Adv Physiol Educ 37: 370–376, 2013;
doi:10.1152/advan.00015.2013.

A comparison of student performance on discipline-specific versus integrated exams in a medical school course

Andrew R. Thompson,[1,2] Mark W. Braun,[2] and Valerie D. O'Loughlin[2]

[1]*Department of Biomedical Sciences, West Virginia School of Osteopathic Medicine, Lewisburg, West Virginia; and* [2]*Medical Sciences Program, Indiana University School of Medicine, Bloomington, Indiana*

Submitted 8 February 2013; accepted in final form 9 August 2013

Thompson AR, Braun MW, O'Loughlin VD. A comparison of student performance on discipline-specific versus integrated exams in a medical school course. *Adv Physiol Educ* 37: 370–376, 2013; doi:10.1152/advan.00015.2013.—Curricular reform is a widespread trend among medical schools. Assessing the impact that pedagogical changes have on students is a vital step in review process. This study examined how a shift from discipline-focused instruction and assessment to integrated instruction and assessment affected student performance in a second-year medical school pathology course. We investigated this by comparing pathology exam scores between students exposed to traditional discipline-specific instruction and exams (DSE) versus integrated instruction and exams (IE). Exam content was controlled, and individual questions were evaluated using a modified version of Bloom's taxonomy. Additionally, we compared United States Medical Licensing Examination (USMLE) step 1 scores between DSE and IE groups. Our findings indicate that DSE students performed better than IE students on complete pathology exams. However, when exam content was controlled, exam scores were equivalent between groups. We also discovered that the integrated exams were composed of a significantly greater proportion of questions classified on the higher levels of Bloom's taxonomy and that IE students performed better on these questions overall. USMLE step 1 exam scores were similar between groups. The finding of a significant difference in content complexity between discipline-specific and integrated exams adds to recent literature indicating that there are a number of potential biases related to curricular comparison studies that must be considered. Future investigation involving larger sample sizes and multiple disciplines should be performed to explore this matter further.

systems integrated scheduling; medical school curricula; curricular comparison

TRADITIONAL MEDICAL SCHOOL CURRICULA are based on guidelines developed over a century ago (13) and generally involve 2 yr of basic sciences education followed by 2 yr of instruction in a clinical setting. The first 2 yr are frequently taught in a didactic, discipline-based format. Recently, there has been a trend to revise this well-established tradition (8, 44) in favor of approaches that integrate subject material, such as problem-based learning (PBL) (7, 37, 42, 43), team-based learning (TBL) (16, 28, 32), and systems-integrated scheduling (SIS) (23, 34). SIS differs from PBL and TBL in that the focus is on teaching information across disciplines in a coordinated, block format based on anatomic systems (e.g., reproductive, renal, and cardiovascular) without necessarily involving team- or problem-based activities. However, a goal shared by all three instructional methods is to coordinate and integrate topics in multiple courses to better prepare students for the challenges involved with practicing medicine in a clinical setting.

SIS is considered favorable to discipline-based instruction because it requires students to consciously integrate information that might otherwise have been disconnected by large gaps in the timing of instruction (8). Current literature suggests that implementing SIS requires careful planning (34), but the impact on students is largely positive (27, 41). Students report that learning is more meaningful and relevant when learned in an integrated context (34). A change from discipline/course-specific examinations to integrated examinations (i.e., exams that combine questions from several courses) often accompanies SIS (27, 49). The effect that these changes have on students is not fully understood because many of the existing studies have focused on survey data and/or on describing trends and challenges involved with curricular revision (e.g., Refs. 11, 14, 17, 23, 27, and 33). To better appreciate the impact of SIS and integrated exams, it is important to consider testing performance, the primary mode of assessment in medical school and licensing.

Traditional medical school curricula with discipline-specific courses often (although not always) use discipline-specific exams (8, 44). For example, a pathology course will typically use exams that assess pathology content, and a pharmacology course typically will assess pharmacology course material. While there may be some questions on each exam that overlap with other course material, a discipline-specific exam primarily assesses the course material for each respective discipline. In contrast, an integrated exam assesses knowledge across multiple disciplines. For example, an exam on the cardiovascular or respiratory system may assess students' knowledge of these topics as they relate to pharmacology, pathology and medicine (23, 34). The United States Medical Licensing Examination (USMLE) step 1 exam is an example of a high-stakes integrated exam. As medical schools revise their curricula to integrate subject material, it is assumed that the assessment of this material will move from discipline-specific exams to integrated exams. However, there have been no systematic studies to assess how this change may impact student performance or whether directly comparing exam scores between these different exams formats is appropriate when medical curricular reform is evaluated.

In 2011, the second-year courses at the Bloomington center of the Indiana University School of Medicine (IUSM-Bloomington) transitioned from a purely discipline-specific format to one of more SIS. The main original courses (pathology, medicine, and pharmacology) remained, but the course content among all three overlapped based on body systems. In addition, course faculty members began implementing integrated exams to assess students over content from all three courses.

Address for reprint requests and other correspondence: A. R. Thompson. Dept. of Biomedical Sciences, West Virginia School of Osteopathic Medicine. 400 N. Lee St., Lewisburg, WV 24901 (e-mail: athompson1@osteo.wvsom.edu).

In this study, we followed this transition from discipline-specific exams to integrated exams in the second year of coursework at IUSM-Bloomington. We chose to focus our research on the course in pathology because pathology is often viewed as the bridge between basic sciences and clinical medicine (12, 31). The primary goal of this project was to explore the effect that integrated exams had on student exam performance in pathology and on the USMLE step 1 exam. Considering student performance on both in-house exams and the USMLE allowed us to evaluate both immediate and more long-term effects of the curricular and assessment changes.

Table 1. *Intergroup comparison of DSE and IE populations*

	2010 DSE Group	2011 IE Group	*P* Value
Sex (men/women)	16/11	16/14	N/A
Incoming age, yr	23.3 (2.1)	22.4 (1.9)	N/A
Undergraduate grade point average	3.79 (0.18)	3.72 (0.21)	0.24
Medical College Admission Test score	30.8 (2.6)	31.6 (2.6)	0.27
Class exam average (C601)	91 (3)	89 (4)	0.14

Values are means (SD). DSE, discipline-specific exam; IE, integrated exam. No significant differences were found between the two groups.

METHODS

Study sample. The second-year curriculum at IUSM-Bloomington consists primarily of three year-long courses: pathology, medicine, and pharmacology. Average total weekly contact hours range from 24 to 28 h. A typical week in pathology consists of 5 h of lecture and discussion, with 2 h of either laboratory or TBL assignments. In this study, we followed two cohorts of medical students in the second semester of pathology (C602). The control sample ($n = 27$ students), the discipline-specific exam group (DSE group), took C602 in spring 2010, before the change in exam format. The experimental sample ($n = 30$ students) consisted of the first cohort of students to be assessed with integrated exams. These students took C602 in spring 2011 and are referred to as the IE group.

Discipline-specific exams versus integrated exams. Before 2011, IUSM-Bloomington second-year medical courses met each week, and discipline-specific exams (consisting primarily of ~115 multiple-choice questions and 1–2 essay questions) were scheduled individually throughout the semester. A casual effort was made to coordinate core content of the three major courses. This resulted in the alignment of approximately one-half of the topics.

When SIS was initiated in spring 2011, the semester was divided into three 5-wk blocks. Each block consisted of 4 wk of classroom and laboratory instruction and 1 wk devoted solely to studying and assessment. Since SIS was implemented, a concerted effort improved coordination to 75% of the topics covered. Exams were modified from a discipline-specific format to an integrated format, whereby the three disciplines (pathology, medicine, and pharmacology) were assessed using a single integrated exam during the assessment week.

Integrated exams contained 200 total multiple-choice questions (approximately one-half from the pathology course) plus 3–5 essay-style questions. The multiple-choice questions were delivered in 2 segments of 100, and students were allotted 2 h for each 100-question segment. Many pathology and medicine questions were presented as case-based scenarios (clinical vignettes). Each course maintained its own exam key. There was no "aggregate" exam score; rather, only questions designated by each respective course director were scored for a particular course. Some questions were simultaneously scored by more than one course, but these generally represented <20% of the total number of questions.

Before 2011, the pathology course had four discipline-specific exams. After the implemention of integrated exams, topics that were previously covered over the third and fourth pathology exams were combined into the third integrated exam. Thus, there were only three integrated exams in 2011 versus four discipline-specific exams in 2010. For this study, the third and fourth discipline-specific exams were combined and compared with the third integrated exam. Only multiple-choice questions were examined for this study.

Comparison between DSE and IE groups. The nature of this study necessitated a comparison between two different cohorts of medical students (vs. a longitudinal study of a single group). According to Schmidt et al. (42), a basic requirement of curricular comparison (CC) studies is to verify that no substantial intrinsic differences in prior academic performance or general demographics exist between the groups being compared. To test for differential enrollment, we compared incoming grade point average, Medical College Admission Test score, and age/sex information (Table 1) and found no statistically significant intergroup differences. In addition, both cohorts took the first semester of pathology (C601) in the discipline-specific exam format, which permitted a more direct and recent comparison of aptitude. There were no statistically significant differences between class averages in C601 ($t = 1.51$, $P = 0.14$). Taken together, these results suggest that any observed intergroup test score differences reflect the changes made in exam format and curriculum.

Several additional concerns related to CC study design have been outlined by Schmidt et al. (42). The first is differential sampling, which refers to whether participants in a CC study are drawn from similar groups or have a similar motivation for volunteering to be part of a study. This issue does not apply to our study since the two groups that were compared were represented by the entire cohort of each respective group, not on a volunteer basis. Differential attrition is another form of bias that can impact cross-sectional CC studies when the curricular change has been implemented for an extended period of time. However, since this study focused on the first semester after the curricular change, there should be no difficulties related to student attrition. The last sampling bias described by Schmidt et al. (42) is differential exposure. Again, the design of this study removed this as a potential problem since both groups had identical training in the first year of medical school and the study took place at the inception of the curricular change.

Data preparation and analysis. Exam performance on pathology multiple-choice questions was explored in two ways. First, we investigated whether there were any overall differences in exam scores between DSE and IE groups for each exam. Second, we identified both identical and comparable pathology questions between the two sets of exams and then compared class averages on these subsets of questions. Questions were considered comparable if the target information was the same but the wording or scenario of the questions differed slightly (Table 2). For example, test questions that assessed risk factors for lymphoma were considered comparable because the subject matter was the same, even if the wording or scenarios differed slightly between questions. All questions that were identified as being comparable were verified by the author of the questions (M. W. Braun). Sample sizes for the pooled set of comparable and identical exam questions were as follows: $n = 31$ questions (*exam 1*), $n = 33$ questions (*exam 2*), and $n = 43$ questions (*exam 3*).

SPSS version 19 (22) was used perform all statistical analyses. Student's *t*-tests were used to compare exam averages between cohorts. Because raw data were in the format of a proportion, an arcsin transformation was performed before the statistical analyses were run. Statistical significance was set at $P < 0.05$. Although cohort sizes were slightly different (DSE: $n = 27$ students and IE: $n = 30$ students), the probability making a type I error due to the differences in sample size is low given the ratio of students and the selected α-level (15).

Table 2. *Sample of comparable pathology questions and lower- and higher-level pathology questions determined using a revised version of Bloom's taxonomy*

Comparable But Not Identical Questions	
A 10-yr-old black boy had gone camping with a group of Boy Scouts at a state park (6,000 ft above sea level). He became ill and was brought down to sea level, where you saw him in your hospital's emergency room. His hemoglobin is 9 gm/dl, hematocrit is 24%, and red blood cell count is 2.4 million/mm^3. The patient probably has: *A*. Chronic renal disease *B*. Glucose-6-phosphate deficiency *C*. Hereditary spherocytosis *D*. Sickle cell anemia*	A 17-yr-old black male has a history of severe lifelong anemia requiring many blood transfusions. He has nonhealing leg ulcers and recurrent episodes of abdominal and chest pain. His spleen is nonpalpable, and he has had two serious infections of encapsulated bacteria. These signs and symptoms are most likely associated with which one of the following laboratory abnormalities? *A*. Decreased erythrocyte glucose-6-phosphate dedydrogenase *B*. Increased erythrocyte osmotic fragility *C*. A positive direct Coombs test *D*. A positive sickle cell test* *E*. Severe hypochromic microcytic anemia
Lower-Level Pathology Question	**Higher-Level Pathology Question**
An epidemiological study is done of men who have developed breast cancer. The demographic data included studies of past medical history, family history, and laboratory data in an effort to identify risk factors. Which of the following is the factor most consistently associated with the development of breast cancer in men? *A*. Gynecomastia *B*. Asian ancestry *C*. Age > 70 ys* *D*. BRCA-1 gene mutation *E*. Alcoholism	A 40-yr-old woman complaining of abdominal fullness and discomfort is found to be mildly hypertensive. Occasional red blood cells are present in the urine sediment. Abdominal sonography demonstrates bilaterally loculated, large kidneys. An additional problem that this patient might be at risk for is: *A*. Emphysema *B*. Berry aneurysm* *C*. Mycotic aneurysm *D*. Osteodystrophy *E*. Urolithiasis

*Correct answer.

RESULTS

Initial study results. Exam averages from all pathology questions were compared between the DSE and IE groups (Fig. 1). DSE students had higher averages compared with IE students on all exams. In three instances (*exam 1*, *exam 3*, and overall), DSE students scored significantly higher than IE students (*exam 1*: $t = 4.15$, $P < 0.001$; *exam 3*: $t = 2.99$, $P = 0.003$; overall: $t = 4.85$, $P < 0.001$). However, after the first exam, IE students had improved exam performance. This is not unexpected as it is reasonable to anticipate that IE students would need to acclimate to the new testing format. The decreased disparity in performance on the second and third exams suggests that this adjustment period occurred rather quickly.

We next compared the pooled sample of comparable/identical pathology questions from each set of corresponding exams (Fig. 2). Results were mixed, but in most cases IE students did similar to or better than their DSE counterparts. IE students scored significantly higher on the second exam ($t = 1.93$, $P = 0.05$), but overall the subset of comparable/identical questions produced nearly identical class averages. These findings present a very different picture than the complete exam averages and suggest that additional variables could be influencing the results.

To further assess the effects of the curricular change, we compared USMLE step 1 exam scores between DSE and IE groups. In 2010, the DSE group had an average score of 234 (SD 21), whereas 2011 IE students averaged 235 (SD 19). This slight difference was not statistically significant ($t = 0.28$, $P = 0.77$). Comparison of scores from each cohort to overall U.S./Canada averages for that respective year did not add any

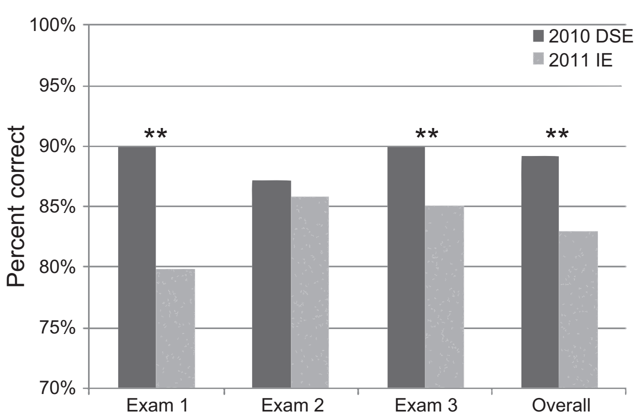

Fig. 1. Discipline-specific exam (DSE) versus integrated exam (IE) student performance on complete exams. While DSE groups scored higher on all of the exams, IE students showed improvement after *exam 1*. Statistical significance assessed using Student's *t*-tests. *$P < 0.05$; **$P < 0.01$.

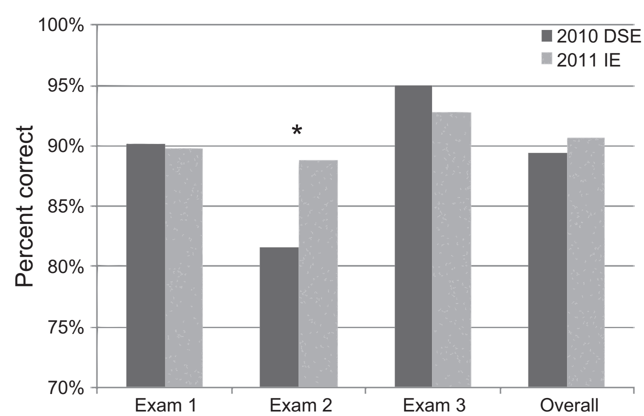

Fig. 2. DSE versus IE student performance on comparable/identical pathology questions. When only comparable/identical pathology questions were analyzed, DSE and IE performance was similar except on the second exam, where IE students scored significantly higher than DSE students. Statistical significance assessed using Student's *t*-tests. *$P < 0.05$; **$P < 0.01$.

additional insight since both DSE and IE groups scored significantly higher than the nationwide average.

Followup study: evaluation of differences between exam question complexity. One of the authors (M. W. Braun) commented that in addition to an increased number of exam questions, the questions used in integrated exams seemed more complex than in previous years. Most notably, more clinical vignettes were used in integrated exams. These types of questions often required students to answer a series of related questions that targeted information from different courses. To investigate whether there was a difference in the complexity of questions between discipline-specific exams and integrated exams, A. R. Thompson and V. D. O'Loughlin scored all the pathology questions from both sets of exams (n = 816 questions) using a modified version of Bloom's taxonomy (2, 6). The process of examining complexity using Bloom's taxonomy has been performed in other science disciplines. For example, Bissell and Lemons (4) evaluated introductory biology exam questions using Bloom's taxonomy and Zheng et al. (51) used Bloom's taxonomy to explore the degree to which critical thinking was assessed in various exams leading up to and in medical school. Additionally, Crowe et al. (9) developed a Blooming Biology Tool to assist biology faculty in aligning their assessments with course objectives. Unfortunately, the Blooming Biology Tool was not the most appropriate tool for us to evaluate the pathology questions in this study, which is why we used a separate, modified version of Bloom's taxonomy.

It should be noted that the various levels of Bloom's taxonomy are not directly analogous to degrees of difficulty. For example, a question that requires basic recall can be extremely difficult if the content is obscure or very detailed. Similarly, it would be incorrect to assume that item discrimination indices would directly correspond to the levels of Bloom's taxonomy. Instead, Bloom's taxonomy is used to measure the cognitive level associated with or targeted by a particular measure of competency (in this case, multiple-choice exam questions). While it may be the case that some questions classified as higher level according to Bloom's taxonomy are more "difficult" compared with lower-level questions, hierarchical ordering of question based on the intrinsic level of difficulty is not only problematic to measure but also is not the direct intent of Bloom's taxonomy.

Our initial attempt to classify exam questions into the four Bloom's categories that can typically be used to assess the pathology multiple-choice questions (21) proved difficult and inconsistent since there is considerable overlap between category boundaries. To circumvent this issue, we collapsed the first two levels (remembering and understanding) together and the third and fourth levels (applying and analyzing) together. In the resulting organization scheme, we categorized questions into what we term lower level and higher level (Table 2). The practice of combining categories is common when using Bloom's taxonomy (51) and provides more consistent results both within and between observers (46).

Two authors (A. R. Thompson and V. D. O'Loughlin) scored each of the 816 pathology questions independently and then compared findings. When a discrepancy in scoring arose, the exam question was discussed and a consensus was reached. Intraobserver reliability was tested by comparing the scores we assigned to identical questions between integrated and discipline-specific exams (n = 64 questions). A Cohen's κ-test (29) was used to assess error. A κ-value of 0.845 was obtained, which indicates strong reliability (3) and suggests that we were consistent in applying our ranking system across all exams.

Followup results. We first investigated whether there was a difference in the proportion of higher-level questions between exam sets (Fig. 3) using χ^2-tests. The results demonstrated there was a significantly greater proportion of higher-level questions on each of the integrated exams (*exam 1*: $\chi^2 = 12.37$, $P < 0.001$; *exam 2*: $\chi^2 = 4.22$, $P = 0.05$; *exam 3*: $\chi^2 = 15.75$, $P < 0.001$). In fact, more than half of the pathology questions were categorized as higher level on the first two integrated exams and overall. This finding might explain why Figs. 1 and 2 do not quite align: there were, in fact, differences in the levels of competency being tested between integrated exams and discipline-specific exams. Thus, it is not surprising that there were considerable differences in performance on the complete exams but similar performance when exam content was controlled.

Next, we explored whether there was a difference in how DSE and IE students performed on higher-level questions. To examine this, we compared class averages on the sample of comparable/identical higher-level questions between exams using Student's t-tests (Fig. 4). Sample sizes of comparable/identical higher-level questions varied both by group and exam since comparable questions might not necessarily be scored the same level of Bloom's taxonomy (i.e., the core content of the questions were comparable, but the structure or way the question was presented differed). The sample size for the number of comparable/identical questions for each exam and group are as follows: *exam 1* (DSE: n = 13 questions and IE: n = 17 questions); *exam 2* (DSE: n = 16 questions and IE: n = 16 questions); and *exam 3* (DSE: n = 13 questions and IE: n = 18 questions).

DSE students scored slightly better than IE students on higher-level questions on the first exam, but the difference was not statistically significant. On the second exam, IE students scored significantly better than DSE students on higher-level questions ($t = 2.21$, $P = 0.03$). IE students continued to do better on higher-level questions on the third exam, although results were not statistically significant. Over the course of the entire semester, IE students answered correctly on 5% more higher-level questions than DSE students.

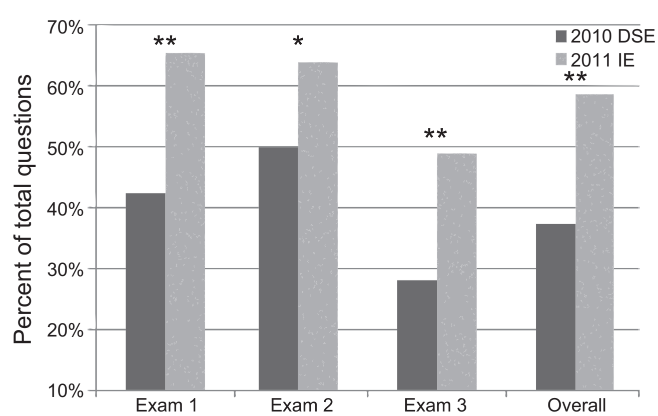

Fig. 3. Percentage of higher-level Bloom pathology questions on discipline-specific versus integrated exams. IEs were composed of a significantly larger proportion of higher-level questions. Statistical significance assessed using χ^2-tests. $*P < 0.05$; $**P < 0.01$.

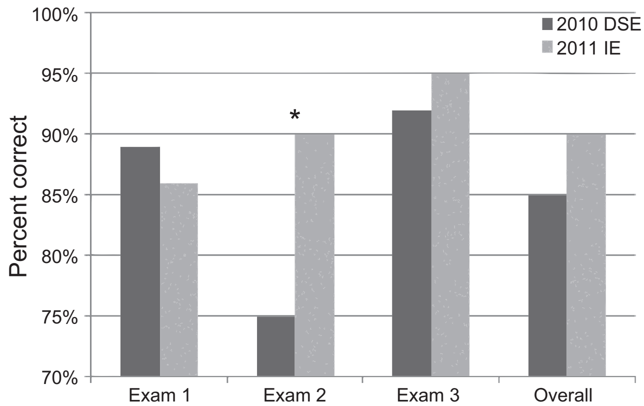

Fig. 4. DSE versus IE student performance on comparable/identical higher-level pathology questions. IE students outperformed DSE students on the second and third exams and overall. Statistical significance assessed using Student's *t*-tests. $^*P < 0.05$; $^{**}P < 0.01$.

DISCUSSION

If a goal of a medical school is for students to assume responsibility for their own learning and graduate with the ability to integrate and apply knowledge, then the traditional compartmentalized presentation and assessment of information may be counterproductive to meeting these objectives (36, 47).

In the 1950s, Western Reserve University (now Case Western Reserve) implemented an "organ-based" approach that required interdepartmental and interdisciplinary coordination (30). The results of this transition were not only favorable for student learning but also underscored the importance of evaluating the educational impact of curricular changes through educational research. The trend of revising curricula in medical school has continued into the present decade (8, 23, 24, 35, 44, 50), and numerous resources exist that outline the challenges and lessons learned during curricular reform (11, 14, 17, 18, 27, 33).

In the present study, we investigated the impact that a change from a traditional curriculum and discipline-based exams to SIS and integrated exams had on student performance at IUSM-Bloomington from 2010 to 2011. Our findings indicate that, on the surface, IE students had more difficulty with the pathology portion of the integrated exams compared to DSE students. Although overall exam performance improved after the first exam, IE students never quite caught up to the level of DSE students. Our initial comparison using complete exam scores proved to be problematic after we demonstrated that a substantial difference in exam complexity existed between discipline-specific and integrated exams. With this in mind, we took a more detailed approach by making a controlled comparison of comparable/identical exam questions using a modified version of Bloom's taxonomy. We have noted that other biology education researchers have used Bloom's taxonomy to develop exam analysis rubrics and tools for instructors to determine if course objectives aligned with their assessments (4, 9, 51). In our study, the more controlled comparison of comparable exams questions, when further categorized with Bloom's taxonomy, showed that IE students did as well or better than DSE students over the course of the semester.

A previous study (42) has warned against the superficial comparison of curricula. Our initial findings add credence this concern since exam question complexity appears to have played an important role in shaping our initial results. From a research standpoint, this discovery shows that the format, composition, and complexity of exam questions should be taken into consideration when student performance is assessed, especially after curricular changes.

A limited number of studies have looked directly at the implementation of integrated exams. Hudson and Tonkin (20) evaluated the impact of a transition from discipline-specific to integrated practical exams during the third year of medical school. The authors found overall positive results from this shift, although no measures of student performance change or long-term effects were investigated. Sodhi et al. (45) described the integration of gross anatomy and physical diagnosis courses during the first year of medical school. Compared with students where information and testing occurred in each course separately, students in the integrated environment scored significantly better on integrated practical exams. Yet, the authors acknowledge that limitations imposed by the study design require the findings to be interpreted tenuously. Although integrated exams were incorporated into the curricular revisions outlined by Klement et al. (27), there was no direct assessment of changes in student performance on course examinations. However, the authors did find that students in the integrated curriculum scored higher on National Board of Medical Examiners subject exams than students in the discipline-based format.

Higher test scores with integrated curriculum formats, such as PBL and TBL, have been demonstrated in several studies. Koles et al. (28) found that TBLs improved exam scores, especially for those in the lowest academic quartile. Clinical diagnoses, which often require an integrated understanding of several disciplines, were found to be more accurate among students in PBL and integrated curricula versus those in discipline-based instruction (41). Despite early studies that suggested that PBL strategies have a negative impact on standardized test scores (1, 7, 48), Blake et al. (5) found a trend for USMLE step 1 and step 2 exam performance to increase with the adoption of PBL. Although the authors do not discuss the testing format in their particular PBL curriculum, they suggest that the nature of PBL activities (often clinical vignettes) is similar to the USMLE exams, which might put PBL students at an advantage. A similar argument can be made for the use of integrated exams since the coordination of material often results in questions based around clinical scenarios (49). The presence of more vignette-style questions is a major reason why we found the integrated exams in our study to be composed of a greater number of higher-level questions.

The discovery that students exposed to a PBL curriculum tend to do better on clinical questions is not surprising. Studies have shown the ability to recall learned information is related to the number of repeated retrievals (25, 26). In addition, there is a positive correlation between the ease of recall and memory strength (38). While repeated retrieval of specific information might not necessarily increase in an integrated curriculum, students with more consistent exposure to the complexities involved with clinical diagnoses are more likely to be adept at approaching these cases with confidence and accuracy. This occurrence has been demonstrated with a PBL approach (19).

Complications associated with widespread comparisons of curricular changes, particularly with PBL strategies, have been described in detail (42). According to this research, the vast majority of earlier and even recent CC studies fall short on meeting the requirements of scientific validity due to a number of confounding factors that were not appropriately controlled. As a result, most review studies have demonstrated mixed outcomes related to curricular reform (1, 40). Schmidt et al. (42) reinvestigated a number of these earlier studies and found that after controlling for intergroup biases, positive outcomes of nontraditional curricula were apparent.

The present study adds to the findings of Schmidt et al. (42) in two ways. First, we showed that exam question complexity, as determined by Blooming exam questions (4, 9), should be an important consideration in CC studies that look at student performance, especially when comparisons are made on a small scale. As such, intergroup differences in even the most carefully controlled setting could be masked when exam format and complexity differences are not adequately considered. Second, we found support that an integrated style curriculum and assessment promotes aspects of knowledge acquisition that are favorable in a clinical environment. Most notably, implementation of an integrated-style form of instruction and assessment better prepared students in this study for exam questions that required information to be analyzed and/or applied.

Additional research, especially from other institutions, would be helpful in confirming our findings. The most notable limitations in our research are the relatively modest sample sizes and narrow focus of our study (a single discipline at a single institution). Increasing the number of students by continuing the study in future cohorts and in other courses would help resolve these shortcomings. In addition, there is inherent subjectivity involved with categorizing exam questions using Bloom's taxonomy. We attempted to minimize this issue by *1*) collapsing the categories into two groups (lower-level and higher-level questions) and *2*) having two separate individuals blindly classify questions and resolve discrepancies through discussion. However, there were still several instances when questions were on the border of higher- and lower-level categories. Developing a detailed rubric for classifying questions and including additional observers may be a helpful step in making the process more objective.

Conclusions. Given the widespread trend of curricular reform, education-based research is becoming increasingly more important. Transitioning to a new teaching and structural format in medical school is by no means an easy or straightforward process. This is underscored by the fact that most researchers suggest continual assessment and reappraisal of pedagogical strategies is a vital component of ongoing curricular development (10, 20, 33, 39, 44).

In this study, we assessed how the shift to a more integrated-style curriculum and integrated exams impacted student exam performance. After verifying that the two cohorts examined in our study were comparable, we found that an integrated learning environment facilitated student performance on questions that required information to be analyzed and/or applied (see Fig. 4). The discovery of a considerable difference in complexity between the two exams sets being compared adds to recently literature suggesting that there are several important factors to consider when CC studies are conducted (42). We also found that the transition to an integrated curriculum and testing format did not impact USMLE performance. It may take several semesters or even years of adjustments to achieve widespread effectiveness (and student acceptance) of innovative teaching strategies. Additional research using larger samples and multiple learning contexts should be performed to shed more light on the subject.

ACKNOWLEDGMENTS

The authors thank the Indiana University School of Medicine classes of 2012 and 2013 for participating in this study. The Indiana University Institutional Review Board approved this research (no. 1012004417). Dennis Deal (Indiana University School of Medicine) assisted us with United States Medical Licensing Examination averages.

DISCLOSURES

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

AUTHOR CONTRIBUTIONS

Author contributions: A.R.T., M.W.B., and V.D.O. conception and design of research; A.R.T. and V.D.O. analyzed data; A.R.T., M.W.B., and V.D.O. interpreted results of experiments; A.R.T. prepared figures; A.R.T. drafted manuscript; A.R.T., M.W.B., and V.D.O. edited and revised manuscript; A.R.T., M.W.B., and V.D.O. approved final version of manuscript.

REFERENCES

1. **Albanese M, Mitchell S.** Problem-based learning: a review of literature on its outcomes and implimentation issues. *Acad Med* 68: 61–82, 1993.
2. **Anderson L, Krathwohl D, Airasian P, Cruikshank K, Mayer R, Pintrich P, Raths J, Wittrock M.** *A Taxonomy for Learning, Teaching and Assessing: a Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman, 2001.
3. **Bernard H.** *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Lanham, MD: AltaMira, 2006.
4. **Bissell A, Lemons P.** A new method for assessing critical thinking in the classroom. *Bioscience* 56: 66–72, 2006.
5. **Blake RL, Hosokawa MC, Riley SL.** Student performances on step 1 and step 2 of the United States medical licensing examination following implementation of a problem-based learning curriculum. *Acad Med* 75: 66–70, 2000.
6. **Bloom B, Englehard M, Furst E, Hill W, Karathwohl D.** *Taxonomy of Educational Objectives: Cognitive Domain*. New York: McKay, 1956.
7. **Colliver J.** Effectiveness of problem-based learning curricula: research and theory. *Acad Med* 75: 259–266, 2000.
8. **Cooke M, Irby DM, Sullivan W, Ludmerer KM.** American medical education 100 years after the Flexner report. *N Engl J Med* 355: 1339–1344, 2006.
9. **Crowe A, Dirks C, Wenderoth MP.** Biology in bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sci Educ* 7: 368–381, 2008.
10. **Cunningham J, Freeman R, Hosokawa MC.** Integration of neuroscience and endocrinology in hybrid PBL curriculum. *Adv Physiol Educ* 25: 233–240, 2001.
11. **Dienstag J.** Evolution of the New Pathway curriculum at Harvard medical school: the new integrated curriculum. *Perspect Biol Med* 54: 36–54, 2011.
12. **Du Boulay C.** Learning pathology. Why? How? When? *J Clin Pathol* 50: 623–624, 1997.
13. **Flexner A.** *Medical Education in the United States and Canada: a Report to the Carnegie Foundation for the Advancement of Teaching*. Boston, MA: Merrymount, 1910.
14. **Giffin B, Drake R.** Gross anatomy of the head and neck and neuroscience in an integrated first-year medical school curriculum. *Anat Rec* 261: 89–93, 2000.
15. **Glass G, Peckham P, Sanders J.** Consequences of failure to meet assumptions underlying the fixed effects of variance and covariance. *Rev Educ Res* 42: 237–288, 1972.
16. **Haidet P, Levine RE, Parmelee DX, Crow S, Kennedy F, Kelly PA, Perkowski L, Michaelsen L, Richards BF.** Perspective: guidelines for reporting team-based learning activities in the medical and health sciences education literature. *Acad Med* 87: 292–299, 2012.
17. **Harden R.** The integration ladder: a tool for curriculum planning and evaluation. *Med Educ* 34: 551–557, 2000.

18. **Harden R, Crosby F, Davis M, Howie P, Struthers A.** Task-based leaning: the answer to integration and problem-based learning in the clinical years. *Med Educ* 34: 391–397, 2000.
19. **Hmelo C.** Cognitive consequences of problem-based learning for the early development of medical expertise. *Teach Learn Med* 10: 92–100, 1998.
20. **Hudson JN, Tonkin AL.** Evaluating the impact of moving from discipline-based to integrated assessment. *Med Educ* 38: 832–843, 2004.
21. **Huxham G, Naerra N.** Is Bloom's taxonomy reflected in the response pattern to MCQ items? *Med Educ* 14: 23–26, 1980.
22. IBM. *IBM SPSS Statistics for Windows (version 19.0).* Armonk, NY: released 2010.
23. **Irby DM, Wilkerson L.** Educational innovations in academic medicine and environmental trends. *J Gen Intern Med* 18: 370–376, 2003.
24. **Jones R, Higgs R, de Angelis C, Prideaux D.** Changing face of medical curriula. *Lancet* 357: 699–702, 2001.
25. **Karpicke JD, Roediger HL.** The critical importance of retrieval for learning. *Science* 319: 966–968, 2008.
26. **Karpicke JD, Roediger HL.** Repeated retrieval during learning is the key to long-term retention. *J Mem Lang* 57: 151–162, 2007.
27. **Klement BJ, Paulsen DF, Wineski LE.** Anatomy as the backbone of an integrated first year medical curriculum: design and implementation. *Anat Sci Educ* 4: 157–169, 2011.
28. **Koles P, Stolfi A, Borges N, Nelson S, Parmelee D.** The impact of team-based learning on medical students' academic performance. *Acad Med* 85: 1739–1745, 2010.
29. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174, 1977.
30. **Ludmere K.** *Time to Heal: American Medical Education From the Turn of the Century to the Era of Managed Care*. New York: Oxford Univ. Press, 1999.
31. **Marshall R, Cartwright N, Mattick K.** Teaching and learning pathology: a critical review of the English literature. *Med Educ* 38: 302–313, 2004.
32. **Michaelsen L, Sweet M.** Fundamental principles and practices of team-based learning. In: *Team-based Learning for Health Professions Education: a Guide to Using Small Groups for Improving Learning*, edited by Michaelsen L, Parmelee DX, McMahon K, Levine RE, Billings D. Sterling, VA: Stylus, 2008, p. 9–34.
33. **Muller J, Shore WB, Martin P, Levine M, Harvey H, Kelly P, McCarty S, Szarek J, Veitia M.** What did we learn about interdisciplinary collaboration in institutions? *Acad Med* 76: S55–S60, 2001.
34. **Muller JH, Jain S, Loeser H, Irby DM.** Lessons learned about integrating a medical school curriculum: perceptions of students, faculty and curriculum leaders. *Med Educ* 42: 778–785, 2008.
35. **Neville A, Norman GR.** PBL in the undergraduate MD program at McMaster University: three iterations in three decades. *Acad Med* 82: 370–374, 2007.
36. **Newble D, Entwistle N.** Learning styles and approaches: implications for medical education. *Med Educ* 20: 162–175, 1986.
37. **Norman GR, Schmidt HG.** Effectiveness of problem-based learning curricula: theory, practice and paper darts. *Med Educ* 34: 721–728, 2000.
38. **Roediger H, Karpicke JD.** Learning and Memory. In: *Encyclopedia of Social Measurement*, edited by Kempf-Leonard K. San Diego, CA: Academic, 2005, vol. 2, p. 479–486.
39. **Satterfield J, Mitteness L, Tervalon M, Adler N.** Integrating the social and behavioural sciences in an undergraduate medical curriculum: the UCSF essential core. *Acad Med* 79: 6–15, 2004.
40. **Schmidt H, Dauphinee W, Patel V.** Comparing the effects of problem-based and conventional curricula in an international sample. *J Med Educ* 62: 305–315, 1987.
41. **Schmidt H, Machiels-Bongaerts M, Hermans H, Ten Cate OT, Venekamp R, Boshuizen H.** The development of diagnostic competence: comparisonof a problem-based, an integrated, and a conventional medical curriculum. *Acad Med* 71: 658–664, 1996.
42. **Schmidt HG, Muijtjens AM, Van der Vleuten CP, Norman GR.** Differential student attrition and differential exposure mask effects of problem-based learning in curriculum comparison studies. *Acad Med* 87: 463–475, 2012.
43. **Schmidt HG, Rotgans JI, Yew EH.** The process of problem-based learning: what works and why. *Med Educ* 45: 792–806, 2011.
44. **Shatzer J.** Instructional methods. *Acad Med* 73: S38–S45, 1998.
45. **Sodhi G, Ranniger C, Slaby F, Blatt B.** A-PEX: a case-based integrated physical diagnosis/anatomy program for first year medical students. *Med Sci Educator* 21: 330–334, 2011.
46. **Thompson AR, O'Loughlin VD.** *Evaluating an Anatomy-Specific Tool for Blooming Exam Questions.* Presented at American Association of Anatomists Annual Meeting. Boston, MA: 2013.
47. **Tosteson D.** New pathways in general medical education. *N Engl J Med* 322: 234–238, 1990.
48. **Vernon D, Blake R.** Does problem-based learning work? A meta-analysis of evaluative research. *Acad Med* 68: 550–563, 1993.
49. **Wallach P, Crespo L, Holtzman K, Galbraith R, Swanson D.** Use of a committee review process to improve the quality of course examinations. *Adv Health Sci Educ* 11: 61–68, 2006.
50. **Wong BM, Levinson W, Shojania KG.** Quality improvement in medical education: current state and future directions. *Med Educ* 46: 107–119, 2012.
51. **Zheng AY, Lawhorn JK, Lumley T, Freeman S.** Application of Bloom's taxonomy debunks the "MCAT myth". *Science* 319: 414–415, 2008.