Please provide feedback

Issues of Question Equivalence in Online Exam Pools

By Cody Goolsby-Cole, Sarah M. Bass, Liz Stanwyck, Sarah Leupen, Tara S. Carpenter, and Linda C. Hodges

During the pandemic, the use of question pools for online testing was recommended to mitigate cheating, exposing multitudes of science, technology, engineering, and mathematics (STEM) students across the globe to this practice. Yet instructors may be unfamiliar with the ways that seemingly small changes between questions in a pool can expose differences in student understanding. In this study, we undertook an investigation of student performance on our questions in online exam pools across several STEM courses: upper-level physiology, general chemistry, and introductory physics. We found that the difficulty of creating analogous questions in a pool varied by question type, with quantitative problems being the easiest to vary without altering average student performance. However, when instructors created pools by rearranging aspects of a question, posing opposite counterparts of concepts, or formulating questions to assess the same learning objective, we sometimes discovered student learning differences between seemingly closely related ideas, illustrating the challenge of our own expert blind spot. We provide suggestions for how instructors can improve the equity of question pools, such as being cautious in how many variables one changes in a specific pool and "test driving" proposed questions in lower-stakes assessments.

The switch to remote instruction during the COVID-19 pandemic created unique challenges for students and instructors compared with traditional online teaching and learning, such as the lack of preparation for, and choice of, the online environment; issues of equity and accessibility; and concerns about academic integrity online. Strategies for enforcing academic integrity when testing online include using lockdown browsers and proctoring software. However, using these systems raises issues of student access to the technology and students' reactions against, and possible anxiety about, intrusion into their personal space (e.g., Asgari et al., 2021; Eaton & Turner, 2020).

Before the pandemic, results of research on students cheating on unproctored online exams were mixed (e.g., Alessio et al., 2017, 2018; Beck, 2014: Harris et al., 2020; Watson & Sottile, 2010). The trauma and societal unrest during the COVID-19 pandemic, however, created additional pressures that were conducive to academic dishonesty. News reports chronicled universities' stories of increased cheating online (e.g., Cruise, 2020). One study found that the number of queries and responses posted to Chegg, a homework help website, increased by almost 200% between April and August 2020 compared with the same time interval the year before (Lancaster & Cotarlan, 2021). Students requested exam-style questions and received answers well within the examination time frame, suggesting uses that violated academic integrity expectations. Certainly, our experiences and those of other researchers

(Adams, 2021) corroborated those findings. Instructors thus sought solutions for encouraging academic integrity when administering unproctored online exams, such as drawing from question pools or test banks; presenting questions singly without allowing backtracking; restricting the time for an exam; permitting open notes; and writing more creative questions that could not easily be googled (e.g., Budhai, 2020; Raje & Stitzel, 2020).

In this article, we discuss our use and analysis of question pools in exams. Question pools allow instructors to use their learning management system (LMS) to draw a question randomly for each student from a set of questions designed to address the same learning objective or concept. Students then each receive a distinct exam consisting of a question from each pool as well as any nonpooled questions.

Research on the effective design of question pools is concentrated in the field of standardized testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Additional work has been done in computer science courses (Butler et al., 2020; Denny et al., 2019; Sud et al., 2019) and in examining statistical factors that affect question duplication among students (Murdock & Brenneman, 2020). In this study, we specifically analyzed pool types across science courses for variation in student performance. When instructors wish to make comparable questions for exam pools, recognizing factors that affect question analogy is important for equity. Based on our results, we

TABLE 1

Course descriptions and assessment methods.

Course name	Course number	Total enrollment	Student population	Question pool delivery	Open notes
Upper-Level Physiology	BIOL 305, 307	90	Biology majors	SQB	No
General Chemistry I and II	CHEM 101, 102	200–720	STEM majors	GQ	Yes
Introductory Physics II	PHYS 122	260	STEM majors	GQB	Yes

Note. GQ = groups of questions visible, no backtracking between groupings; GQB = groups of questions delivered singly, backtracking within groups; SQB = single question, backtracking allowed.

propose common categories of question types pooled and their associated caveats, share our findings for comparing analogous questions using student outcome data, and suggest implications for instructors to consider when they design question pools.

Methods

Context

This work was conducted at a medium-size minority-serving mid-Atlantic research university and involved multiple science, technology, engineering, and mathematics (STEM) courses: upper-level physiology, both semesters of general chemistry, and second-semester introductory physics. The courses served major and nonmajor undergraduate STEM students and enrolled between 90 and more than 300 students per section (Table 1). Our results reflect practices beginning with the rapid shift to remote instruction that took place in March 2020 and continuing through the completion of the spring 2021 semester.

Course assessments and delivery

Prior to the pandemic, we all taught primarily face-to-face classes and administered exams in proctored in-person sessions. After moving to remote instruction, we administered online exams via the LMS, opting for different approaches to deter cheating based on our personal teaching philosophies, such as the following:

TABLE 2

Example of instructors' processes for formulating comparable exam questions in pools.

Difficulty level	Description	Estimated time (min.)	Requires calculation
Easy	Simple concept or calculation that does not require much if any application	1	Yes/No
Medium	Application required; similar to many versions seen; calculation not lengthy	2–4	Yes/No
Difficult	Deeper application/analysis required; longer calculation	5+	Yes/No

- using an honor code pledge at the beginning of the exam
- allowing open notes
- delivering questions in random order
- drawing some questions randomly from pools
- delivering one question on the entire exam at a time and allowing backtracking
- grouping questions in sections and using the adaptive release function of the LMS to enable students to move to the next section

Creating question pools

Strategies for developing question pools depend on how the pools are being used. In some cases of high-stakes multiple choice exams, instructors may use the same question stem and vary the answer options in pools, mixing various correct and in-correct options (Denny et al., 2019). We focus on pools in which we varied question stems, composing questions meant to be isomorphic—that is, "problems whose solutions and moves can be placed in one-to-one relation with the solutions and moves of the given problem" (Simon & Hayes, 1976, p. 165). When creating such question pools for our high-stakes exams, we considered factors such as learning objectives addressed, topics covered, question difficulty, calculation required, and ease of making multiple versions. For example, some of us drew on our experience to rank questions addressing a specific learning objective as easy, medium, or difficult using criteria such as those in Table 2 for the chemistry classes. Considering such factors helped prevent disproportionate distribution of question

pools for an assessment (e.g., no part was all calculations or consisted mostly of difficult questions). Instructors also considered equal distribution of questions from each chapter in each section grouping so that no part of the exam would be heavier on earlier or later content.

One author teaching applied statistics used question pools differently, with the primary goal of providing students with scaffolded practice throughout the course. She based her pools on topic or concept, not intentionally designing analogous questions but exposing students to a variety of such questions throughout the semester by administering mini-quizzes, chapter quizzes, and finally unit quizzes, pulling from the same pools. This approach to question pools was not included in this analysis.

Statistical comparison of question comparability

When examining the item analysis of question pools in high-stakes exams during the time the courses were taught, we sometimes found that specific question pools were not actually analogous based on student performance. We made adjustments at the time to promote equitable grading, but in discussing this issue, we realized the need for a systematic method of determining if question pools were fair based on student performance.

We performed statistical hypothesis tests to determine whether the difficulty level was equivalent for each question in a high-stakes exam pool. *Difficulty level*, also known as item difficulty, is defined as the proportion of students who got each question correct (as per Towns, 2014). If the pool consisted of two questions, we used a large-sample hypothesis test for equality of proportions; if the pool had more than two questions, we used a chi-square test of homogeneity. In cases where the sample sizes were not sufficient to perform these tests, we used Fisher's exact test. A

significance level of 5% was used throughout.

Results

Assessing question pools

When we analyzed our exams, we mainly recognized four different ways in which we varied questions within pools to be analogous. The challenge to creating comparable questions varied by question type, with quantitative problems being the easiest to vary without altering average student performance. Creating comparable conceptual questions was more difficult, and we found several unexpected pitfalls in permuting these questions, as we describe in the following sections.

Quantitative questions

We define *quantitative questions* as those in which the calculation and concept remain the same across the pool, but the exact numbers inserted change (for examples, see the online appendix). These types of questions may sometimes be described as *algorithmic* (e.g., Hartman & Lin, 2011). Quantitative question pools may require students to provide their own numerical response or choose from multiple options. In either case, instructors need to be cautious when using multistep problems in pools to avoid unduly penalizing students for mistakes early in the process and creating wide variability in the percentage of correct responses between questions (Hartman & Lin, 2011; Towns, 2014). Overall, our quantitative questions usually showed low variability in student performance across versions of questions. Students in a physiology course were able to calculate action potential velocity with equal ease when the distances over which the action potential traveled were changed ($p = 0.7787$). In chemistry, students were able to calculate the half-life of a reaction in minutes and seconds comparably when the rate constant and concentration

were varied ($p = 0.2248$). Likewise, in the introductory physics course, students had similar test performance on a question calculating the charge of a particle when the distance between two charged particles and their individual charges were changed ($p = 0.9702$). Anomalies in student performance occasionally arose on these kinds of questions when there were small numbers of students receiving any particular question.

Questions that are easy to rearrange or make substitutions

Like quantitative questions, some questions may be easily varied by rearranging entities or variables (for examples, see the online appendix). For example, in physiology, one way to rearrange questions about the endocrine system is to change which hormone within a negative feedback loop is abnormal or modified. One example, in which different hormones are ultimately responsible for changes in a lemur's testosterone levels, is included in the online appendix. In this three-question pool (one for each hormone in the negative feedback loop), there were no significant differences among the question versions, with percentage correct ranging from 62% to 76% ($p = 0.3881$). Thus, the understanding of endocrine negative feedback loops required by the different versions of the question appeared to be roughly equivalent.

In chemistry, substitutions and rearrangements can be more sensitive. For example, a common approach to creating question pools may be to substitute different chemical reactions between variations. However, when students were asked to calculate the overall cell potential of an oxidation-reduction reaction where the half-reactions are varied, the percentage correct ranged from 57% to 90% ($p < 0.0001$).

In electricity, electric field and equipotential plot questions can be made into many different versions simply

by rotating the plots and changing the positions. The instructor can then ask students to find quantities such as the change in electric potential energy, work done by the electric field, or work done between different points. Student performance on this question showed no significant differences between the question versions, with percentage correct ranging from 77% to 82% ($p = 0.9904$). An example of an easy-to-rearrange question showing wide variability in student performance, however, comes from magnetism, where a charged particle is moving in both a magnetic and an electric field. Both the direction of the particle's motion and the direction of the applied electric field are changed between each version, with students calculating the direction of the applied magnetic field and magnetic force. This question showed significant differences among versions, with percentage correct ranging from 24% to 70% ($p = 0.0485$). Although this seems like an ideal question for pooling, student performance indicates otherwise.

Questions using conceptual opposite counterparts

For certain concepts, a simple and appealing method of creating another version of the same question is to pose its opposite or substitute its opposite conceptual counterpart; for example, if the question stem refers to high blood pressure, change it to low blood pressure, or if it refers to an acid, change it to a base (for examples, see the online appendix). These questions exemplify a special subset of rearrangement and substitution questions. Because they apparently simply present the same questions in the reverse direction or ask about the opposite counterpart of a conceptual pair, they seem fair to instructors, but students may not find them equally difficult. For example, in one of the physiology courses, when testing students on the effect on the action potential of varying the sodium concentration inside or outside of the cell, 93% of students correctly identified how the action potential would change if the amount of sodium *outside* the cell membrane was doubled, but only 37% correctly identified how the action potential would change if the amount of sodium *inside* the cell membrane was doubled ($p = 0.0014$). Similarly, 91% of students identified the correct order through which oxygen would diffuse through a series of body locations, but only 66% could correctly identify the order in which carbon dioxide would (which is the exact reverse of oxygen; $p = 0.0084$).

When designing questions on acid and base concepts in general chemistry, we recognized variations in students' ability to calculate pH and pOH directly or indirectly from the respective ion concentrations and allocated such questions to separate pools. We noted issues later, however, when pooling questions about the opposite conceptual counterparts of oxidation and reduction. For example, when analyzing the notation for a galvanic cell, students could easily identify the reaction taking place at the anode or cathode using standard convention when the magnitude of charge was the same in all iterations ($p = 0.7965$). By contrast, when students were asked to identify from a list of half-reactions which ones would take place at the anode or cathode compared with the standard hydrogen electrode, the results were not comparable ($p = 0.0003$). The percentage correct in this four-question pool ranged from 55% to 76%, with a distinction between which reactions occur at the anode (55%–60%) versus the cathode (71%–76%).

Not all questions in this category, however, suffered from these apparent contradictions in students' conceptual understanding. For example, in physics, students were asked to determine how the electric field changes for a parallel plate capacitor when the distance between the plates is either increased or decreased. Students were equally able to identify how the electric field would change when the distance decreased (51%) or increased (59%; $p = 0.3935$). This is a more straightforward application for the students, assuming they know the correct relationship between the electric field and distance between the plates.

Questions that are different but address the same learning objective

Exam questions ideally target specific learning objectives, and questions testing the same learning objective at the same level of Bloom's taxonomy of the cognitive domain (Anderson & Krathwohl, 2001; Bloom & Krathwohl, 1956; Crowe et al., 2008) may be seen by the instructor as fair to pool (for examples, see the online appendix). For example, in one of the physiology classes, students were required to know the definitions of *basal metabolic rate* and *standard metabolic rate*. From the instructor's perspective, these are both simple definitions (lowest Bloom's level), equivalent, and poolable. However, the definition of one of them is actually much easier for students to identify: More than 95% of the students correctly chose the definition for basal metabolic rate, but only 45% chose the correct definition for standard metabolic rate ($p = 0.0001$), with the answer patterns suggesting that about half of students lumped the two phrases together under the actual definition of basal metabolic rate.

Similarly, in a question testing the learning objective "Identify or describe the physiological conditions under which neural versus endocrine control of a variable is used," students were asked what the best piece of evidence was for uterine labor contractions being regulated by nervous signaling. This question was easy for students, but simply changing "nervous" to "endocrine" with exactly the same answer choices—an arrangement that may feel parallel to the instructor—made the question much

more difficult ($p = 0.0047$).

In the introductory unit for chemistry, students were given the learning objective of differentiating between chemical and physical changes for various processes. Although most questions in the pool were similar in difficulty (93%–100% correct), one question asking students whether dissolving a compound is a physical or chemical change was disproportionately more challenging (71% correct). On the other hand, a four-question pool in a later unit asking students whether a given gas property increases or decreases relative to other given properties at constant pressure was not statistically different ($p = 0.7237$).

Likewise, for the thermodynamics unit in physics, students were tested on the learning objective "Determine relevant thermodynamic parameters for any process depicted on a PV diagram" by being asked to show how heat energy is transferred in a Carnot cycle. Similar percentages of students were able to identify how heat energy is transferred for an adiabatic process (88%) and for an isothermal process (81%; $p = 0.1344$), despite these being two fundamentally different processes.

Discussion

During the pandemic, the use of question pools for online testing was recommended to mitigate cheating, exposing multitudes of STEM students around the world to this practice. Nevertheless, instructors may be unfamiliar with the ways that seemingly small changes between questions in a pool can expose differences in student understanding. Our investigation of our own online exams allowed us to uncover ideas that can help instructors approach this practice with confidence or caution depending on their goals for students' learning. As one might expect, manipulating quantitative questions usually resulted in comparable student performance across pools. However, in some cases of question pool types, faculty may suf-

fer from expert blind spot (Nathan et al., 2001). For example, when rearranging aspects of a question, posing opposite counterparts of concepts, or formulating questions assessing the same learning objective, our analysis sometimes revealed student learning differences between seemingly closely related ideas. Posing possible explanations for variation in student understanding between such questions is beyond the scope of this article and warrants future research. Our results, however, can alert instructors to subtle differences in cognitive level between question versions as well as student misconceptions, learning bottlenecks (Middendorf & Pace, 2004), and disciplinary threshold concepts (Meyer & Land, 2005).

When considering the equity of question pools, if many pooled questions are used throughout the semester and most are comparable, it may be tempting to assume that an individual student would receive as many difficult versions as they would easy ones over the term. However, an analysis of our pools suggests only about 70% of questions are statistically similar in difficulty, raising concerns about that premise. Our study provides insights that can help increase the likelihood that the distinct tests created for each student from pools are analogous. Our analysis does not provide any answers as to whether the practice actually decreases cheating, but other research suggests that it may (Chen et al., 2018; Silva et al., 2020).

Conclusions

Although promoting assessment integrity was the primary reason for using pool questions, we found other benefits as well. Instructors may wish to assess student learning frequently for formative as well as summative purposes, drawing from pooled questions by topic and exposing students to a wide variety of questions over time, as per the example of one author. Doing so may be one way for

instructors to address the problem of noncomparable questions. Pool questions can also inform instructors on which version of a question to ask for future assessments. For example, formative online assessments using pool questions could be used to ascertain the difficulty of similar versions of the same question. The results could then be used to determine the appropriate level of challenge for summative, single-version, paper-and-pencil assessments.
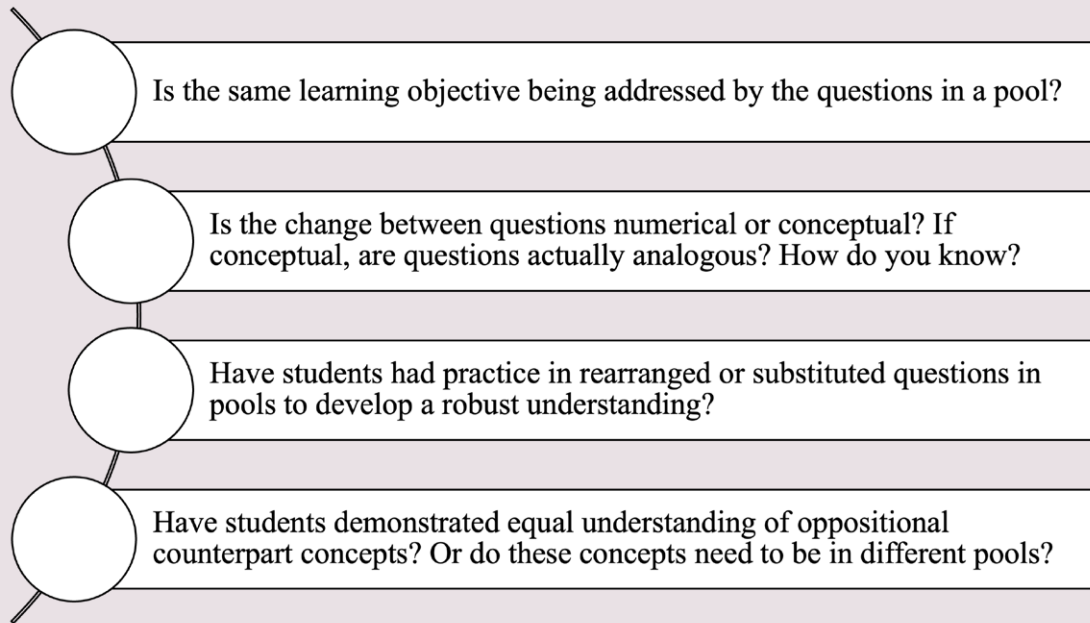
Based on our study, we offer several questions (Figure 1) and suggestions for practice when designing online exam pools for high-stakes assessment, including the following:

- Do not assume that questions are equivalent in difficulty unless they differ only quantitatively.
- Change only one variable in a multivariable question (e.g., the reaction or the number) to make questions comparable.
- Pool questions that have been shown to be of similar levels of difficulty.
- "Test drive" pools in lower-stakes assessments and perform an item analysis to determine if questions are equitable enough for future use.
- Evaluate performance on pool questions after an exam, and consider adjusting grading or discarding questions that are not comparable.
- Use the results of a pool question to better inform question design for an in-class single version of an exam.

Although many instructors used online exams and pools of questions for the first time during the pandemic, not all instructors returned to paper testing when they resumed face-to-face classes. Many instructors noticed several advantages of online exams using the LMS, such as thorough and automatic item analysis, ease of grading anonymously, ease of having instructors and teaching assistants

FIGURE 1

Questions to guide question pool design.

Is the same learning objective being addressed by the questions in a pool?

Is the change between questions numerical or conceptual? If conceptual, are questions actually analogous? How do you know?

Have students had practice in rearranged or substituted questions in pools to develop a robust understanding?

Have students demonstrated equal understanding of oppositional counterpart concepts? Or do these concepts need to be in different pools?

grade the same exams simultaneously, and ability to accommodate some students' preferences for typing over hand-writing answers to free response or essay questions. If online testing—even, in some cases, while sitting in a physical classroom—continues, instructors need to ensure that tests and quizzes are as equitable as possible. Using question pools is an important tool in the toolbox of creating assessments that are fair and promote exam integrity. Our work provides actionable information about how to create pools that help online tests meet these criteria.`

References

Adams, S. (2021, March 31). This $12 billion company is getting rich off students cheating their way through Covid. *Forbes*. https://www.forbes.com/sites/susanadams/2021/01/28/this-12-billion-company-is-getting-rich-off-students-cheating-their-way-through-covid/?sh=73c7268363ff

Alessio, H. M., Malay, N., Maurer, K., Bailer, A. J., & Rubin, B. (2017). Examining the effect of proctoring on online test scores. *Online Learning*, *21*(1), 146–161. http://dx.doi.org/10.24059/olj.v21i1.885

Alessio, H. M., Malay, N., Maurer, K., Bailer, A. J., & Rubin, B. (2018). Interaction of proctoring and student major on online test performance. *International Review of Research in Open and Distributed Learning*, *19*(5). https://doi.org/10.19173/irrodl.v19i5.3698

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives: Complete edition*. Longman.

Asgari, S., Trajkovic, J., Rahmani, M., Zhang, W., Lo, R. C., & Sciortino, A. (2021). An observational study of engineering online education during the COVID-19 pandemic. *PLOS One*, *16*(4), e0250041. https://doi.org/10.1371/journal.pone.0250041

Beck, V. (2014). Testing a model to predict online cheating—Much ado about nothing. *Active Learning in Higher Education*, *15*(1), 65–75. https://doi.org/10.1177/1469787413514646

Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook 1: Cognitive domain*. Longman.

Budhai, S. S. (2020, May 11). Fourteen simple strategies to reduce cheating on online examinations. *Faculty Focus*. https://www.facultyfocus.com/articles/educational-assessment/fourteen-simple-strategies-to-reduce-cheating-on-online-examinations

Butler, L., Challen, G., & Xie, T. (2020, November 9–12). *Data-driven investigation into variants of code writing questions* [Paper presentation]. Institute of Electrical and Electronics Engineers 32nd Conference on Software Engineering Education and Training, Munich, Germany. http://doi.org/10.1109/CSEET49119.2020.9206195

Chen, B., West, M., & Zilles, C. (2018, June 26–28). *How much randomization is needed to deter collaborative cheating on asynchronous exams?* [Paper presentation]. Association for Computing Machinery 5th Annual Conference on Learning at Scale, London, United Kingdom. https://doi.org/10.1145/3231644.3231664

Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education*, *7*(4), 368–381. https://doi.org/10.1187/cbe.08-05-0024

Cruise, R. (2020, May 2). College cheating scandal amid coronavirus: Georgia Tech, Boston University probing students after online exams. *International Business Times*. https://www.ibtimes.com/college-cheating-scandal-amid-coronavirus-georgia-tech-boston-university-probing-2968945

Denny, P., Manoharan, S., Speidel, U., Russello, G., & Chang, A. (2019, February 27–March 2). *On the fairness of multiple-variant multiple-choice examinations* [Paper presentation]. Association for Computer Machinery 50th Technical Symposium on Computer Science Education, Minneapolis, MN, United States. https://doi.org/10.1145/3287324.3287357

Eaton, S. E., & Turner, K. L. (2020). Exploring academic integrity and mental health during COVID-19: Rapid review. *Journal of Contemporary Education Theory & Research*, *4*(2), 35–41. http://doi.org/10.5281/zenodo.4256825

Harris, L., Harrison, D., McNally, D., & Ford, C. (2020). Academic integrity in an online culture: Do McCabe's findings hold true for online, adult learners? *Journal of Academic Ethics*, *18*(4), 419–434. https://doi.org/10.1007/s10805-019-09335-3

Hartman, J. R., & Lin, S. (2011). Analysis of student performance on multiple-choice questions in general chemistry. *Journal of Chemical Education*, *88*(9), 1223–1230. https://doi.org/10.1021/ed100133v

Lancaster, T., & Cotarlan, C. (2021). Contract cheating by STEM students through a file sharing website: A COVID-19 pandemic perspective. *International Journal for Educational Integrity*, *17*, Article 3. https://doi.org/10.1007/s40979-021-00070-0

Meyer, J. H., & Land, R. (2005). Threshold concepts and troublesome knowledge (2): Epistemological considerations and a conceptual framework for teaching and learning. *Higher Education*, *49*(3), 373–388. https://doi.org/10.1007/s10734-004-6779-5

Middendorf, J., & Pace, D. (2004). Decoding the disciplines: A model for helping students learn disciplinary ways of thinking. *New Directions for Teaching and Learning*, *98*, 1–12.

Murdock, M., & Brenneman, M. (2020). Designing tests from question pools with efficiency, reliability, and integrity. *Journal of Contemporary Chiropractic*, *3*(1), 92–100. https://journal.parker.edu/index.php/jcc/article/view/128

Nathan, M. J., Koedinger, K. R., & Alibali, M. W. (2001, August 27–31). *Expert blind spot: When content knowledge eclipses pedagogical content knowledge* [Paper presentation]. International Association for Cognitive Science Third International Conference on Cognitive Science, Beijing, China.

Raje, S., & Stitzel, S. (2020). Strategies for effective assessments while ensuring academic integrity in general chemistry courses during COVID-19. *Journal of Chemical Education*, *97*(9), 3436–3440. https://doi.org/10.1021/acs.jchemed.0c00797

Silva, M., West, M., & Zilles, C. (2020, March 11–14). *Measuring the score advantage on asynchronous exams in an undergraduate CS course* [Paper presentation]. Association for Computing Machinery 51st Technical Symposium on Computer Science Education, Portland, OR, United States. https://doi.org/10.1145/3328778.3366859

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology*, *8*(2), 165–190. https://doi.org/10.1016/0010-0285(76)90022-0

Sud, P., West, M., & Zilles, C. (2019, June 15–19). *Reducing difficulty variance in randomized assessments* [Paper presentation]. American Society for Engineering Education Annual Conference and Exposition, Tampa, FL, United States. https://www.asee.org/public/conferences/140/papers/25513/view

Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. *Journal of Chemical Education*, *91*(9), 1426–1431. https://doi.org/10.1021/ed500076x

Watson, G. R., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses? *Online Journal of Distance Learning Administration*, *13*, 1.

Cody Goolsby-Cole is *a lecturer in the Department of Physics,* ***Sarah M. Bass*** *is a lecturer in the Department of Chemistry and Biochemistry,* ***Liz Stanwyck*** *is a principal lecturer in the Department of Mathematics and Statistics,* ***Sarah Leupen*** *is a principal lecturer in the Department of Biological Sciences,* ***Tara S. Carpenter*** *is a principal lecturer in the Department of Chemistry and Biochemistry, and* ***Linda C. Hodges*** *(*lhodges@umbc.edu*) is the director of the Faculty Development Center, all at the University of Maryland, Baltimore County, in Baltimore, Maryland.*