Testing in the Age of Active Learning: Test Question Templates Help to Align Activities and Assessments

Gregory J. Crowther,[1] Benjamin L. Wiggins,[2] and Lekelia D. Jenkins[3]
[1]Department of Life Sciences, Everett Community College, Everett WA; gcrowther@everettcc.edu
[2]Department of Biology, University of Washington, Seattle WA; benlw@uw.edu
[3]School for the Future of Innovation in Society, Arizona State University, Tempe AZ; kiki.jenkins@asu.edu

Abstract
Many undergraduate biology instructors incorporate active learning exercises into their lessons while continuing to assess students with traditional exams. To better align practice and exams, we present an approach to question-asking that emphasizes templates instead of specific questions. Students and instructors can use these Test Question Templates (TQTs) to generate many variations of questions for pre-exam practice and for the exams themselves. TQTs are specific enough to show students which material they should master, yet general enough to keep the exact exam questions a surprise and easy to change from term to term. TQTs generate biology problems analogous to other STEM disciplines' standard problems whose general format is known to students in advance. TQTs thus help instructors ask more exam questions at the higher-order cognitive levels of Bloom's taxonomy, while empowering students to prepare actively and creatively for such questions. https://doi.org/10.21692/haps.2020.006

Key words: study guides, problem sets, rubrics

Introduction
Six years ago, a comprehensive meta-analysis argued that active learning was unequivocally superior to traditional lecturing in terms of student learning gains in STEM courses (Freeman et al. 2014). There is growing acceptance of this conclusion among biology and STEM faculty, as reflected in faculty surveys (Patrick et al. 2016) and the expectation of many search committees that teaching demonstrations should include active-learning techniques (Smith et al. 2013). The actual implementation of active-learning techniques may lag behind perceived best practices (Stains et al. 2018), due in part to many instructors' belief that they must cover more content than active learning can accommodate (Miller and Metz 2014, Silverthorn 2020).

The momentum of the active learning movement does not address the fact that the term "active learning" encompasses many distinct teaching and learning strategies (Bonwell and Eason 1991). In fact, in the above-mentioned meta-analysis (Freeman et al. 2014), active learning was operationally defined to include almost any student activity other than reading, listening, or verbatim copying of notes. This diversity of active learning options is, overall, a good thing, since different instructional goals may be served best by different approaches (Tanner 2013). However, this diversity also raises the question of how such learning can best be demonstrated in summative assessments, especially traditional comprehensive fact-based tests, which are the primary determinant of students' final grades in most STEM courses (Goubeaud 2010, Momsen et al. 2010).

As STEM educators, we are interested in the extent to which these traditional assessments align with active learning as currently practiced in undergraduate courses (Pellegrino 2006, Reeves 2006). While we are unaware of comprehensive empirical data on this issue, we suspect that the active learning movement has not reformed testing to the same extent that it has reformed classroom lectures. Traditional tests may not be an ideal inventory of the fruits of active learning; for instance, if active-learning activities help students improve their higher-order cognition (HOC), such improvements may not be captured by typical undergraduate biology tests, which consist mostly of lower-order cognition (LOC) questions (Momsen et al. 2010). Improving summative assessments to better align with class assignments is a current major effort of national directives like the Next Generation Science Standards (NGSS Lead States 2013).

In considering the alignment between learning activities and subsequent tests, it is helpful to apply the principle of backwards design (Wiggins and McTighe 2005), which advises teachers to first define how they want their students to demonstrate mastery in summative assessments, and then to design learning activities that lead naturally to those assessments. Like most teachers, we want our students to be able to solve problems that go beyond the recognition and recitation of specific facts (Songer and Kali 2014). Backwards design would suggest that we identify appropriate problems on which students can be assessed, and then give the students numerous opportunities (e.g., classroom activities and homework assignments) to practice such problems prior to formal summative assessments. The practice should resemble the assessment, though not so closely that the practice gives away the exact nature of the assessment, which would then allow students to pre-prepare memorized answers without

continued on next page

necessarily understanding them. Best practices indicate that it is useful to create opportunities for students to do the work of cognition both in performance and in their own independent practice (Pellegrino 2014). Specifically, the goal is to build novel, efficient assessments that reveal students' abilities to operate at HOC levels of engagement, as described in Bloom's taxonomy (Crowe et al. 2008, Anderson et al. 2001).

Introducing Test Question Templates

Here we describe an approach, which we call Test Question Templates (TQTs), for improving the alignment of learning activities and summative assessments. The word "template" is used to indicate that a TQT is not itself a question, but rather a question generator. A TQT can be used to generate a large number of distinct questions that conform to the template format (discussed below). The questions generated are suitable for tests as well as pre-test practice.

At its core, a TQT defines a relationship between an Input (the information given to the student) and an Output (what the student will do with the information given). For maximum clarity and transparency, a TQT should also include an Example (a specific case of an Input and the corresponding requested Output), a Key (a correct answer for the Example), and Other Answers (some imperfect responses along with notes on scoring them). Table 1 illustrates the structure of a TQT with an example in short-answer format; however, TQTs can just as easily be used to create multiple-choice questions, as shown in the footnote to Table 1.

TQT element	Example of element	Rationale
Input: Information to be given to the student.	"Given a table of the genetic code and a point mutation (reported in terms of the DNA coding strand OR the DNA template strand OR the mRNA)..."	This example indicates that students should know how to use a table of the genetic code, and should also be able to interconvert between the two DNA strands and mRNA.
Output: What the student will do with the information given.	"... determine the likelihood that the mutation will affect the function of the corresponding protein."	This example indicates that students should be able to convert codons to amino acids before and after the mutation, and thus see whether the amino acid changes (to another amino acid, or to a stop codon).
Example: A specific case of a possible question in the Input/Output format.	"Example: In an exon of the coding region of the coding strand for a particular gene, most people have the codon 5′-AGG-3′, but Jesse has codon 5′-CGG-3′. Is the corresponding protein likely to function differently in Jesse than in most other people? Show your work and explain your reasoning."*	An Example is vital for helping students see how a general Input-Output pair can be translated into a specific test question. The Example should ask for the same Output format (e.g., short answer or multiple-choice selection) as the test will.
Key: An answer to the Example that would earn full credit.	"Old DNA coding strand codon AGG => mRNA codon AGG => amino acid Arg. New DNA coding strand codon CGG => mRNA codon CGG => amino acid Arg. Since the old codon and the new codon both code for the same amino acid, Arg (arginine), no change in protein function is expected."	In addition to being a rudimentary rubric, students can use the Key for pre-test practice. If possible, put the Key in a separate file and encourage students to work through the Example before checking the Key.
Other Answers: examples of answers that would earn partial credit, with explanations of the scoring.	Example A: "The amino acid changes from Ser to Ala, so function might be impaired." [Shows understanding that a change in amino acid may alter function, but the conversions to amino acids were done wrong.] Example B: "Even though the amino acid is Arg in both cases, the change in DNA and RNA may cause problems." [The codons were translated correctly, but the answer does not convey that amino acid sequence dictates function.]	This optional TQT component can help students recognize and avoid common mistakes.

Table 1: *The structure and function of a Test Question Template*

**A multiple-choice version of this example might ask the same question and might provide choices like the following:*

- *(A) The amino acid changes from Ser to Ala, so the protein's function might be impaired.*
- *(B) The amino acid changes to a stop codon, so the protein is likely nonfunctional.*
- *(C) The change in DNA codon did not change the corresponding amino acid, so the protein's function should remain the same. [correct]*
- *(D) The change in DNA sequence might interfere with transcription.*

Additional TQTs for sophomore-level A&P courses -- with multiple-choice and short-answer examples -- are provided in the Appendix

continued on next page

TQTs are given to students prior to tests, and thus constitute a kind of study guide. In general, study guides and other practice problems, whether delivered in class or not, are useful to students in flagging certain information as important (Lieu et al. 2017). However, for a typical practice question, once students have arrived at a good answer, it is unclear what additional steps (if any) they should take to prepare for a test. Should they commit this particular answer to memory? Should they think of variations on the original question, and, if so, which aspects of the original question should remain fixed, and which should be varied? Or should they just move on to the next question in order to get through as many questions as possible?

TQTs avoid such dilemmas by explicitly showing students the relationship between the practice questions and the test questions. If an instructor delineates key course content in the form of a TQT, the students can use the TQT to practice for the test, and the instructor can use the same TQT to generate actual test questions. Therefore, the alignment between the practice and the test is excellent, even though the exact details of the test questions are appropriately hidden from students and can easily be changed in subsequent iterations of the course. This close alignment should benefit all students, but especially at-risk groups such as first-generation college students, who might otherwise struggle to understand the instructor's expectations for tests (Wright et al. 2016). Furthermore, students who are language learners are likely to benefit from increasingly explicit modes of practice similar in format to the assessments used (Abedi, 2001).

Table 2 compares TQTs with two common study question formats: the "fact check" and the "mini-essay." To be clear, we find all three formats useful. However, as detailed in Table 2, we believe that TQTs have several advantages: they are arguably more specific than mini-essays, more general than fact checks, and more transparent and adaptable than either of these other formats.

	Fact Check	Mini-Essay	Test Question Template
Description	Students are asked to name one or more correct specific facts.	Students are asked to write a short correct narrative (describe a process, compare two things, etc.).	Given some specific information (textual or graphical), students answer a question about that information.
Typical cognitive level in Bloom's taxonomy	1 (Knowledge), 2 (Comprehension)	2 (Comprehension), 3 (Application), 4 (Analysis), 5 (Synthesis), 6 (Evaluation)	2 (Comprehension), 3 (Application), 4 (Analysis), 5 (Synthesis), 6 (Evaluation)
Nervous System example (for sophomore-level A&P)	What are the two general types of Na^+ channels found in the cell membranes of neurons, and in which parts of the neuron (dendrites, soma, axon) is each type located?	Explain how an electrical signal is passed from one neuron to another.	Given a graph of membrane potential versus time at the axon hillock of a neuron, identify the EPSPs, IPSPs, and action potentials.
Cardiovascular System example (for sophomore-level A&P)	List the cardiac structures specialized for electrical conduction, in the order that they are depolarized.	Why do the pacemaker cells of the heart depolarize spontaneously? Explain in terms of ion channels.	Given a specific alteration in the cardiac conduction pathway, explain which ECG time intervals would be most affected, and in which direction (increased or decreased).
Specificity: Is it usually clear which specific facts are needed to answer the stated question?	Yes	Maybe	Yes
Generality: Does the question usually highlight a theme of general importance?	Maybe	Yes	Yes
Transparency: Does the study question familiarize students with the format of test questions?	Maybe	Maybe	Yes
Adaptability: Does the question usually permit many variations appropriate for student practice and testing?	No	No	Yes

Table 2: *A comparison of Test Question Templates with conventional practice question formats*

continued on next page

Tips for creating good TQTs

Those interested in the TQT format may wonder how they might create TQTs suitable for their own courses.

TQTs may be created for any topic in which many possible Inputs connect to many possible Outputs. Even a very straightforward template – for example, "Given the name of an organelle, summarize its function in a few words" may qualify as a TQT according to the definition outlined in Table 1, and may provide useful transparency to students. However, our main interest is in TQTs that encourage HOC, and that, therefore, are most appropriate for courses whose Learning Objectives (LOs) go beyond the mastery of factual details. If the LOs ask students only to *identify*, *list*, or *describe*, verbs commonly associated with the Knowledge and Comprehension levels of Bloom's taxonomy, there will not be many ways to ask about this information, and thus not enough variations to generate TQTs that require HOC. On the other hand, LOs that ask students to *analyze*, *critique*, *interpret*, or *predict*, verbs associated with the upper levels of Bloom's taxonomy, may be more readily translated into TQTs that require HOC.

An alternative to thinking in terms of LO verbs per se is the following; just about all HOC involves combining or reconciling two or more sources of information. Often, previously mastered background information (obtained from "source 1", often the textbook or teacher) is used to interpret a brand-new example (from "source 2", often the test itself). For the example in Table 1, background information on the genetic code (from source 1) is juxtaposed with the specifics of Jesse's DNA (from source 2). If we identify types of information that may be juxtaposed in numerous interesting ways, such that students need to think analytically, rather than simply memorizing the outcomes of all possible combinations, we may be able to craft a TQT that requires HOC.

An example of this "combinatorial" approach, showing how several aspects of the integumentary system might be combined in a TQT, is illustrated in Table 3. A conventional study question might ask students about relationships between UV light and other factors, with each relationship considered independently. The question can be made into a TQT by asking students to apply those known relationships to a new-to-them scenario about a particular patient. Only in the TQT version of the question do interesting interactions among the factors emerge, e.g., high melanin levels are beneficial in that they protect folate supplies and protect against skin cancer, but are undesirable in that they reduce endogenous vitamin D production and thus calcium absorption (HHMI Biointeractive 2015).

	Conventional study question	**Test Question Template**
Example	Indicate the impact (stimulatory, inhibitory, or neither) that UV light has on skin cancer incidence, melanin levels in the skin, rate of endogenous vitamin D production, rate of dietary calcium absorption, and plasma folate levels.	Given the results of an interview and physical exam of a patient -- including information on her natural melanin levels, sunlight exposure, and dietary (calcium, vitamin D, folate) habits and needs -- offer her sound medical or nutritional advice.
Comments on example	This is a straightforward five-part question: what effect does UV light have on each of five other variables?	This template covers the same variables as the conventional question, but is more interesting because there are more than five possible combinations to consider.
Number of possible variations of question	Few	Many
"Memorizability" of answers to all possible questions	High	Low
Importance of reasoning ability in answering the question	Low	High

Table 3: *Test Question Templates may motivate students to pursue true understanding rather than pure memorization*

continued on next page

While the TQT example in Table 3 is a clinical one, some clinical issues cannot be readily translated into TQTs. For example, if a sophomore-level physiology course covers only a couple of neuromuscular disorders, it would not make much sense to have a TQT of the form of, "given some symptoms, diagnose a patient's neuromuscular disorder." And if a course mentions numerous disorders, but only does so in passing, students may not have the knowledge needed to think analytically about the disorders. Thus, while clinical applications are often interesting to students and good for stimulating critical thinking, instructors who create clinical TQTs should take care not to presume background knowledge or analytical skills that their students do not yet have.

Whether clinically focused or not, most instructors probably have existing practice questions that can be converted into TQTs. The key is to recognize questions that represent specific examples of a general pattern. For example, imagine that your study guide asks students to calculate cardiac output from an end-systolic volume of 60 mL, an end-diastolic volume of 140 mL, and a heart rate of 50 beats per minute. Presumably you would like your students to be able to solve any similar problem involving the cardiac output equation. This expectation can be made more transparent by giving students a TQT such as the following: "Given values for three of the following four variables -- end-systolic volume, end-diastolic volume, heart rate, and cardiac output -- solve for the fourth variable." The original study-guide problem could then serve as an example of a question generated by this TQT.

An additional consideration is that, ideally, a TQT would reflect the type of active learning activity that was originally used to teach the content. For example, if students initially learned about histology by examining microscope slides, an ideal TQT might also involve the analysis of new (but related) microscope slides. However, if this is not feasible, the TQT could instead involve electronic images and/or text descriptions.

These last two examples -- the cardiac output calculation and the histology images -- illustrate the fact that simple mathematical analyses and figures are often an excellent basis for TQTs. For equations and graphs, the exact numbers and curve shapes can be varied endlessly, allowing students to get unlimited practice on important mathematical relationships. This also furthers the important goal of integrating more quantitative skill development into biology (Brewer and Smith 2011). Likewise, for qualitative figures, there are numerous versions that students have not seen before, yet should be able to analyze after previous experience with similar figures.

A variation on the use of qualitative figures is the following, which we have borrowed from the "Public Exam" system (Wiggins 2019). Give students a specific figure to study in advance, often a complex but important one (e.g., of the blood clotting cascade), along with examples of questions that could be asked about the figure. Assure students that they will have a copy of the figure during the test. The clear message to students is that this figure should be understood in depth, but does not need to be memorized. The advantage to instructors, in this case, is that they do not need to search for novel figures; they can simply use existing figures that are already central to their lessons.

Supporting students' use of TQTs

Many undergraduate students think of biology as an endless series of specific, unique questions, each with its own specific, unique answer. Since TQTs, by contrast, involve more general patterns of questions, students deserve explicit guidance on this alternative approach. If TQTs will be used to generate test questions, they should be introduced and explained during formative assessments. Cooperative learning formats such as think-pair-share (Mazur 1996), and jigsaws such as those used in Theobald et al. 2017, can be especially beneficial because students can help each other with the format of the questions, as well as the relevant content. Students who encounter TQTs early and often will be well-equipped to handle TQT-generated questions on actual tests.

Even if the TQT format is explained carefully, with examples, many students may simply study the instructor-provided examples without creating their own additional examples. However, such students would miss a primary advantage of TQTs, namely, the opportunity to create their own additional practice questions. For this reason, students should be explicitly assigned the task of creating novel TQT-based questions. This assignment can benefit from a group-learning context; students can check each other's creations, try to solve the ones that seem most plausible as test questions, and grade each other's answers. Instructors should monitor these efforts, if possible, to ensure that no group is straying too far from the original intent of each TQT.

Preliminary student feedback on TQTs

In order to examine student views on topics covered in this paper, we administered a brief questionnaire to students who had taken a course that used TQTs. (The Everett Community College IRB affirmed that this survey was exempt from formal IRB review.) Of the 76 students enrolled in two courses, 35 opted to read this manuscript and respond to the questionnaire. The questionnaire consisted of three open-ended questions:

1. In your view, what are the main CLAIMS (central assertions or arguments) being made by this paper? Please list at least two claims.
2. To what extent do you agree or disagree with the claims listed in the previous response? Please be assured that it is OK to disagree! Explain why you agree and/or disagree with each claim.
3. Aside from the claims covered by the previous two questions, what other comments (if any) do you have about this paper?

continued on next page

We coded students' responses to the first question as high, moderate, or low with regard to their level of comprehension of the manuscript. We excluded from the full analysis any student whose responses reflected a low level of comprehension, in that the response did not identify a claim of the paper or lacked sufficient detail to assess comprehension. We then coded the remaining 30 student responses with the qualitative data analysis software program, MAXQDA, applying a grounded theory approach to text analysis (Strauss and Corbin 1998), which is an iterative inductive process of coding concepts that arise from the data and linking these concepts into themes. Our analysis yielded four themes which consistently arose across student responses.

The most general of these four themes was the idea that active learning in general (not necessarily TQTs in particular) is helpful for student learning. For example, one student wrote, "We need more active learning in classrooms so students better understand material and not just memorize facts."

Two additional themes centered more specifically on TQTs. These themes were, first, that TQTs especially help students learn new material more deeply, and, second, that TQTs are useful for reviewing previously covered material in preparation for exams. The first theme was exemplified by student comments such as the following: "Incorporating TQTs … will better promote active learning in students by not just memorizing facts given, but to further understand the context and apply it to deeper thinking problems." The second theme, focusing more on exam preparation, was echoed in comments such as, "As a student, when I have sample question [sic] that include the same material that will be on the exam I study more and make sure that I actually know the material even if the sentence gets rearranged or numbers get changed. It makes it easy for students to know exactly what to study and focus more on, instead of studying the whole book. I think its [sic] really unrealistic when a teacher makes you study absolutely everything they covered in class."

All of the students agreed in part or in whole with the claims they identified in the manuscript. The few who agreed only in part largely expressed an additional theme, namely, that students may struggle with how to use TQTs when first introduced to them. One representative comment was the following: "TQTs can be good as long as students understand how to solve them and understand them. And so I believe for TQTs to work, it would need to be integrated in school learning…. It can be very frustrating and confusing when given something like a TQTs question on the exam and you never came across that type of question before." This insight, that even an otherwise helpful format such as a TQT can be confusing in the absence of repeated exposure, underscores our suggestion above that students be given explicit, extensive training in TQTs well before the first exam.

Other approaches

As noted earlier, promoting HOC, both during personal practice and during summative assessments, is considered an educational best practice (Pellegrino 2014). The literature in developing assessments that promote higher-order thinking and learning in the K-12 system is rich (e.g., Darling-Hammond and Adamson 2014). TQTs thus represent one of many possible ways of aligning practice and assessment of HOC. Analogous approaches in fields that habitually use algorithmic or equation-based calculations may routinely do TQT-like work without explicitly stating it. Specifically, physics and engineering exams commonly include questions on the application of formulae to physical problems. The TQT framework may help to make the design of these assessments more transparent for students. Learning progressions are an effective way to formalize assessment-ready pieces of curriculum by investigating how students proceed towards mastery (Scott et al. 2019). Development of learning progressions is laborious but highly useful and will achieve goals overlapping with those of TQTs. Other formulations of exam-presentation tools for students, such as the "Public Exam" system (Wiggins 2019), may also be useful to educators looking to move beyond traditional exam styles.

Summary

Biology instructors want to help their students go beyond isolated facts to master fundamental, general patterns and skills. In-depth active-learning activities are great for inspiring deep learning, but students may not fully engage in these activities if they do not see strong connections between these learning activities and subsequent high-stakes exams. TQTs represent an approach, newly formulated, yet similar to other existing frameworks, to bridge the gap between practice questions and exam questions. By improving the transparency of instructors' exam-writing, TQTs help instructors and students avoid the cat-and-mouse game of "What's going to be on the exam?" and instead provide students with rich opportunities for creative, analytical practice with the course's most important material.

Acknowledgments

Mentorship of GJC by LDJ was supported in part by a PALM Fellowship (PI: Susan Wick, University of Minnesota). We thank Susan Wick for encouragement and comments on a draft of this paper, and Subin Hyun for illustrations in the Appendix. We also thank Jonathan Pottle for feedback on GJC's ideas for an "A&P manifesto," which indirectly fueled the writing of the present article.

continued on next page

About the Authors

Greg Crowther teaches human anatomy and human physiology at Everett Community College. When time permits, he studies the incorporation of the arts into STEM education.

Kiki Jenkins is based at Arizona State University, where she teaches various interdisciplinary courses and conducts research on marine conservation, marine technology, and the role of science dance in social change.

Ben Wiggins is the Manager of Instruction at UW Biology, where his research focuses on active learning in large classrooms. He teaches in molecular biology, civilizational biology, and science teaching methods.

Literature cited

Abedi J and Lord C. 2001. The language factor in mathematics tests. *Appl Meas Educ* 14(3): 219-234.

Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich, PR, ... & Wittrock MC. 2001. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition. White Plains, NY: Longman.

Bonwell CC and Eison JA. 1991. Active learning: creating excitement in the classroom. ASHE-ERIC Higher Education Reports. Washington, DC: The George Washington University, School of Education and Human Development.

Brewer CA, Smith D. 2011. Vision and change in undergraduate biology education: a call to action. Washington, DC: American Association for the Advancement of Science.

Crowe A, Dirks C, Wenderoth MP. 2008; Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sci Educ* 7(4):368-381.

Darling-Hammond L, Adamson F. 2014. Beyond the bubble test: How performance assessments support 21st century learning. San Francisco, CA: John Wiley & Sons.

Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci U S A* 111(23):8410-8415.

Goubeaud K. 2010. How is science learning assessed at the postsecondary level? Assessment and grading practices in college biology, chemistry and physics. *J Sci Educ Technol* 19(3): 237-245.

HHMI Biointeractive. 2015. The biology of skin color. https://www.biointeractive.org/classroom-resources/biology-skin-color.

Lieu R, Wong A, Asefirad A, Shaffer JF. 2017. Improving exam performance in introductory biology through the use of preclass reading guides. *CBE Life Sci Educ* 16(3): ar46.

Mazur E. 1996. Peer instruction: A user's manual. Boston: Pearson.

Miller CJ, Metz MJ. 2014. A comparison of professional-level faculty and student perceptions of active learning: its current use, effectiveness, and barriers. *Adv Physiol Educ* 38(3): 246-252.

Momsen JL, Long TM, Wyse SA, Ebert-May D. 2010. Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE Life Sci Educ* 9(4):435-440.

NGSS Lead States. 2013. Next generation science standards: For states, by states. Washington, DC: National Academies Press.

Patrick LE, Howell LA, Wischusen W. 2016. Perceptions of active learning between faculty and undergraduates: Differing views among departments. *J STEM Educ Innov Res* 17(3): 55-63.

Pellegrino JW. 2006. Rethinking and redesigning curriculum, instruction and assessment: What contemporary research and theory suggests. Chicago: National Center on Education and the Economy.

Pellegrino JW. 2014. A Learning sciences perspective on the design and use of assessment in education. In: Sawyer RK, editor. The Cambridge handbook of the learning sciences. Cambridge, UK: Cambridge University Press.

Reeves TC. 2006. How do you know they are learning? The importance of alignment in higher education. *Int J Learn Technol 2*(4): 294-309.

Scott EE, Wenderoth MP, Doherty JH. 2019. Learning Progressions: An Empirically Grounded, Learner-Centered Framework to Guide Biology Instruction. *CBE Life Sci Educ* 18(4): es5.

Silverthorn DU. 2020. When active learning fails … and what to do about it. In: Mintzes JJ, Walter EM, editors. Active learning in college science: The case for evidence-based practice. Berlin: Springer, p. 985-1001.

Songer NB, Kali Y. 2014. Science education and the learning sciences as coevolving species. In: Sawyer RK, editor. The Cambridge handbook of the learning sciences. Cambridge, UK: Cambridge University Press.

Smith MK, Wenderoth MP, Tyler M. 2013. The teaching demonstration: what faculty expect and how to prepare for this aspect of the job interview. *CBE Life Sci Educ* 12(1): 12-18.

Stains M et al. 2018. Anatomy of STEM teaching in North American universities. *Science* 359: 1468-1470.

Strauss AL and Corbin J. 1998. Basics of qualitative research: Techniques and procedures for developing grounded theory. Thousand Oaks, CA: SAGE Publications, Inc.

continued on next page

Tanner KD. 2013. Structure Matters: Twenty-One Teaching Strategies to Promote Student Engagement and Cultivate Classroom Equity. *CBE Life Sci Educ* 12(3): 322-331.

Theobald EJ, Eddy SL, Grunspan DZ, Wiggins BL, Crowe AJ. 2017. Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PLoS ONE* 12(7): e0181336.

Wiggins B. 2019. The Public Exam System: Simple Steps to More Effective Tests. https://www.coursehero.com/faculty-club/classroom-tips/benjamin-wiggins/.

Wiggins G, McTighe J. 2005. Understanding by Design [2nd edition]. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Wright CD, Eddy SL, Wenderoth MP, Abshire E, Blankenbiller M, Brownell SE. 2016. Cognitive Difficulty and Format of Exams Predicts Gender and Socioeconomic Gaps in Exam Performance of Students in Introductory Biology Courses. *CBE Life Sci Educ* 15: 1-16.

Appendix Hyperlink

Follow the link below to view and download the pdf appendix to this article.

https://www.hapsweb.org/page/crowtherappendix

Back to TOC