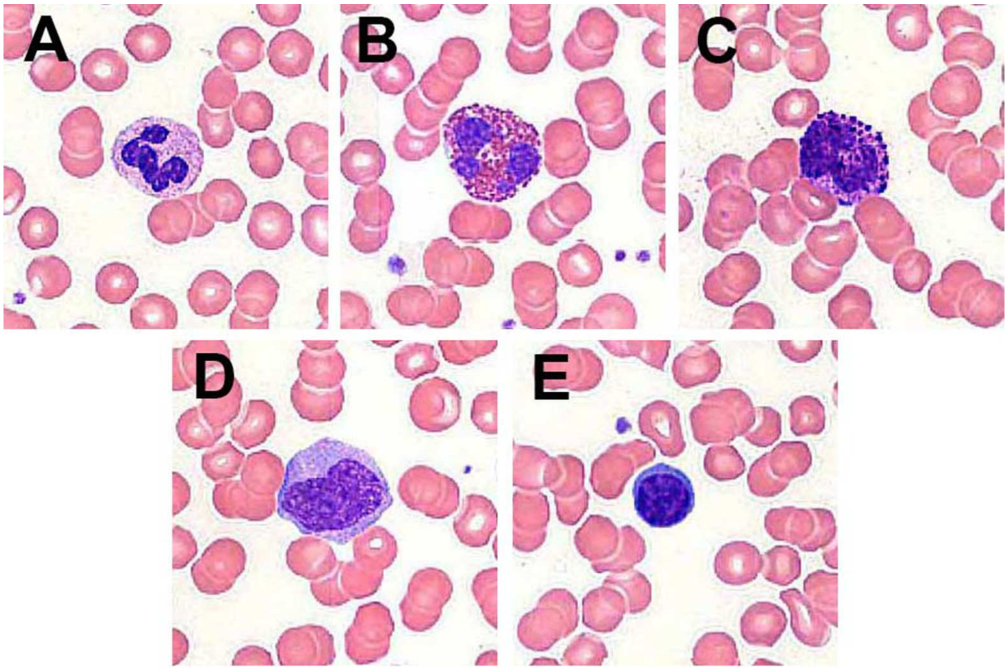
Type of student:	**N**	**First half of course**	**Second half of course**	**Entire course**
		Mean % (±SD)	**Mean % (±SD)**	**Mean % (±SD)**
Undergraduate students	51	85.23 (±10.12)	81.71 (±10.15)	83.46 (±9.36)
Graduate students	71	90.13 (±6.55)	87.76 (±9.11)	88.96 (±7.15)
t -value		3.03[a]	3.45[a]	3.68[a]
Cohen's *d*		0.57	0.63	0.66

[a]P < 0.005



Is part of Table 2

42x42mm (300 x 300 DPI)



Is part of Table 2

78x51mm (300 x 300 DPI)