Use of the NBME Comprehensive Basic Science Examination as a progress test in the preclerkship curriculum of a new medical school

Teresa R. Johnson,1 Mohammed K. Khalil,1 Richard D. Peppler,1 Diane D. Davey,2 and Jonathan D. Kibble1

1Department of Medical Education, University of Central Florida College of Medicine, Orlando, Florida; and 2Department of Clinical Sciences, University of Central Florida College of Medicine, Orlando, Florida

Submitted 11 April 2014; accepted in final form 11 August 2014

Assessments constructed from multiple-choice questions (MCQs) are the mainstay of knowledge testing in the preclinical years of medical school because they provide high levels of validity and reliability (2). It is a recognized problem in health sciences education that high-stakes MCQ tests frequently suffer from flaws that threaten their validity (7, 9, 11, 28). Moreover, several authors have written about the need to address this problem with faculty development interventions (15, 20, 34, 35). When a new curriculum is implemented, it is particularly important to establish the validity of such high-stakes tests to reassure students and to satisfy external regulatory authorities.

When a new curriculum starts, especially in a new school, there are no senior students with an established track record of success to inform progress. In this setting, it is important to provide evidence in real time to students and faculty members that the new curriculum is producing meaningful learning. Furthermore, finding metrics other than student evaluation on which to base curriculum evaluation decisions is a challenge in the early stages of a new curriculum. To meet these challenges, the idea of using the CBSE as a progress test was born.

Progress testing is an approach in which assessment instruments are used at intervals throughout the program to test student performance at the level expected at the end of a program of study (16, 37). The emphasis is on monitoring student improvement over time toward the achievement of competency. This approach to testing medical students emerged in the literature as early as 1977 (36) and has since been used extensively in undergraduate problem-based curricula (1, 4, 6, 12, 18, 22, 32, 33) and postgraduate medical education (8, 27) around the world. The CBSE is a 200-item MCQ exam made available by the NBME as a representative test blueprinted against the United States Medical Licensing Examination (USMLE) Step 1 content outline. The items are similar in style and difficulty to the USMLE Step 1, and NBME publishes performance standards that predict USMLE Step 1 performance. A previous study (10) in one medical school also demonstrated strong correlations between CBSE administrations toward the end of the M1 year and during the M2 year with subsequent USMLE Step 1 scores. As such, the CBSE represented to us a gold-standard external validation tool to show how student learning was progressing in a brand new integrated curriculum and to provide a means to validate the new summative MCQ assessments being created by faculty members.

The project served a number of secondary aims, including the provision of an external formative assessment tool for students. Formative assessment is widely regarded as an effec-
How We Teach: Generalizable Education Research

PROGRESS TESTING IN PRECLERKSHIP MEDICAL EDUCATION

Advances in Physiology Education • doi:10.1152/advan.00047.2014 • http://advan.physiology.org

A comprehensive mechanism for providing prescriptive feedback to students, guiding future learning, encouraging self-regulation and active learning strategies, promoting reflection and developing self-evaluation skills, maintaining motivation and self-esteem, and focusing the learner’s attention (3, 21, 23, 24, 26, 38). A final aim of this study was to establish benchmark CBSE scores from the start of a curriculum and thereby provide the medical education community with a measure to compare patterns of learning progression produced by different curriculum designs.

The present investigation aimed to assess the utility of the CBSE in addressing the aforementioned purposes, through an analysis of internally developed MCQ examinations and CBSE and USMLE Step 1 scores from students in the first two graduating classes of 2013 and 2014.

METHODS

Study setting and participants. The University of Central Florida (UCF) College of Medicine was established in 2006 and began educating the charter class of students in August of 2009. We offer an integrated curriculum across 4 yr of undergraduate medical education leading to the MD degree. During the first year, foundational basic sciences are taught in Human Body modules that leverage traditional synergies between disciplines (e.g., physiology is taught together with anatomy in a 17-wk structure and function module). The second academic year uses a systems-based approach (S modules), which focus on the study of disease process and culminates in students completing the USMLE Step 1. The last phase of the curriculum is the translation of knowledge and skills into practice and is represented by clerkship and elective rotations in the third and fourth academic years. The preclerkship curriculum phase of the first 2 yr of the MD program (Fig. 1) was the focus of this study. Participants for this study included students who matriculated in the graduating classes of 2013 (n = 41) and 2014 (n = 60). Our holistic admissions practices ensure that each class is diverse in regard to ethnicity, sex, and educational background. The present study was reviewed and exempted by the Institutional Review Board of UCF.

Overview of the assessment system. For each module shown in Fig. 1, MCQ exams formed the mainstay of knowledge testing. It is important to note that many other forms of testing are in use to characterize student progression toward competency in the multiple areas of knowledge, skills, attitudes and behaviors expected of medical students. These include a series of clinical skills assessments, culminating in an end-of-preclerkship objective structured clinical examination. Students maintain a portfolio documenting reflections from clinical preceptor visits as well as providing evidence of milestones from a required research project. Novel assessment tools are also being developed to address integrative learning, such as a group autopsy report and presentation as well as pathophysiology reports based on clinical postencounter notes. Attributes of professionalism will be monitored via a smartphone app we have developed. The utility of these newer assessments is the subject of other research. Student Evaluation and Promotion Committees meet 2–3 times/yr to consider each student’s progress and to recommend remediation as needed. The same committee stays with a student cohort through all 4 yr to become very familiar with each student’s progress in the cohort. Weekly formative MCQ assessments are available online, and narrative assessments are provided in every module for students to monitor their own learning.

Internal MCQ examinations. While our assessments are specifically constructed to address learning objectives at the level of individual lessons, all UCF faculty members are expected to create multiple-choice items in accordance with the NBME item-writing guide (19) to model the style of items appearing on the CBSE and USMLE Step 1. Each item is vetted by an interdisciplinary team of faculty members to ensure that it covers the appropriate material at the right level of difficulty, as recommended by other authors (15, 34). Each item is described with metadata that include what basic science disciplines and body systems are represented in the question as well as specific topic keywords. All items are stored in a central database managed by assessment office personnel. Module faculty members

![Fig. 1. The first 2 yr of the MD Program at the University of Central Florida College of Medicine with the timing of the five Comprehensive Basic Science Examination (CBSE) administrations. HB modules, Human Body modules; S modules, modules using a systems-based approach; OSCE, objective structured clinical examination.](http://advan.physiology.org/10.220.33.3 on June 18, 2017 http://advan.physiology.org/ Downloaded from http://advan.physiology.org/)
select items to appear on the respective summative module examinations. All exams were deployed through Questionmark software (Questionmark, Norwalk, CT) and were completed by students on their UCF-issued laptop computers under supervised conditions in large lecture halls. The summative assessments of interest for this investigation included a midterm and final examination in each of three Human Body modules and the final examinations given in each of six S modules (Fig. 1).

Progress testing procedure. Students in the graduating classes of 2013 and 2014 were offered five opportunities to complete the CBSE during their first 2 yr of undergraduate medical education. The progress testing initiative was adopted by the MD Program Curriculum Committee as a formal element of the curriculum, and UCF assumed the cost to deliver the CBSE ($44-$45/student for each administration). The purpose of the progress test was explained to students during orientation, and, in particular, it was communicated that performance data would not be used in any grading or promotion decisions. Students were advised not to engage in any special study for the exam to avoid conflicting with their normal coursework and to rely on what they were learning in regular classes. Exam administration dates were scheduled to occur at the following intervals (Fig. 1): 1) at the beginning of the first year as a baseline measure, before any instruction in August (time 1); 2) approximately at the middle of the first year in December (time 2); 3) at the end of the first year in May (time 3); 4) approximately at the middle of the second year in December (time 4); and 5) at the end of the second year in March, ~4–6 wk before students completed USMLE Step 1 (time 5). In each case, administration dates were posted on student class schedules in the learning management system, and e-mails were sent out to remind students about the progress test before each administration.

For each test administration, students assembled in a large lecture hall to complete the CBSE at the same time, such that makeup exams at alternate times were not available. Students were given 4 h to complete the CBSE, which consists of 200 single-best answer type multiple-choice items. All exams were computer based and were supervised by a UCF representative who is trained specifically in NBME proctoring procedures. The CBSE was scored by the NBME, and score reports were made available by the NBME in a secure environment within 7 days after each exam administration. CBSE scores are reported by the NBME as scaled scores, the majority of which range from 45 to 95. A scaled score of 70 is equivalent to a USMLE Step 1 score of 200; as such, a scaled score of 65 on the CBSE was considered to approximate a passing score on USMLE Step 1. Scaled scores permit direct comparison across multiple test forms and administration dates.

Data analysis. Categorical variables are reported as frequencies and percentages, and continuous variables are expressed as means ± SD, with 95% confidence intervals presented for the mean. Mixed repeated-measures ANOVA, using graduating class as a between-subjects factor and time of exam administration as a within-subjects factor, was conducted to examine the main effects of class and time as well as class × time interaction effects on CBSE scores. All followup pairwise comparisons were made with Bonferroni corrections. Effect sizes in association with mixed repeated-measures ANOVA are reported as η² values, in accordance with the following formula commonly used by statisticians: η² = SS_between / SS_total, where SS is the sum of squares (14). η² values are useful as effect sizes in that they may be interpreted as the percentage of variance accounted for by a variable or model (14). Between-class comparison of mean USMLE Step 1 scores was conducted using an independent samples t-test. Bivariate correlations between pairs of continuous variables were made using Pearson’s r. Exam reliability for our module exams was assessed as internal consistency using Cronbach’s coefficient α. All tests were two-sided, and P values of <0.05 were considered statistically significant. Statistical analyses were conducted using SPSS 21.0 (IBM, Chicago, IL).

RESULTS

Study participants. Although students were required to take all summative examinations and CBSE examinations, a small number of excused absences occurred for illness or other reasons; complete data sets related to UCF College of Medicine high-stakes exam and USMLE Step 1 scores were available for 39 of 41 students (95.1%) in the class of 2013 and for 58 of 60 students (96.7%) in the class of 2014. Complete data sets for 5 administrations of the CBSE were available for 39 of 41 students (95.1%) in the class of 2013 and for 55 of 60 students (91.7%) in the class of 2014.

Student performance: CBSE and USMLE Step 1 scores. Mean scores on the CBSE increased in a relatively linear fashion across the five administrations for both classes (Fig. 2 and Table 1), with scores at each time point being significantly different from the preceding time point. There was also a significant time by class interaction effect, albeit with a small effect size (η² = 0.01), that is, the class of 2014 demonstrated significantly higher scores compared with the class of 2013 at times 4 and 5. However, overall, there was no significant main effect of class, and the two classes did not differ significantly on USMLE Step 1 scores (P = 0.10; Table 1).

Exam performance: UCF exams, CBSE, and USMLE Step 1 score correlations. Table 1 shows data related to UCF exams by class to provide additional details about our internally developed assessments, including the number of items per exam, estimates of exam reliability, and mean scores. Mean scores were relatively consistent across exams and between classes, ranging from 82.8% to 90.3% for the class of 2013 and from 83.8% to 89.4% for the class of 2014. Exam reliability was moderate to strong across exams for each class, as the UCF College of Medicine considers α = 0.80 to be a minimum acceptable standard for the internal consistency of our high-stakes assessments.

UCF exam scores were moderately to strongly and, in most cases, significantly correlated to CBSE scores at times 2–5. For the class of 2013, significant correlations between UCF exam scores and CBSE scores were observed for 8 of 12 UCF exams at time 2, 9 of 12 UCF exams at time 3, and 11 of 12 UCF exams at times 4 and 5 (significant r = 0.32 to 0.74, P = 0.048 to <0.001). For the class of 2014, scores from 9 of 12 UCF exams were significantly associated with CBSE scores at time 2, and scores from all 12 UCF exams were significantly
correlated with CBSE scores at times 3–5 (significant \( r = 0.31 \) to 0.74, \( P = 0.02 \) to <0.001). The general trends are that correlations between UCF exam scores and CBSE scores primarily increased with time as the curriculum progressed.

UCF exam scores were moderately to strongly and significantly (i.e., in all but one instance) correlated to USMLE Step 1 scores (Table 1). For the class of 2013, correlations ranged from 0.30 (i.e., Human Body module 1 midterm, \( P = 0.07 \)) to 0.82 (S module 4 final, \( P < 0.001 \)). For the class of 2014, correlations ranged from 0.50 (i.e., Human Body module 3 midterm) to 0.81 (S module 3 final); all correlations were significant at \( P < 0.001 \). Correlation between CBSE and USMLE Step 1 scores at time 5 (\( r = 0.78 \)) may serve as the "gold standard" against which we can consider the strength of our UCF exams. Correlations between CBSE and USMLE Step 1 scores ranged from 0.03 (i.e., at time 1, \( P = 0.85 \)) to 0.78 (at time 5, \( P < 0.001 \)) for the class of 2013 and from 0.19 (i.e., at time 1, \( P = 0.16 \)) to 0.78 (at time 5, \( P < 0.001 \)) for the class of 2014. Correlations for both classes at times 2–5 were significant at \( P \leq 0.006 \). For the class of 2013, two of the UCF exams surpassed the standard of 0.78 (S module 2 final: \( r = 0.81 \) and S module 4 final: \( r = 0.82 \)). For the class of 2014, two of the UCF exams were very near or exceeded the standard of 0.78 (Human Body module 2 final: \( r = 0.77 \) and S module 3 final: \( r = 0.81 \)). Of note, the internal examinations showing higher correlation to the USMLE Step 1 than the final CBSE exam all had prominent components of physiology or pathophysiology content.

**DISCUSSION**

**Validity evidence for internal MCQ assessments.** Several lines of validity evidence support our ability to interpret student test scores on our newly developed internal MCQ assessments as meaningful measures of learning. First of all, we have developed a culture of careful test construction that includes faculty vetting of questions and tests to ensure high quality, appropriate difficulty, and correct representation of content. High statistical reliability is also a consistent feature of all the high-stakes tests (Table 1). Most central to the present study is the significant correlations observed between our internal MCQ assessments and as percentage of total number of items answered correctly for University of Central Florida College of Medicine exams.

### Table 1. Mean scores on the CBSE across the five administrations for both classes

<table>
<thead>
<tr>
<th>Exam</th>
<th>Class of 2013 (n = 39)</th>
<th>Class of 2014 (n = 58)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( k )</td>
<td>Mean ± SD* (95% CI)</td>
</tr>
<tr>
<td>CBSE Time 1</td>
<td>200</td>
<td>39.7 ± 3.0 (38.7–40.7)</td>
</tr>
<tr>
<td>HB module 1 midterm</td>
<td>116</td>
<td>86.7 ± 5.6 (84.8–88.5)</td>
</tr>
<tr>
<td>CBSE Time 3</td>
<td>125</td>
<td>82.8 ± 6.3 (80.7–84.8)</td>
</tr>
<tr>
<td>HB module 2 midterm</td>
<td>130</td>
<td>85.8 ± 6.3 (83.7–87.8)</td>
</tr>
<tr>
<td>HB module 2 final</td>
<td>200</td>
<td>47.7 ± 5.6 (45.9–49.6)</td>
</tr>
<tr>
<td>S module 1 final</td>
<td>138</td>
<td>87.1 ± 6.7 (84.9–89.3)</td>
</tr>
<tr>
<td>HB module 3 midterm</td>
<td>115</td>
<td>86.3 ± 7.0 (84.0–88.5)</td>
</tr>
<tr>
<td>CBSE Time 4</td>
<td>200</td>
<td>54.8 ± 5.8 (52.9–56.7)</td>
</tr>
<tr>
<td>HB module 4 final</td>
<td>130</td>
<td>85.8 ± 7.2 (83.5–88.1)</td>
</tr>
<tr>
<td>S module 5 final</td>
<td>84</td>
<td>87.5 ± 5.2 (85.8–89.2)</td>
</tr>
<tr>
<td>CBSE Time 5</td>
<td>200</td>
<td>61.3 ± 6.8 (59.1–63.5)</td>
</tr>
<tr>
<td>S module 6 final</td>
<td>130</td>
<td>83.0 ± 8.8 (81.2–84.9)</td>
</tr>
<tr>
<td>S module 6 final</td>
<td>126</td>
<td>85.8 ± 5.3 (84.1–87.5)</td>
</tr>
<tr>
<td>USMLE Step 1</td>
<td>118</td>
<td>88.9 ± 5.1 (87.2–90.5)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>68.2 ± 8.3 (65.5–70.9)</td>
</tr>
<tr>
<td></td>
<td>–322</td>
<td>221.2 ± 20.3 (214.6–227.8)</td>
</tr>
</tbody>
</table>

*Mean ± SD [95% confidence interval (CI)] values are reported as scaled score units for the National Board of Medical Examiners (NBME) Comprehensive Basic Science Examinations (CBSEs) and as percentage of total number of items answered correctly for University of Central Florida College of Medicine exams.

†Exam was assembled through the use of the NBME customized assessment services for the class of 2013, \( n \), number of students; \( k \), number of items; \( \alpha \), Cronbach’s coefficient \( \alpha \) for internal consistency; \( r \), Pearson correlation; USMLE Step 1, United States Medical Licensing Examination Step 1.

Downloaded from [http://advan.physiology.org](http://advan.physiology.org) by 10.220.33.3 on June 18, 2017
with a restricted sample size, in the fairly unique setting of a new medical school. While we would not claim the data are generalizable, we have demonstrated that the innovative use of the CBSE as a progress test has utility as a validation tool. Our data also represent an opportunity for other schools to perform comparative studies. This is a time of dynamic change nationally in medical curriculum design. For example, in a 2010 Academic Medicine supplemental issue (5), >80% of 128 United States medical schools were surveyed and described some level of integration of basic and clinical sciences as a core characteristic of their curriculum and future planning. Finding ways to compare learning progression and efficiency among the many variants of curriculum design is a daunting task. Our progress test initiative offers one avenue for comparative studies. In hindsight, another limitation is that we did not include further CBSE trials in the third or fourth years of medical school. This would be an improvement in our design that will allow learning retention to also be evaluated during the clinical years among different curricula.

**Progression of student learning.** At the most foundational level as a developing medical school, we needed to confirm that our students were learning. While there was a steady and significant upward trend in scores at each CBSE administration, we did not expect to see such a linear trend. Due to the comprehensive nature of the CBSE to assess a broad range of content typically covered through basic science medical education coursework (i.e., our first 2 yr of the curriculum), we expected the progression to be curvilinear. That is, it was anticipated that CBSE scores would increase markedly between times 3 and 4 (compared with gains made from times 1 and 2 and with times 2 and 3) and, even more so, between times 4 and 5. It is during this second year when students have the opportunity to acquire and assimilate the full scope of material appearing on the CBSE through systems-based coursework. This steady rate of change was a source of concern for us, especially for the class of 2013. At the final trial, after two academic years of basic science education, 13 of 39 students (33.3%) in the class of 2013 scored below the approximate passing score of 65 on the CBSE. In the end, 95% of this class passed USMLE Step 1 at the first attempt, with a mean score of 221, which is equivalent to a CBSE score of 79. This indicates to us that students were not retaining information during the curriculum as well as we would like and were relying on the short preexam review period to assure success on USMLE Step 1. This is probably a rather typical picture nationally; for example, in a study of 1,217 first-time takers of USMLE Step 1 from United States medical schools, Thadani et al. (30) observed that students studied for an average of 5.8 wk after the completion of their second year of education and for an average of 53 h/wk. However, we are now looking at ways to improve the horizontal and vertical integration of content and to better reinforce and rehearse material as we progress through the courses, with a view to aiding retention of learning.

The sharp upturn in student performance during the USMLE Step 1 self-study period is ample demonstration that student factors are as important as any changes faculty members can make to the curriculum. Another feature of the progress testing data that caught our eye was the steady increase in score variance with time, a phenomenon that has been noted previously in progress testing (18). At the first progress test trial, before the curriculum started, students in both cohorts scored very similarly, resulting in a small SD. In other words, the playing field of knowledge for end-of-second year material is level at the start. However, student performance quickly starts to spread out and continues to do so over time. This suggests to us that learner characteristics and behaviors are more critical than prior knowledge. Supporting the theme that a focus on generic learner expertise may be more critical than just improving teaching is our observation of how soon exam correlations to USMLE Step 1 emerge in the curriculum. For example, correlations between the structure and function exams and USMLE Step 1 are already very high, when students are only 3–4 mo into medical school. One explanation is that knowledge of physiology and anatomy is critical to USMLE Step 1 performance. However, large parts of USMLE Step 1 do not specifically test this material. More likely is the fact that this is a challenging course that creates meaningful variance in student performance, which reflects student capabilities in the same way that USMLE Step 1 will differentiate students over a year later. In a previous study (13) looking only at our structure and function course, we found that participation and performance on formative practice quizzes correlated to the summative examinations but that such correlations could be dissociated from the topic of study (e.g., performance on an internally developed quiz about reproduction would correlate just as well with performance on cardiovascular topics as it would on reproductive topics). Therefore, it is more likely that conceptually demanding areas identify “strong” versus “weak” students and that these distinctions transcend discipline or topic barriers.

In physiology education, Modell (17) has long expounded the view that teachers should focus energy on “helping the learner to learn.” In medical education, the importance of embracing the student’s ability to engage in self-regulated learning is also well recognized (25). Certainly, this viewpoint seems consistent with our data. For these reasons, we are now placing greater emphasis on student academic support. Our school has hired professional student academic support staff who now counsel students from day 1 about general wellness and study skills, including time management, self-testing, and test-taking strategies. The unit has also collaborated with the student body to create a peer-coaching program and other support systems aimed at enhancing learner expertise. We will monitor the impacts of these initiatives going forward.

**Utility of the CBSE as a progress test.** According to Van der Vleuten’s framework for evaluating the utility of assessments (31), six factors are important: reliability, validity, feasibility, cost effectiveness, acceptance, and educational impact. Reliability refers to the consistency of a measurement, whereas validity basically asks whether assessments test what they are supposed to test. With respect to the CBSE, the psychometric properties of items and validity are already established by the NBME and can effectively be taken as a gold standard. Considering feasibility and cost, the tests are fairly easy to administer but do require resources in terms of formally trained proctors and an Information Technology infrastructure for a large group of students to simultaneously take an online exam. The cost is approximately $45 per student per trial, which is not insignificant and must be budgeted. With respect to acceptability, we found, anecdotally, that the first two cohorts were enthusiastic about having the opportunity to take the CBSE, to gain experience of taking national board questions. Students

*How We Teach: Generalizable Education Research*

PROGRESS TESTING IN PRECLERKSHIP MEDICAL EDUCATION 319

**Advances in Physiology Education** • doi:10.1152/advan.00047.2014 • http://advan.physiology.org

Downloaded from http://advan.physiology.org by 10.220.33.3 on June 18, 2017
reported more appreciation for later trials in the progress testing sequence as USMLE Step 1 approached. Through our curriculum committee, the most recent classes have expressed that only CBSE trials 3 and 5 are needed (now that the internal testing program is validated). From the viewpoint of educational impact, we feel there are several positive outcomes of this study: 1) it has provided strong validity evidence for our internal assessment program; 2) reflecting on the data has led to program enhancements, such as increased efforts at integration as well as investment in student academic support services; 3) we have developed benchmarks for students to gauge their own progress as part of our formative assessment; and 4) we are providing the wider medical education community with benchmarks to compare learning progression in different curriculum models.

ACKNOWLEDGMENTS
Present address of M. K. Khalil: Department of Biomedical Sciences, University of South Carolina School of Medicine, Greenville, SC.

DISCLOSURES
No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

REFERENCES