

## Article

# Verbal Final Exam in Introductory Biology Yields Gains in Student Content Knowledge and Longitudinal Performance

Douglas B. Luckie,<sup>\*†</sup> Aaron M. Rivkin,<sup>\*</sup> Jacob R. Aubry,<sup>\*</sup> Benjamin J. Marengo,<sup>\*</sup> Leah R. Creech,<sup>†</sup> and Ryan D. Sweeder<sup>†</sup>

<sup>\*</sup>Department of Physiology, Michigan State University, East Lansing, MI 48825; <sup>†</sup>STEM Learning Laboratory, Lyman Briggs College, Michigan State University, East Lansing, MI 48824-3320

Submitted April 25, 2012; Revised May 7, 2013; Accepted May 8, 2013  
Monitoring Editor: Michelle Kathleen Smith

We studied gains in student learning over eight semesters in which an introductory biology course curriculum was changed to include optional verbal final exams (VFs). Students could opt to demonstrate their mastery of course material via structured oral exams with the professor. In a quantitative assessment of cell biology content knowledge, students who passed the VF outscored their peers on the medical assessment test (MAT), an exam built with 40 Medical College Admissions Test (MCAT) questions (66.4% [ $n = 160$ ] and 62% [ $n = 285$ ], respectively;  $p < 0.001$ ). The higher-achieving students performed better on MCAT questions in all topic categories tested; the greatest gain occurred on the topic of cellular respiration. Because the VF focused on a conceptually parallel topic, photosynthesis, there may have been authentic knowledge transfer. In longitudinal tracking studies, passing the VF also correlated with higher performance in a range of upper-level science courses, with greatest significance in physiology, biochemistry, and organic chemistry. Participation had a wide range but not equal representation in academic standing, gender, and ethnicity. Yet students nearly unanimously (92%) valued the option. Our findings suggest oral exams at the introductory level may allow instructors to assess and aid students striving to achieve higher-level learning.

## INTRODUCTION

The spoken word is a unique device. Our words can affect others' behavior in dramatic ways and can profoundly affect our own thoughts and learning. Whether one reviews the classic work of Piaget and Vygotsky or other literature of cognitive development (Inhelder and Piaget, 1958; Vygotsky, 1962; Bruner, 1973), learning (Posner *et al.*, 1982; Guest and Murphy, 2000; Krontiris-Litowitz, 2009), memory (Baddeley and Hitch, 1974), or metacognition (Schunk, 1986; Pintrich, 1988; Labuhn *et al.*, 2010), the data support that the action of speaking thoughts aloud strongly impacts one's own un-

derstanding and mastery. In fact, for many years, the literature has even suggested that a student who speaks aloud his or her answer to a question will have higher retention and comprehension than the student who silently thinks the same answer (Gagne and Smith, 1962; Carmean and Wier, 1967; Bruner, 1973). Overt verbalization is a powerful tool for learning and increases self-evaluative judgment and cognitive commitment (Wilder and Harvey, 1971; Schunk, 1986; Labuhn *et al.*, 2010).

The literature suggests that exams and other high-stakes assessments tend to drive student learning in the classroom (Tobias, 1992; Tobias and Raphael, 1997). The data are persuasive, and upon reflection, most experienced instructors would concede this is very likely to be true. No matter how much time and energy a professor puts into challenging students to think critically, no matter how much we implore them to think like a scientist, if the exams they face do not test higher-level thinking and problem solving, students are not likely to master the material (Traub and MacRury, 1990; Momsen *et al.*, 2010). Students perceive that the information on our high-stakes assessments is what we truly value (Entwistle and Entwistle, 1992), and the exam format itself can change how students study and approach problem solving.

DOI: 10.1187/cbe.12-04-0050

Address correspondence to: Douglas Luckie (luckie@msu.edu).

© 2013 D. B. Luckie *et al.* CBE—Life Sciences Education © 2013 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution-Noncommercial-Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

As instructors, our exams and other assessments certainly also impact us. We genuinely want our students to learn things that Benjamin Bloom (1956) would have considered higher level, but we struggle with how to assess it. Consider this learning goal: to think like a scientist and be able to adaptively negotiate a question or problem. It is an appropriate learning goal, but if we attempt to align it with our traditional assessments (quizzes, exams, homework, etc.), we rarely find robust sources of evidence that demonstrate mastery (Momsen *et al.*, 2010). In mentoring our graduate students, we tend to teach with a questioning style much like that of Socrates (Stumpf, 1983). We interact one-on-one and ask questions to evaluate as well as to determine how to best guide our students. In this setting, we orally assess whether our students are mastering or achieving “how to think like a scientist” (Paul and Elder, 2006).

A number of years ago, a new assessment was introduced in our biology classroom, a structured clinical interview or oral exam with Socratic questioning called the “verbal final” (VF). It was used to assess higher-level thinking, while at the same time allowing verbalization to better impact learning. We observed students’ ability to think critically about topics taught and adaptively negotiate a posed question or problem as addressed in the literature (Schubert *et al.*, 1999; Thorburn and Collins, 2006). The assessment was predicted to allow students to also participate in a metacognitive process as they grappled with the topic and how to explain it to others (Baddeley and Hitch, 1974). The buy-in from students created a novel culture of learning outside the classroom driven by this high-stakes assessment (DiCarlo, 2009) and appears to have a lasting impact on student learning.

## METHODS

### *Population and Context*

Lyman Briggs College is an undergraduate science program established at Michigan State University (MSU) in 1967 (Sweeder *et al.*, 2012). It is a residential college modeled after those at Oxford University in the United Kingdom and has a focus of educating undergraduates in a liberal science curriculum defined as a “solid foundation in the sciences and a significant liberal education in the history, philosophy, and sociology of science.” Introduction to Cell and Molecular Biology (Bio II) is a five-credit, freshmen-level course. It is the second in a core two-semester introductory biology sequence for science majors. It has an enrollment of ~100 students and each student attends two lectures and one recitation (50 min each) and two laboratory sections (3 h each) per week.

With the approval of MSU Institutional Review Board (IRB X00-475 and 10-543), student data were collected from 445 students who completed our Introduction to Cell and Molecular Biology course at MSU; the data were from cohorts in the eight semesters studied between the years 2002 and 2011. All students were enrolled at MSU, and participant consent was obtained while students were enrolled in the course. The registrar provided ACT science and composite scores and demographic data for 407 of the 445 students who participated. With the approval of MSU-IRB (IRB 07-446), student data were also collected for longitudinal tracking and compari-

son purposes from all 21,528 students who completed select upper-level science courses at MSU from 1997 to 2010.

The VF process is structured to incentivize students to participate with high reward and to decrease anxiety by having no penalty. The 1-h exam is graded on a pass/not-pass basis. Students who pass the VF earn a 100% for the final exam score in the course. Students may retake the VF a number of times. If they do not pass, they take the written final exam. In addition to the method listed below (and in greater detail in the Supplemental Material), we have made additional resources available, including audio and video recordings of students taking the VF, on the public VF website ([www.msu.edu/~luckie/verbalfinals](http://www.msu.edu/~luckie/verbalfinals)).

### *VF Protocol*

During the exam process, instructors keep notes of questions asked and student errors that occur (Figure 1). They also make an effort to frequently use the phrase “What do you predict is likely?” to communicate being open to imperfect answers as long as students provide reasonable support for their predictions. If students are performing at a very high level, then assessment of even more advanced concepts, such as experimental design, can be attempted; for example, “Okay, great prediction, now design an experiment to test that prediction, what could you try?”

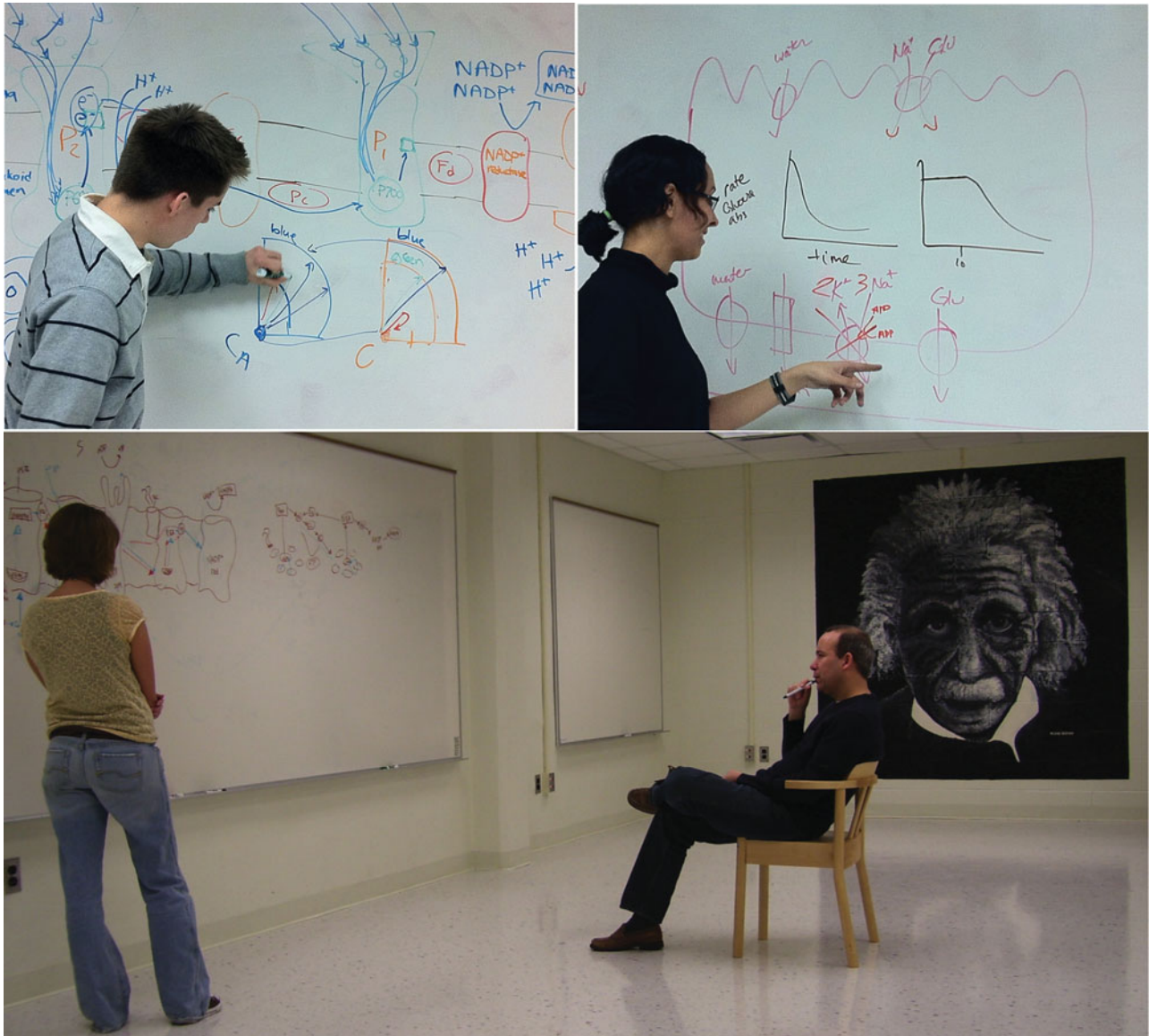
At the beginning of the exam, the instructor reads an introductory statement aloud to outline what is expected and give the guidelines for the process. After this, several topics are discussed in sequence. Below are the topics discussed in the VF used for this study (photosynthesis, digestion/absorption, central dogma/biosynthesis).

### *Phase 1. Explanation of Process of Photosynthesis*

*Light Reactions.* The first phase of the exam is designed to require the student to draw and explain the light reactions of photosynthesis. At the start of the official exam time period, the student is told to draw an illustration of the photosystems and carriers important for light reactions. This illustration serves as an aid as the student explains the process. After the student is done, the instructor asks probing questions; first about whatever the student said that might not have been completely clear or accurate, then to determine the depth of student understanding. Example questions are: What is an absorption spectrum? How do the pigments absorb light in a photosystem? If there were 100 protons in the stroma and 200 in the thylakoid lumen how much ATP do you predict would be made? What do you predict would happen if we put a hole in the membrane?

When the student successfully completes phase 1, he or she is told of any strikes (significant errors) he or she may have accrued. If there are fewer than three strikes and time remains in the hour, the student may proceed.

*Calvin Cycle.* The student is instructed to draw and explain the Calvin cycle, including names of enzymatic and structural compounds. When the student is done explaining the process, the instructor again asks questions to clarify and check for errors and omissions, and then follows this up with probing



**Figure 1.** The event of the VF gives the student one-on-one time with the professor. They both follow a structured exam protocol outlined in advance. The topics discussed during the VF are framed in one, albeit simplistic, story with three parts: how photosynthesis converts light into sugar/food, how humans digest and circulate food, and how a pancreatic cell makes insulin in response to the presence of sugar. The student creates an illustration and explains the way he or she believes the biology works. Once the student completes the explanation about each topic, the professor then asks questions for clarification and follows up with probing questions to test the student's confidence and depth of knowledge. (Permissions for publication of photos were obtained.)

questions: What are the names of phases and why? What is reduction? What would happen if we fixed only one  $\text{CO}_2$  molecule? What does G3P taste like, and why does that make sense from an evolutionary perspective?

If the student successfully completes the discussion of the Calvin cycle, he or she is then told of any additional strikes he or she may have accrued. As was the case before, if there are fewer than three strikes and time remains in the hour, the student may proceed. As a transition, we state: "Okay, let's pretend the glucose you just made in the Calvin cycle turns into a donut. Let's discuss its digestion and absorption."

### *Phase 2. Explanation of the Process of Digestion and Absorption*

*Digestion and Absorption of Macromolecules.* In the second phase, the student is asked to explain digestion and absorption. We require students to know a sample set of organs and their functions (oral cavity, stomach, small intestine, liver, gall bladder, and pancreas, as well as a set of events; organ function; relevant cells; relevant enzymes; hormones; processes), but for efficiency sake, we often ask them to explain how digestion works in organ X, or alternatively, to explain how food group X is digested and absorbed. Randomizing which

organ or macromolecule they are asked to discuss requires students to prepare for all questions, but less time is required for the actual exam. When the student has completed his or her explanation of digestion the instructor may ask questions for clarification or to check for errors and omissions, but does not probe the organ level of this topic deeply. Instead, the instructor asks the student to then focus on cell biology. Our goal is to more quickly get to the topics of membrane transporters, gradients, and enzymes in the domain of cell biology, using the parietal cell and/or the villus epithelial cell as models in context.

*Parietal or Villus Cell Biology.* In this section of the digestion phase, we probe students' understanding of how a cell works and ask students to make predictions about how membrane transporters and chemical reactions may respond to perturbations. For example, we will ask a student to draw a stomach's parietal cell and explain how it makes HCl or to draw the small intestine's villus epithelial cell and explain how it absorbs glucose. We evaluate the explanation and probe for understanding with questions such as: What would happen to this process in the parietal cell if a drug inhibited the enzyme carbonic anhydrase? Indicate which reaction rates would change by placing a check mark next to its arrow in your drawing, how might that change the pH in the stomach lumen or the bloodstream? Similar transporter questions can be used for the villus epithelial cell; for example: What would happen if a drug inhibited the Na/K pump? How would this affect other transporters? How would this affect the membrane potential? How would this affect the rate of glucose absorption/uptake? We vary and alter these questions for different students by, for example, first asking what might happen if the membrane potential is changed.

As an affective transition to the last phase of the test, we request that the student tell us some detail of the path an absorbed macromolecule (e.g., glucose) must take to travel from a capillary at the basolateral side of the villus cell to a capillary beside a beta cell in the pancreas. The student is expected to name some basic vessels and the path (valves/chambers) through the heart. We often request they do this without an illustration, if possible. When the student is done, we ask questions only if errors or omissions occurred; otherwise, we talk about the last steps, in which glucose exits the capillary and enters the beta cells. We probe for understanding of diffusion via questions such as: How does glucose exit the capillary as well as into the pancreatic cell? Does it need a carrier or channel or something? Why does it move?, What drives it? Why do you predict this would occur?

### **Phase 3. Central Dogma—Biosynthesis of Insulin**

*Biosynthesis of Proteins.* The last phase of the VF requires the student to explain central dogma using the biosynthesis of a particular protein, usually insulin, as we discuss insulin biosynthesis frequently in class. Students are required to draw a eukaryotic cell and explain the processes of transcription, translation, and biosynthesis in terms of steps that occur at each organelle. When the student has completed his or her explanation of insulin biosynthesis and secretion, we ask follow-up questions for clarity and to check for errors and omissions, and then probe with queries such as: What if

there's a mutation? What do you predict would happen if we add some hydrophobic domains to the gene?

*Biosynthesis of Other Proteins.* After the student shows an acceptable understanding of how insulin is created, we ask what path would be taken in the biosynthesis of an apical channel protein (e.g., CFTR), or a cytoplasmic protein (e.g., tubulin). We ask the student to draw a large transport vesicle and indicate where insulin or CFTR might be found. We also often request that a student create an illustration to show how a transport vesicle behaves when it reaches the cell membrane.

### **Timing and Constraints of VF**

We have refined the process to avoid wasting time when student performance indicates that a student was unable to prepare adequately. We start promptly and end promptly. If the student has not completed the exam by the end of 60 min, we just stop and require the student to set up another attempt. If it quickly becomes clear during the exam that a student does not yet know basic information, we just end the exam and ask the student to make another appointment. To communicate these rigorous guidelines regarding time and other elements, we provide handouts at the start of the semester, and the instructor reads aloud an introductory statement the first time a student takes the exam. To increase student preparation and performance, during the eight-semester period discussed in this study, the exam process was considered a public one, with classroom doors kept open and interested colleagues permitted to observe if granted prior permission by student.

### **Collecting Data on Student Performance at End of Semester**

We administered a small, standardized exam, the Medical Assessment Test (MAT) as a posttest during the final week (week 15) of all semesters. The MAT exam was built from Medical College Admissions Test (MCAT) test questions developed, validated, and purchased from the Association of American Medical Colleges (Luckie *et al.*, 2004). The MAT exam is a 40-question, multiple-choice test composed of relevant passage-style questions. MCAT passage questions have been studied by others and are deemed to assess higher-level content knowledge than typical multiple-choice exams (Zheng *et al.*, 2008). The MAT exam instrument consisted of questions from five general topic categories: cell structure and function, oncogenes/cancer, cellular respiration, microbiology, and DNA structure and function. Performance of each individual student on the exam as a whole and on questions related to each category was examined, and cohorts were compared. The MAT instrument used in this study is provided online (see VF website). In addition to the MAT instrument, the course written final exam was not changed and was therefore standard across semesters and served as an additional assessment of content knowledge.

Raw MAT scores for all semesters were normalized for variations in each cohort's prior academic performance using ACT scores. In our previous tracking research (Creech and Sweeder, 2012), we determined incoming university grade point average (GPA) tends to be the best predictor of how well a student will do in a university course. Yet the first-semester university GPA, much like high school GPA, is less

predictive. Given that our freshmen recently took the ACT, a carefully administered, valid, and reliable exam, we felt the ACT was the best available metric to use for normalization of data for this particular freshmen biology course.

Student ACT science scores ranged from 15–36 and ACT composite scores from 17 to 35. MAT scores were normalized by determining the mean ACT score for each experimental and control group and calculating the numerical multiplier required to elevate the lower score to parity. This multiplier was then used to adjust respective MAT scores proportionally to normalize all results. VF participation and pass rates were calculated for cohorts by using academic standing, gender, and ethnicity (four students did not report ethnicity). Academic standing quartiles were generated by sorting students first for ACT composite score and secondarily for ACT science score. The resulting sorted list was then divided into equal quartiles. The lowest quartile contained students with composite scores below 25. The cutoff for the second-lowest quartile was a composite score of 27 and science score of <26. The next quartile ranged from students with a composite score of 27 and a science score of 26 or higher to a composite of 29 and a science score of 28 or lower. The highest quartile consisted of composite scores of 29 and a science score of 29 or higher and all higher composite scores.

Student's *t* test, analysis of variance (ANOVA), and *z* test of proportions were used to perform statistical comparisons of MAT data. The *t* test was used in comparisons between the experimental cohort (students who passed the VF) and the matched peer control cohort (students who did not pass) as well as comparisons between pooled data sets of experimental and control groups. ANOVA was also used. The figure legends indicate trial numbers and test applied. Error bars on figures were generated by calculating the SEM, unless otherwise indicated.

### Longitudinal Analysis of Student Performance in Upper-Level Courses

Tracking analysis of student performance in their upper-level science courses was performed. Grades earned by the cohort of all students who passed the VF in Introduction to Cell and Molecular Biology (LB 145, Bio II) were compared with grades earned by other MSU students who enrolled in upper-level science courses. The courses used were Physiology I (PSL 431) and II (PSL 432), Biochemistry I (BMB 461) and II (BMB 462), Advanced Introductory Microbiology (MMG 301), Genetics (ZOL 341), Organic Chemistry I (CEM 251) and II (CEM 252), and Advanced Organic Chemistry I (CEM 351) and II (CEM 352). The data set consisted of grades in upper-level science courses, introductory courses (biology, chemistry, and organic chemistry), cumulative GPA by semester, and student classification information (gender, ethnicity, major, honors college, Lyman Briggs College). Data were analyzed using SPSS 20 (IBM, Chicago, IL).

We examined student performance in multiple fashions. We initially compared students who passed the VF with all MSU students between 1997 and 2010; when the data were visualized, we saw higher performance of students who passed the VF in a wide range of upper-level science courses. However, during analysis, we realized that 1) student incoming ACT scores increased over the time and 2) students in the Lyman Briggs College major tend to have better incoming

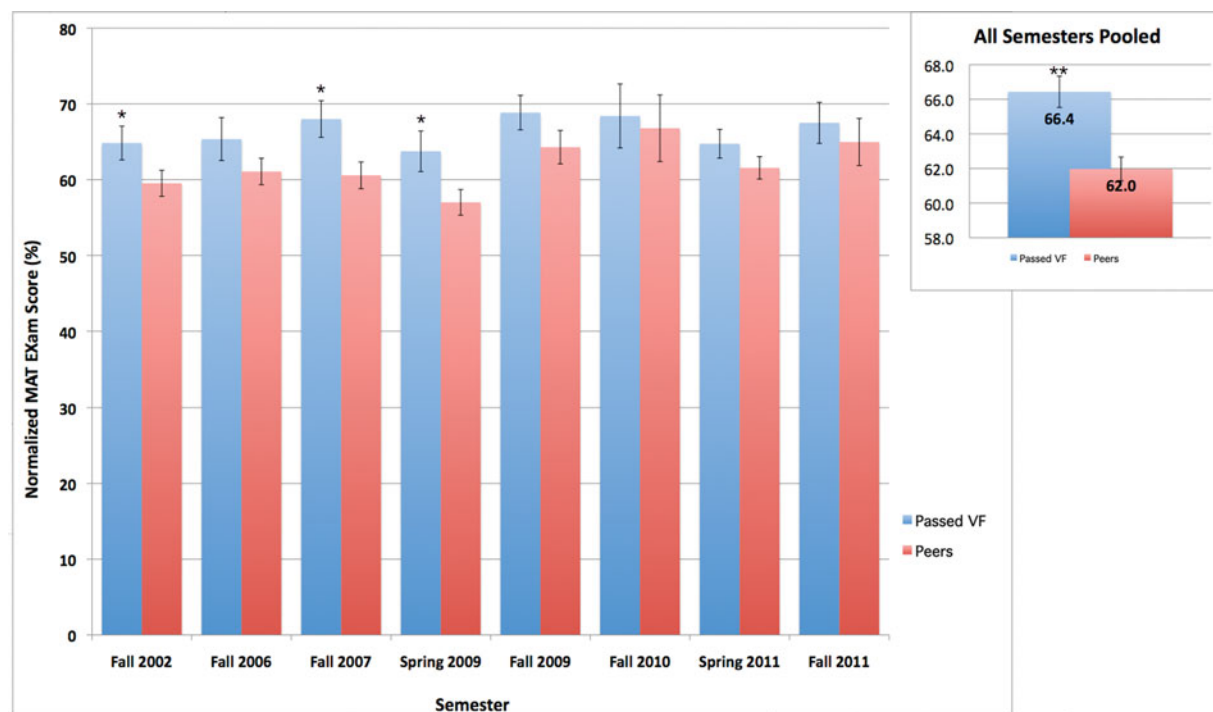
**Table 1.** Tracking performance in upper-level courses of students who passed VF vs. controls<sup>a</sup>

| Course | Cohort  | Average grade (SEM) | <i>n</i> |
|--------|---------|---------------------|----------|
| ZOL341 | VF      | 3.26 (0.12)         | 38       |
|        | Control | 2.80 (0.10)         | 98       |
| MMG301 | VF      | 2.69 (0.18)         | 29       |
|        | Control | 2.69 (0.12)         | 59       |
| PSL431 | VF      | 2.72 (0.21)         | 32       |
|        | Control | 2.42 (0.12)         | 85       |
| PSL432 | VF      | 2.78 (0.21)         | 30       |
|        | Control | 2.47 (0.13)         | 78       |
| CEM251 | VF      | 3.44 (0.10)         | 65       |
|        | Control | 2.88 (0.07)         | 180      |
| CEM252 | VF      | 3.39 (0.12)         | 60       |
|        | Control | 2.79 (0.08)         | 168      |
| CEM351 | VF      | 3.70 (0.13)         | 15       |
|        | Control | 3.18 (0.14)         | 31       |
| CEM352 | VF      | 3.44 (0.19)         | 17       |
|        | Control | 3.26 (0.19)         | 27       |
| BMB401 | VF      | 3.15 (0.33)         | 13       |
|        | Control | 2.23 (0.17)         | 37       |
| BMB461 | VF      | 3.11 (0.10)         | 47       |
|        | Control | 2.51 (0.09)         | 104      |
| BMB462 | VF      | 2.91 (0.12)         | 40       |
|        | Control | 2.49 (0.09)         | 85       |

<sup>a</sup> Average grade in course of students who passed the VF compared with average grade of peer controls from the same classes (*n* = number of students in sample). CEM, chemistry; BMB, biochemistry; PSL, physiology; MMG, microbiology; ZOL, zoology.

GPA. When aligning the time period to 2002–2010 and examining peer students who were Lyman Briggs majors, we still saw higher grades but smaller to nonexistent significance. Given this approach suffers from the fact that there are multiple instructors teaching in Lyman Briggs courses, our final analysis looked exclusively at those peer-matched students who had enrolled in the lower-division LB 145 (Bio II) course over the same time period and with the same instructor who offered the VF exam option. Table 1 lists courses studied in the analysis and side-by-side comparison of final GPA of the experimental and control cohorts in each course. In addition, Figure 5 (later in this article) compares ACT quartile versus earned course grade average in physiology and biochemistry courses. Quartiles were established using ACT scores, as described above. The trial numbers (*n*) as indicated are lower in the analysis, because not all students in the study had taken their upper-level science courses nor do all students enroll in all the courses we studied. An expansive Supplemental Table S1 lists all courses studied in the analysis and side-by-side comparison of final GPA of the experimental and control cohorts in each quartile in each course.

Additional methods of data analysis included linear modeling (LM) and graphical analysis. LM was also used to create a model for estimating students' upper-level science performance (Spicer, 2005; Rauschenberger and Sweeder, 2010). LM used the raw data and was executed stepwise, excluding cases listwise. Graphical representations of the data were created through SPSS 20 to visualize categorical differences in the data. Average grades in specific courses for students who did and did not take the VF are compared for students clustered based on incoming GPA and using scatter plots (Supplemental Figure S1). Students' GPAs were clustered to the nearest



**Figure 2.** Performance of students who passed the VF vs. peers on a standardized MAT content exam. The blue bar corresponds to the collective normalized average on the MAT of students who passed the VF each semester. Likewise, the red bar illustrates normalized MAT scores of peers who did not pass the VF. Fall 2002 ( $n = 77$  students, 31 passed:46 not), Fall 2006 ( $n = 63$ , 21:42), Fall 2007 ( $n = 64$ , 20:44), Spring 2009 ( $n = 70$ , 16:54), Fall 2009 ( $n = 42$ , 13:29), Fall 2010 ( $n = 19$ , 11:8), Spring 2011 ( $n = 77$ , 29:42), and Fall 2011 ( $n = 33$ , 19:14). The inset compares average MAT scores of all semesters pooled for those who passed the VF ( $n = 160$ ) and those who did not ( $n = 285$ ). \*,  $p < 0.05$ ; \*\*,  $p < 0.001$ ;  $t$  test compared students who passed the VF with cohorts who did not for each course/semester and for pooled data in inset.

quarter grade in the plots, and the error bars represent the SEM.

### Collecting Student Opinion Data via Student Self-Report on Course Evaluation Forms

End-of-semester course evaluation surveys were used to evaluate student opinion and assess affective or qualitative elements in response to the optional VF. We evaluated the course using the Student Assessment of Learning Gains (SALG; Seymour *et al.* 2000). Beginning in 2009, a new prompt, “Do you think the VF option helps student learning?,” was added to the course evaluation form administered at the end of each semester in Bio II. Response choices were: 1—no help, 2—a little help, 3—moderate help, 4—much help, 5—great help, 6—not applicable. In addition, in Fall 2010 and 2011, the prompt “Do you think the VF option helps student learning, if so why?” also solicited student feedback via extended written responses. We reviewed the Likert-scale responses to questions posed on the topic of the VF and the written comments prompted by extended-response questions. While only sample feedback is presented in this report, comprehensive student feedback from five semesters (Spring 2009:  $n = 72$ ; Fall 2009:  $n = 43$ ; Fall 2010:  $n = 20$ ; Spring 2011:  $n = 77$ ; and Fall 2011:  $n = 34$ ) is also provided online (see VF website). All student comments are hyperlinked to their full SALG feedback form online.

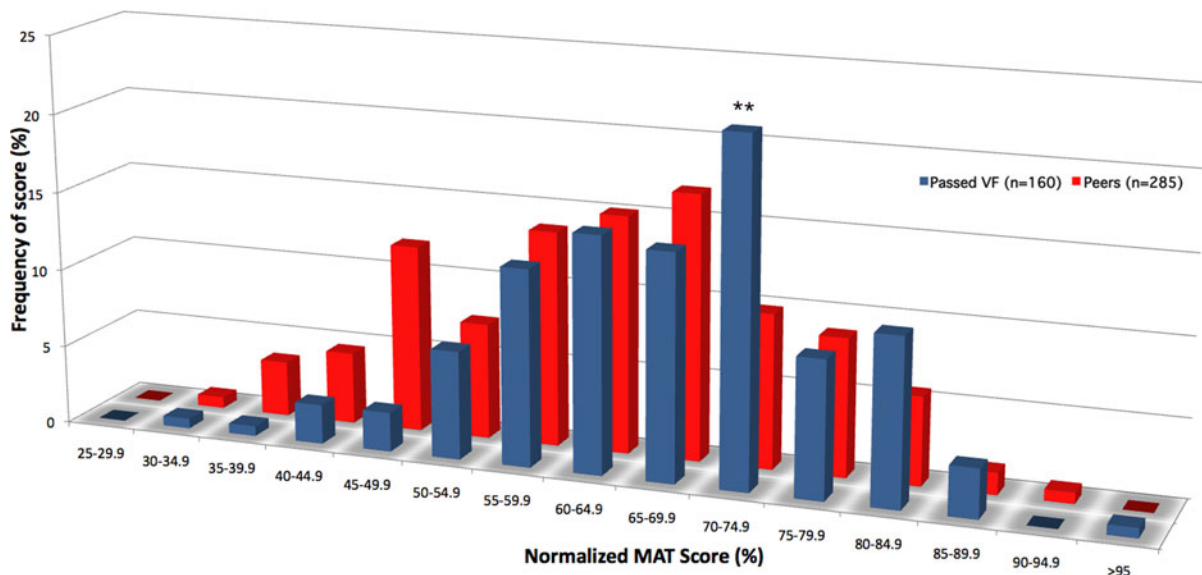
## RESULTS

### Students Who Passed the VF Scored Higher on MCAT Questions

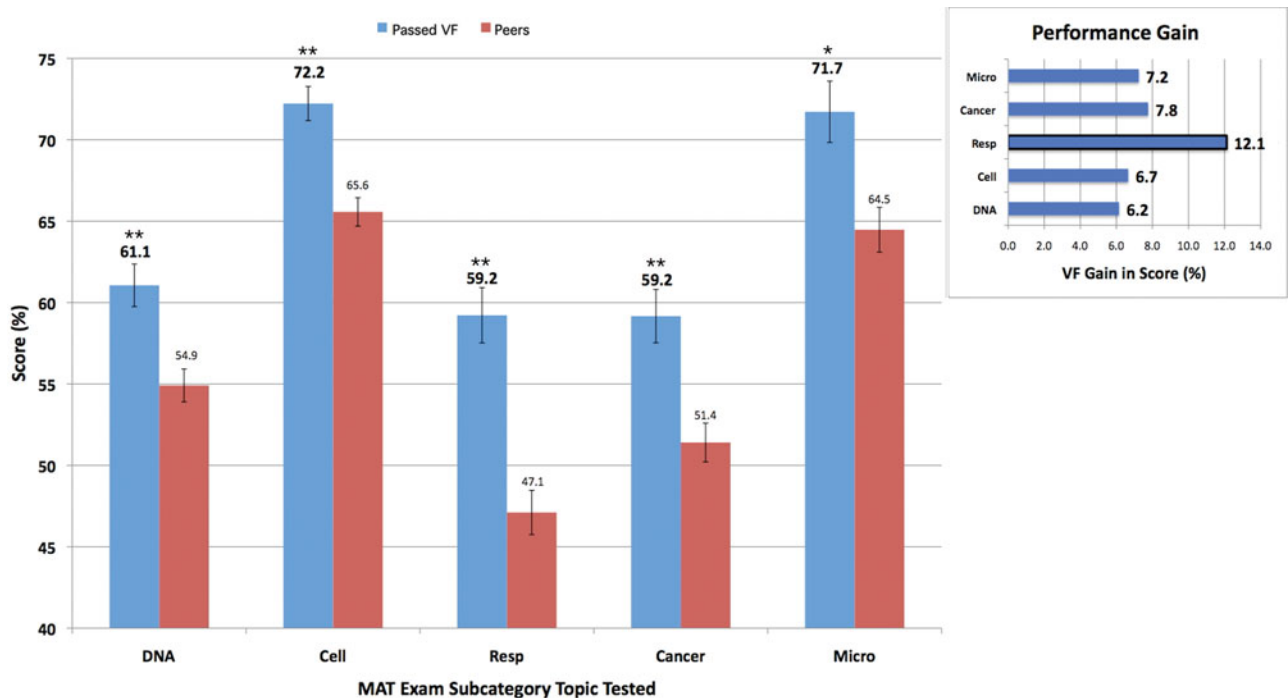
During the VF, students demonstrated their mastery of course material via a structured oral exam with the professor (Figure 1). Students who passed the VF performed better than their peers on MCAT questions included on an end-of-semester standardized MAT instrument (Figure 2; MAT scores were normalized with ACT performance). The average normalized MAT score for all students who participated and passed the VF was 66.4% ( $n = 160$ ), while the mean of peers who did not pass the VF was 62% ( $n = 285$ ;  $p < 0.001$ ; Figure 2, inset). In addition, when the student performance data were replotted in a frequency distribution, a rightward shift was apparent (Figure 3). The mode for performance of students who passed the VF was 72.5%, while those who did not pass had a mode of 65%. Higher levels of content mastery, for example, MAT scores of 80–90%, were more frequent among students who passed the VF.

### Students Who Passed the VF Scored Higher on Every Topic Area and Made Greatest Gains on Cellular Respiration

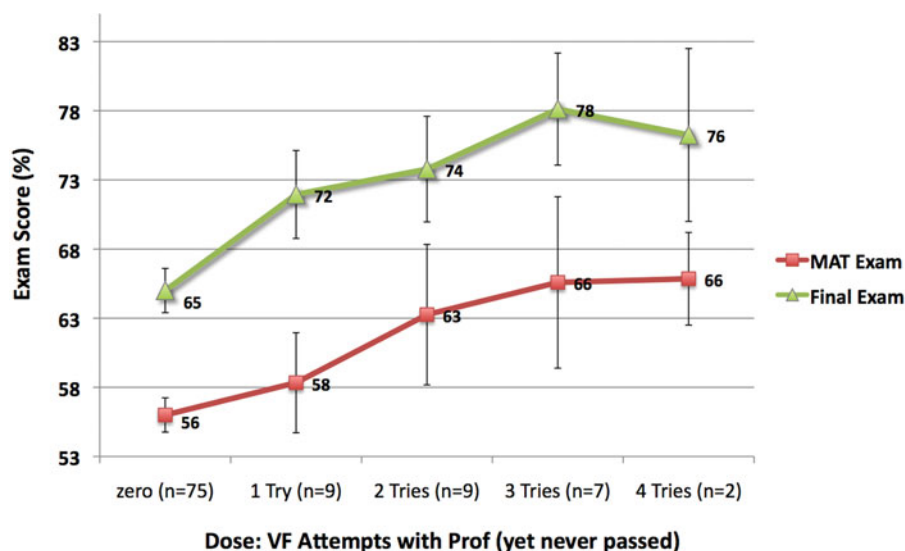
When performance on the MAT exam was examined, those who passed the VF did significantly better than peers on every topic (DNA, cell biology, respiration, cancer, and microbiology; Figure 4). The greatest gain in performance on the MAT



**Figure 3.** Students who passed the VF scored higher on the MAT exam than peers who did not. Normalized MAT performance distribution (score in percentage vs. frequency in percentage) is shown comparing cohort of students who passed the VF ( $n = 160$ ) with students who did not pass ( $n = 285$ ). No students scored below a 30%, while one student scored a 95%. Students who passed the VF are represented in blue. Students who did not pass the VF are represented in red. \*\*,  $p < 0.001$ ;  $t$  test compared pooled data of two cohorts.



**Figure 4.** Performance of students on each topic area tested by the MAT exam. The bars set to the left for each topic (DNA, Cell, Resp, etc.) correspond to the collective average score on that topic of all students who passed the VF ( $n = 160$ ). The bar to the right in each pair represents the average score for that topic of all students who did not pass the VF ( $n = 285$ ). Moving from left to right along the  $x$ -axis labels, the unabbreviated categories are: DNA structure and function, cell structure and function, cellular respiration, cancer, and microbiology topics, respectively. Error bars were generated using SEM. The inset illustrates the gain in learning of content knowledge obtained for each topic area tested by subtracting the lower score from the upper ( $t$  test: \*,  $p < 0.05$ ; \*\*,  $p < 0.001$ ).



**Figure 5.** Just attempting the VF may increase student learning. Students who made a greater number of attempts at the VF (but did not pass) may benefit and increase their learning. Students who did not pass the VF were required to take the traditional written final exam of the course. Thus, for this cohort, both the MAT and final exam scores were used to assess their content knowledge. The number of attempts made by students is compared with their exam performance. MAT scores are represented by the red line with square symbols; the green line with triangle symbols indicates final exam scores. SEM is indicated, and differences are not statistically significant.

was on questions related to the topic of cellular respiration (Figure 4, inset). The VF focuses upon the topics of DNA, cell biology, photosynthesis, and digestion. While cellular respiration is not discussed on the VF, the conceptually similar topic of photosynthesis is given the greatest amount of time and has the most in-depth treatment in the VF. This gain may be evidence of transfer of knowledge of photosynthesis to aid in solving problems on respiration.

### ***Just Attempting the VF Once May Increase Student Learning***

In only three semesters were records retained that indicated number of attempts for students who did not pass the VF. When examining these data, we see correlating increases in average performance on the MAT with number of VF attempts. While the trial numbers are quite low, the trend in the data supports a hypothesis that just participating in the process may lead to increased learning (Figure 5). Because these students participated but did not ultimately pass the VF, they were required to take the written final exam. So, for this cohort, the standardized final exam was a second assessment to compare with and complement the MAT. The performance of students had a near identical trend. With one exception, performance on both the written final exam and MAT instrument increased, suggesting a dosage effect is possible as students make more attempts to pass the VF. If one views each attempt as a formative rather than summative assessment, then the process would be expected to improve student understanding.

### ***Students Who Passed the VF Performed Better in Upper-Level Science Courses***

We tracked student performance in later years, when they took their upper-level science courses. Several rounds of analyses were performed. First, grades earned in upper-level science courses by the cohort of all students who passed the VF ( $n = 160$ ) were compared with those earned in the same courses by all other students from 1997–2010 ( $n = 21,528$ ). This

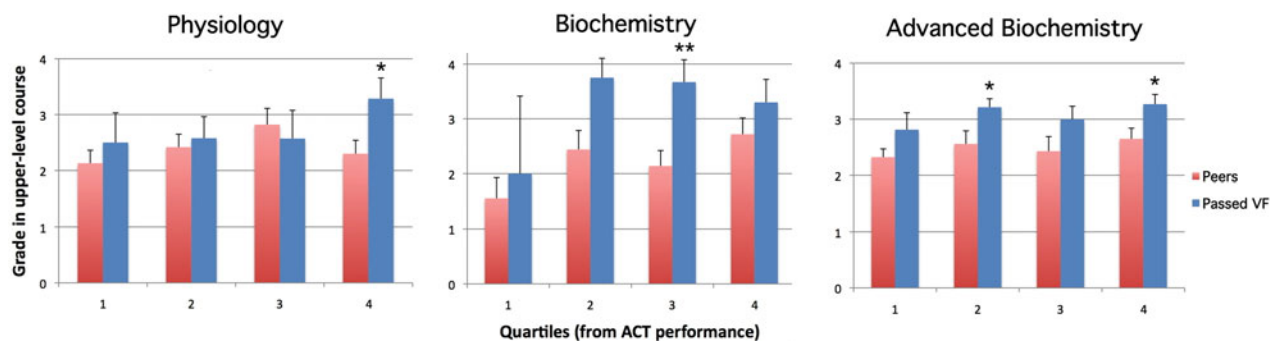
approach showed greater performance by the VF “passers” in all courses; however, the increase in incoming ACT scores over the time period may have conflated these results. Additional studies were done and final analysis was performed comparing two tightly matched cohorts: all students who passed the VF in the same course (Bio II) and other students who had taken Bio II the same semesters with the same instructor. This method again showed higher grades earned by VF passers than “nonpassers” (Table 1). This analysis, while promising, has the potential to be biasing the VF cohort toward students of higher academic potential, as the upper quartiles of students pass at higher rates. An expansive supplemental table lists courses studied in the analysis and side-by-side comparison of final GPA of the experimental and control cohorts in each quartile in each course (Table S1). Figure 6 illustrates a comparison quartile-by-quartile performance on course grade in biochemistry and physiology courses. Although many quartiles suffer from low  $n$  values, we note consistently higher grades earned by the VF passers across the board and significance in physiology and biochemistry courses.

Higher performance in later classes does not demonstrate causation, as it is possible that the results may reflect intrinsic differences in motivation, study skills, and so on. We used LM as a means to dissect this possibility and modeled student performance in the upper-level courses while normalizing for incoming university GPA and gender. In this LM approach, we found the VF to only show statistical significance for organic chemistry, CEM 351 (Figure S1). However, the low number of students in this path ( $n = 41$ ) suggests caution in interpreting these results. On the other hand, passing the VF results in a higher incoming GPA, thus masking its impact somewhat. Using pre-biology course predictors to estimate upper-level science grades resulted in weak models that explained only a small portion of the variance.

### ***Participation Balance from a Perspective of Academic Ability, Gender, and Ethnicity***

Overall, of the 445 students in the study, 160 (36%) passed the VF. In an attempt to tease apart participation from pass





**Figure 6.** A comparison of students who passed the VF vs. peers in upper-level science courses. Performance differences in upper-level science courses comparing students who passed VF with matched peer controls who were in the same major over the time period and took the LB145 course with same instructor. Average course GPA (trial numbers [n], pairs left to right) in PSL 431 (Physiology I:  $n = 27/5, 18/13, 14/7, 15/7$ ), BMB 401 (Basic Biochemistry:  $n = 9/3, 9/2, 7/3, 9/5$ ), and BMB 461 (Advanced Biochemistry I:  $n = 31/8, 17/14, 21/12, 20/13$ ) for VF passers and nonpassers by ACT quartiles. ANOVA: \*\*,  $p < 0.01$ ; \*,  $p < 0.05$ .

rates, we also evaluated three semesters in which records were retained that indicated students who participated in the VF but never passed. Of the 174 students in those courses, 70 (40%) passed the VF, 29 (17%) participated but did not pass, and 75 (43%) did not participate.

To evaluate participation balance from a perspective of academic ability, we stratified students in quartiles by ACT scores. We found the percentage of students who passed the VF was higher in the upper half of distribution, yet a good representation of students from all academic ability levels passed (Figure 7, top). In the three semesters with participation information, participation in each quartile ranged from 19.6 to 30.8%. Albeit from just a sample of the larger data set, these data indicate participation in the VF occurred from students in all levels (Figure 7, top, bottom graph). Students from all quartiles chose to participate, and the pass rates for the top three quartiles were statistically equivalent, not skewed to the top quartile. The full range of students, not just the academically elite, availed themselves of this optional assessment.

When evaluating gender, we found a reasonable balance. While there were no statistical differences overall between the rate of male and female students attempting or passing the VF (Figure 7, bottom), it was interesting that there was a difference between the rate of male and female students attempting the VF in the third quartile (although the rate of passing within this quartile was no different).

Participation balance from ethnic groups was also evaluated. We note that students who self-identified as African American and Asian had pass rates of 10 and 18% respectively, which while not statistically significant, trended below average. This finding gives pause. The pass rate of another ethnicity group, students who self-identified as “two or more races” was 60%, the highest of all cohorts. Given the error introduced by diametrical performance in overlapping categories and small sample size, we do not yet have enough data to make any statistically significant conclusions regarding potential cultural bias but have made this a focus of future study.

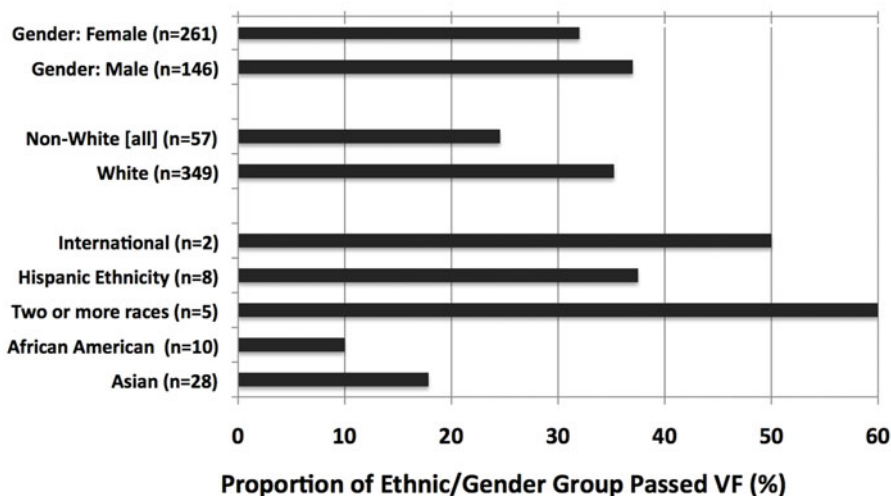
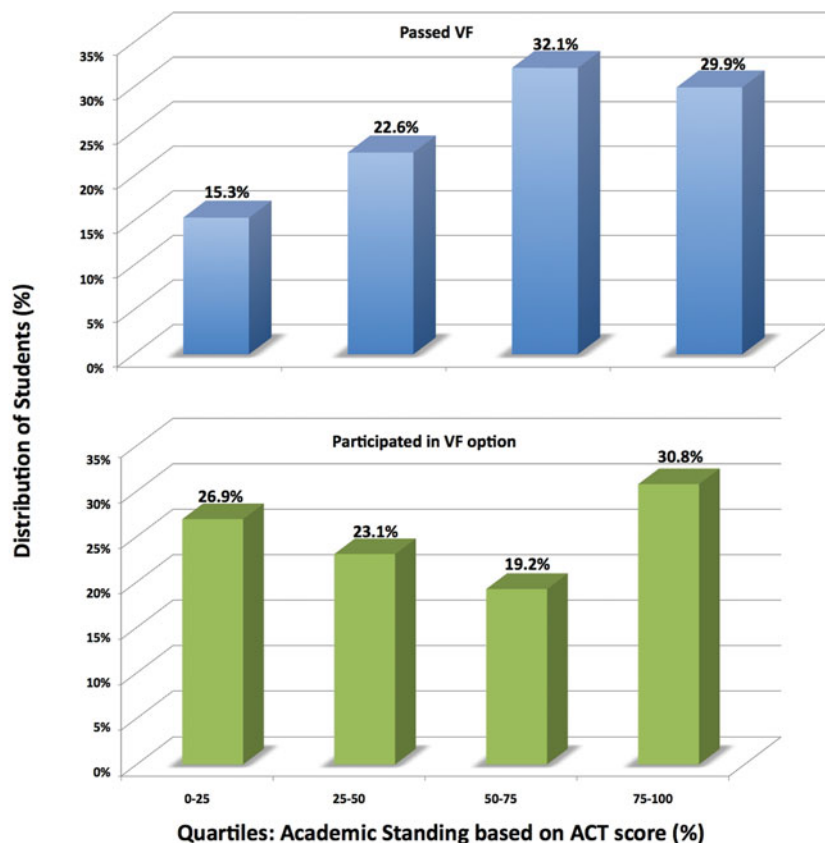
### ***Student Feedback Suggests That the VF Is Valued and Increases Study***

When a challenging intervention is added to a curriculum, the student response can be an important determinant of its

success. The student response has been nearly unanimous in its support for the VF. Starting in 2009, we added new prompts about the VF on the course evaluation forms. Student feedback in response to the question “Do you think the VF option helps student learning?” was unanimously positive (“great or much help”: 92%; “moderate or little help”: 8%; “no help”: 0%;  $n = 243$ ). We are analyzing student quotations as being supportive of predictions from the educational literature or alternative explanations for their learning gains. Representative exemplar comments are provided in Table S2 from both VF participants and abstainers; comprehensive student feedback is available online (Table S2; see VF website). Students’ informal feedback suggested that they recognized the VF had value, because it increased the amount of time they spent studying, changed the way they studied, and required them to verbalize their thinking. Further study is needed to determine the extent to which these variables can explain the gains students made.

## **DISCUSSION**

Over the past decade, the biology group in our residential college has adopted a number of best practices for teaching, among them inquiry laboratories, active and cooperative learning, and a more diverse set of assessments (Luckie *et al.*, 2004, 2011, 2012; Smith and Cheruvilil, 2009, Fata-Hartley, 2011). During this process, we also had opportunities to interview students and became more keenly aware of how little they were actually learning. Even when we spoke to our “best” students, those who did extremely well on traditional exams, we often found shocking gaps in their knowledge on the most fundamental topics. These realizations were a catalyst for change. They were also somewhat liberating, because they freed instructors from a belief that coverage equaled learning (DiCarlo, 2009; Schwartz *et al.*, 2009; Knight and Wood, 2005; Morgan, 2003; Wright and Klymkowsky, 2005). We decided to cover fewer topics but to dwell in each longer and have discussions that went deeper. We also identified the fundamental ideas or concepts that we were really teaching, or more accurately, that we really wanted the students to learn, underneath the conventional notion that students needed to know certain chapters or topics (Boyer, 1998; Ausubel, 2000).



**Figure 7.** Participation occurs from a broad range of students with a variety of academic and cultural backgrounds. Top, the ACT was used to divide the student population of each semester into academic performance-based academic standing quartiles. All students who passed the VF were counted for each quartile ( $n = 160$ ,  $z$  test of proportions at the  $p < 0.05$  level). Pass rates for quartile 1 and 2 are equivalent; quartiles 2, 3, and 4 are equivalent). Distribution of students who participated but did not pass is also represented. Bottom, the self-reported ethnicity and gender of students who passed the VF was also examined. Total number of students in study from each group is represented by  $n$ , percentage of that group who passed the VF is plotted (four students did not report ethnicity).

As a result, we reorganized our lecture sequence so we would be able to cover in the first half of the semester all the ideas critical for students to master in the course. We then could revisit those same concepts a second time by covering a different set of chapters or topics in the second half of the semester. Our goal was to give each student a “double dose” of the important ideas of biology. For example, for a learning objective such as “membranes work as barriers, and cells use transporters to move ions and create

gradients,” either the topic of photosynthesis or respiration can serve to frame the context in which we teach this core principle. A major concept in one topic, such as the role of the ATP synthase in photosynthesis, is revisited in cellular respiration, in which it has the same name and performs the same function. We hoped that reframing and revisiting the same idea or example in different contexts would help model for the students how an expert views the systems and perhaps would lead to a greater opportunity for

knowledge transfer, which is further deepened by the VF experience.

This reorganization of topics taught in our introductory biology course enabled this experiment with the VF to occur. Because we discussed all of the important concepts during the first half of the course, we could administer a VF (repeatedly) to all interested students throughout the entire second half of the semester. The topics discussed during the VF were framed in one story with three parts: how photosynthesis converts CO<sub>2</sub> into sugar/food, how humans digest and circulate sugar/food, and how a pancreatic cell makes insulin in response to the presence of sugar. The emphasis for greatest depth was focused on the beginning and end of the story, photosynthesis and central dogma, while the middle part was used as an affective bridge and opportunity to examine examples of cell biology in context.

### ***Student Content Knowledge Increased***

One goal in this research was to assess student learning and generate data that would be of the relevance and rigor most convincing to scientific colleagues. One important motivator for many of our students is to do well on the MCAT or Graduate Record Exam. They have a great desire to do well on placement exams and profess a desire to learn material in depth and to retain that knowledge beyond the semester. We have also found our science colleagues value student performance on those same exams, in particular the MCAT, far more than any other test we might find or create. As a result, in order to gauge student learning of content, throughout the study period, we examined student performance on a 40-question MAT exam built with MCAT test questions (Luckie *et al.*, 2004). Given that Zheng *et al.* (2008) demonstrated the MCAT assesses more than just lower-level knowledge and that it is a recognized exam for which performance is valued by our students and faculty peers, use of validated reliable passage questions derived from it seemed appropriate.

We found that students who engaged in the process and passed the VF scored the highest on end-of-semester MAT exams. These students significantly outperformed their peers who never participated in the process and those who made attempts but never passed the exam (passed VF: 66%; made attempt: 58–63%; no attempt: 57% score). Those who passed the VF appear to have more content knowledge. We also saw correlating increases in average performance on the MAT with number of VF attempts. With one exception, performance on both the course final exam and MAT instrument trend upward, suggesting a dosage effect is possible as students make more attempts to pass the VF. These data tend to support that the Socratic-styled VF is a useful assessment and learning aid.

Interestingly, students made significant gains in all topic areas tested, and the greatest gain in performance on the MAT was on questions related to the topic of cellular respiration. This topic is discussed in lecture but not during the VF. While cellular respiration is not discussed on the VF, the conceptually similar topic of photosynthesis is given the greatest amount of time and has the most in-depth treatment in the VF. We believe that this finding may be evidence of knowledge transfer, a goal in all teaching but a rather difficult task for students (McKeough *et al.*, 1995; Perkins, 1999; Schonborn and Bogeholz, 2009). Instructors always hope that, if a

student masters the fundamentals and learns some examples, he or she can use that understanding to make predictions and solve problems on related but different topics or tasks. In literature reporting testing of this idea, knowledge transfer is rarely noted to occur (McKeough *et al.*, 1995; Schonborn and Bogeholz, 2009). Students who passed the VF performed better on a broad range of topics, which suggests that they were better able to grasp the fundamentals and apply their knowledge more efficiently to novel situations; this is an indicator of higher-level learning.

We predicted that students who participated in the VF option might make gains in their understanding of biology through the experience, which is a form of cognitive apprenticeship (Brown *et al.*, 1989; Collins *et al.*, 1989). The VF clearly helped them in the course itself, but the idea that the experience may help students integrate and transfer knowledge, combined with long-term effects found by Derting and others, might predict greater success in upper-level classes (Derting and Ebert-May, 2010). We therefore did a longitudinal tracking study to follow how the students did in other science courses. Several rounds of analyses were performed that showed greater performance by the VF passers in a number of courses. An expansive supplemental table lists courses studied in the analysis and side-by-side comparison of final GPA of the experimental and control cohorts in each quartile in each course (Table S1). Passing the VF correlated with higher performance in later classes.

### ***Students Reported Gains But Not All Participate***

Evidence collected indicates that students support the VF as a valued option. Student comments also present opinions that parallel those from learning theory in the literature. Students believe preparation for the VF increases study time, engagement, revisiting concepts in different contexts, and making connections between ideas from different topics. This particular intervention is not controversial and has a positive effect on the affective elements of the classroom.

Analysis revealed that, while the percentage of students who passed the VF is higher in the upper half of distribution, a good representation of students from all academic standing levels passed. In addition, a good balance of students from all quartiles chose to participate in this alternative path. Regarding gender, there is evidence of near balance. The majority of students who passed the VF over the study period were female (61%), and while slightly lower, this percentage was not statistically different from their proportion of enrolled students (64%). When evaluating ethnic background, there is more concern. African-American and Asian students had below-average participation rates, and while the high pass rate of students who self-identified themselves as “two or more races” muddles the data, it seems likely stereotype threat is impacting participation of students from minority groups.

The existence of stereotype threat, speech impairments, test anxiety, and other barriers to success is a strong argument to keep the VF an optional path with no negative consequences (Steele and Aronson, 1995; Steele *et al.*, 2002; Spangler, 1997; Spencer *et al.*, 1999) but more needs to be done. We purposefully designed the process to lower test anxiety and give each student the best chance to engage his or her higher-level thinking without concern that failure would be

catastrophic for his or her grade. Rather than making an oral assessment mandatory in an attempt to increase learning for more students, we recognize its limitations, and our focus is to make it an event wherein demonstration of mastery only leads to rewards. New studies are in the works to test potential stereotype threat effects on participation. Using a private rather than public setting (e.g., closing classroom doors) may alter the participation and performance for female or African-American students. We also would like to understand how the presence of a female or nonmajority instructor administering the VF might improve participation.

### *Time on Task May Be the Mechanism*

Overt verbalization and previously discussed theoretical mechanisms likely contribute to the learning gains seen. Yet an additional mechanism leading to this improvement may simply be time on task (Peters, 2004; Guillaume and Khachikian, 2011), and student self-report feedback supports this rationale. If students spend more time on any topic or task, it is likely they will gain more mastery in it. Interestingly, some education psychology research suggests that when students believe they are pursuing knowledge mastery expertise as a goal, rather than just high performance on an assessment, they will spend far more time, persist longer, and use more elaborate study strategies to succeed (Nolen and Haladyna, 1990). This certainly could be true for the VF. The amount of time students spend preparing for the VF appears to be great. Student practice at available classroom whiteboards is evidenced by illustrations of photosynthesis found throughout the residential college each morning. Students report that they form study groups to practice together, ask students who have just passed the exam to be visiting experts in their study groups, and even recruit help from upper-level students who passed the VF in previous semesters. Teaching assistants also arrange practice VF appointments with interested students. Hence, we believe that time on task is a critical factor in student success.

Given that our goal is increased student learning, we view increasing the motivation of the students as a positive change. We are presenting the students with a different assessment approach that challenges and encourages them to spend more time preparing than they would otherwise. This activity is very public and creates an engaged culture focused on academics and learning that benefits all parties involved (DiCarlo, 2009). Logistical constraints in allowing students to retake exams make it difficult with most tests (Henderson and Dancy, 2007), but the VF is conducive to this philosophy.

### *Limited Faculty Time Leads to Solutions*

While our colleagues may agree it is debatable whether traditional final exams are good methods to evaluate learning (John Harvard's Journal, 2010) and alternative assessments may be worth considering, it is difficult for faculty members to make changes in their teaching practices due to many factors, including limited time (Silverthorn *et al.*, 2006; Henderson and Dancy, 2007; Ebert-May *et al.*, 2011). In the interests of efficiency, instructors implementing the VF have tested a number of different models that, when combined, may offer solutions. During the first few weeks of implementing the VF, the instructor administers exams on only

one day of the week. Limitations are also set during the final weeks; for example, no new students can start the process after a certain deadline, and no more VF appointments are permitted after a certain final date (several days prior the written final exam). These strategies can limit the time the faculty member's time commitment to 5–10 h per week.

Additional time-saving approaches have included one recently used by R.D.S. in his chemistry course, in which each student only gets one chance to take the VF with the professor (Sweeder and Jeffery, 2012). Also, one published approach focuses on the learning benefits of verbal exams administered to students as a group (Boe, 1996; Guest and Murphy, 2000). Either of these alternative approaches alone or in combination could dramatically decrease the time commitment from a faculty member. Alternatively, one could test whether many of the benefits of the VF could still be retained in a nonverbal format. One expert suggested that the kinds of questions asked during the oral exam are quite sophisticated, but they could be asked on a written exam that gave students the opportunity to revise their work continuously until their work was correct. For some faculty members, the written exam idea may well be more desirable and efficient in terms of faculty time.

One very promising and effective approach has been the use of teaching assistants. In recent semesters, we required that students pass with a teaching assistant prior to getting an appointment with the professor. Our teaching assistants, called learning assistants (LAs), are undergraduates who participated in the VF process when students. Because they were outstanding students who demonstrated great interest in teaching others the biology topics they were passionate about, they were selected to be LAs. We have found LAs can be trained to be rigorous in their administration of the VF and can carry the lion's share of the load. They become just as invested in the process as when they were students (and the success of their students is a point of pride). This assistance by LAs enables faculty time to be much more protected, because well-prepared students can pass with the faculty on the first attempt. In addition to helping protect limited faculty time, this approach may provide two other benefits.

The participation by undergraduate teaching assistants in the VF process may increase participation from reluctant (and minority) students and may yield learning gains for the undergraduate LA students themselves. Our LAs are far more approachable and diverse than the typical professor. More than half are female, and a significant number are of nonmajority ethnicity. They can serve as a less threatening starting point for reluctant students and as learning coaches in the shared goal that students master the material and pass the test. Our LAs have also shared with us that participating greatly solidifies their own understanding of the material and connects it to other courses. Here are some of their reflections regarding their own learning:

"I learned a lot from the students. They do so much extra studying which causes them to learn much more than what I will ask them . . . they tell me things that I did not know."

"I believe that the verbal final exam experience has greatly increased my learning as an LA who administered the exams. Students will inevitably discover, and bring to the LA's attention, information that neither

person knew before. Every year, I become a better verbal final exam proctor simply because my students are teaching me biology.”

“I believe administering the VF broadened my understanding of the material. Questioning and evaluation presents another facet to the learning process; additionally, discourse with students often challenge[s] our understanding and require[s] constant de- and reconstruction.”

“Giving verbal finals has in some ways increased my learning. I think the biggest thing is just going over the material that I learned in 145 again, but also seeing how I’ve learned more about certain topics in other classes like PSL and BMB. Students ask a lot of questions and sometimes make you think about/look up things as well.”

“I certainly had information re-introduced that I had forgotten since last taking LB 145. Everything came back quickly after some reviewing (probably because it was so ingrained the first time), but going over the verbal final material once more solidify [sic] certain concepts once again. Just after reviewing them for LB 145, these systems were discussed in my comparative anatomy course.”

### Limitations of the Study

As one considers the implications of this work, it is important to recognize some of the limitations. The course itself was offered to students who were in a science-themed residential college program, so all of the enrolled students were science majors. Although there is no reason to suspect that nonmajors would be any less capable of succeeding on a VF, the participation rate in a general biology class may be anticipated to decrease, as many students may not be as interested in gaining a deeper level of understanding of biology if they do not see an apparent benefit to their long-term career plans.

While ACT normalization helps equalize for motivation and background knowledge in the data presented in Figures 2–4, the higher level of performance of students in upper-level courses in Figure 6 should be interpreted cautiously. The statistically higher grades earned by the VF passers in all quartiles indicate only a correlation, not a causation. Taking the VF may more accurately represent a high level of motivation, dedication, or other characteristic that also leads to success in the upper-level science courses. Except in one case, CEM351, the more powerful statistical approach of using linear modeling did not show the VF as being a crucial indicator of success; however, it relied on the students’ college GPA entering each course. This GPA would already reflect learning gains and concomitant inflation as a result of passing the VF (through the elevated biology grade earned and other classes it may have impacted); thus some evidence of impact would have been removed.

### CONCLUSIONS

An optional VF has provided an alternate venue for assessing student understanding in our introductory biology course. As one might predict, in a quantitative assessment of cell biology content knowledge, students who passed the VF outscored their peers on related MCAT exam questions. The higher-

achieving students performed better on MCAT exam questions from all topic categories tested. Success on the VF also correlated with enhanced success in multiple, subsequent, upper-level science courses.

While we provide detailed *Methods* to enable other instructors to understand the process we used in our course, these will not be very helpful to faculty members who do not teach a similar course. Perhaps more helpful are just the principles that the VF offers. We believe some important elements are these: 1) having students verbalize processes as they illustrate them; 2) allowing students to first explain a process and, when ready, to accept questions; 3) focusing on just a few topics that contain the most important fundamental concepts; 4) enabling students to understand they can repeat attempts and ultimately take the written final exam no matter whether they pass or not; and 5) using a transparent protocol with a clear rationale for when something is incorrect and giving opportunities to revise.

The evidence evaluated in this study supports that Socratic-styled oral exams can allow instructors to assess and aid students striving to achieve higher-level learning. The intervention is a significant one, but student feedback demonstrates it has no negative effect on the affective atmosphere of the classroom. In fact, the opposite is true; students who participated in the option reported it increased their study and learning, and those who chose to not participate remained supportive of the option.

While the affective (happiness) element of students is often carefully tracked in teaching, another rarely tracked element important to the success of student learning is the affective state of the instructor. First and foremost, if you ask your students questions about the topics taught in the class, you will get a much better idea of what they learned and their level of mastery. But perhaps equally important, you will also have enjoyable experiences listening to your students thoughtfully and passionately explain to you the processes taught in your course. You will get opportunities to correct, clarify, and raise each student’s understanding of the material. And you will find great satisfaction when students show increased mastery about the topics so important to their understanding of the discipline.

We hope this publication and other qualitative and quantitative studies (Dressel, 1991; Bairan and Farnsworth, 1997; Schubert *et al.*, 1999; Mills *et al.*, 2000; Ehrlich, 2007; Elfes, 2007) help provide evidence and inspiration to our peers in science on the benefits of moving deliberately to challenge students with active, engaging teaching methods and, in particular, to use more higher-level assessments in the classroom.

### ACKNOWLEDGMENTS

We thank Amanda (Gnau) Seguin for the creative idea that enabled the study presented in this article. We thank Drs. Joseph Maleszewski, John Wilterding, James Smith, Cori Fata-Hartley, Diane Ebert-May, Janet Batzli, John Merrill, Merle Heideman, Karl Smith, Mimi Sayed, Sarah Loznak, and Duncan Sibley for helpful discussions about teaching and learning and assistance during this study. We also thank the many LAs who promoted and supported these interventions. Our cystic fibrosis research laboratory is supported by Cystic Fibrosis Foundation and Pennsylvania Cystic Fibrosis grants for our physiology cystic fibrosis research program; our STEM Learning laboratory is supported by Lilly and Quality grants from MSU and

Course Curriculum and Laboratory Improvement and Transforming Undergraduate Education in Science grants from the National Science Foundation for our education research program.

## REFERENCES

- Ausubel D. (2000). *The Acquisition and Retention of Knowledge: A Cognitive View*, Boston, MA: Kluwer.
- Baddeley AD, Hitch G (1974). Working memory. In: *The Psychology of Learning and Motivation: Advances in Research and Theory*, ed. GH Bower, New York: Academic, 8, 47–89.
- Bairan A, Farnsworth B (1997). Oral exams: an alternative evaluation method. *Nurse Educ* 22, 6–7.
- Bloom BS (ed.) (1956). *Taxonomy of Educational Objectives: Book 1, Cognitive Domain*, New York: Longman.
- Boe BL (1996). Diversity in assessment procedures. Paper presented at the 48th American Association of Colleges for Teacher Education Conference, held 21–24 February 1996, in Chicago, IL.
- Boyer Commission on Education of Undergraduates in the Research University (1998). *Reinventing Undergraduate Education: A Blueprint for America's Research Universities*, Washington, DC: Carnegie Foundation for the Advancement of Teaching.
- Brown JS, Collins A, Duguid P (1989). Situated cognition and the culture of learning. *Educ Res* 18, 32–42.
- Bruner JS (1973). *The Relevance of Education*, New York: Norton.
- Carmean SL, Wier MW (1967). Effects of verbalization on discrimination learning and retention. *J Verbal Learn Verbal Behav* 6, 545–550.
- Collins A, Brown JS, Newman SE (1989). Cognitive apprenticeship: teaching the crafts of reading, writing, and mathematics. In: *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, ed. LB Resnick, Hillsdale, NJ: Lawrence Erlbaum, 453–494.
- Creech LR, Sweeder RD (2012). Gender performance differences in life science courses. *CBE Life Sci Educ* 11, 396–391.
- Derting TL, Ebert-May D (2010). Learner-centered inquiry in undergraduate biology: positive relationships with long-term student achievement. *CBE Life Sci Educ* 9, 462–472.
- DiCarlo SE (2009). Too much content, not enough thinking, and too little fun! *Adv Physiol Educ* 33, 257–264.
- Dressel JH 1991. The formal oral group exam: challenges and possibilities—the oral exam and critical thinking. Paper presented at the 81st Annual Meeting of the National Council of Teachers of English, held 22–27 November 1991, in Seattle, WA.
- Ebert-May D, Derting TL, Hodder J, Momsen JL, Long TM, Jardeleza SE (2011). What we say is not what we do: effective evaluation of faculty professional development programs. *BioScience* 61, 550–558.
- Ehrlich R (2007). Giving bonus points based on oral exams. *Am J Phys* 75, 374.
- Elfes C (2007). Oral exam tests your decision-making skills. *Practitioner* 251, 44–45.
- Entwistle A, Entwistle N (1992). Experiences of understanding in revising for degree examinations. *Learn Instruct* 2, 1–22.
- Fata-Hartley C (2011). Resisting rote: the importance of active learning for all course learning objectives. *J Coll Sci Teach* 40, 36–39.
- Gagne RM, Smith EC (1962). A study of effects of verbalization on problem solving. *J Exp Psychol* 63, 12–18.
- Guest K, Murphy D (2000). In support of memory retention: a cooperative oral final exam. *Education* 121, 350.
- Guillaume DW, Khachikian CS (2011). The effect of time-on-task on student grades and grade expectations. *Assess Eval High Educ* 36, 251–261.
- John Harvard's Journal (2010). Bye-bye, blue books? *Harvard Magazine Jul–Aug*, 64–65. <http://harvardmagazine.com/2010/07/bye-bye-blue-books> (accessed 15 July 2013).
- Henderson C, Dancy MH (2007). Barriers to the use of research-based instructional strategies: the influence of both individual and situational characteristics. *Phys Rev ST Phys Educ Res* 3, 020102.
- Inhelder B, Piaget J (1958). *The Growth of Logical Thinking from Childhood to Adolescence*, New York: Basic.
- Knight JK, Wood WB (2005). Teaching more by lecturing less. *Cell Biol Educ* 4, 298–310.
- Krontiris-Litowitz J (2009). Articulating scientific reasoning improves student learning in an undergraduate anatomy and physiology course. *CBE Life Sci Educ* 8, 309–315.
- Labuhn AS, Zimmerman BJ, Hasselhorn M (2010). Enhancing students' self-regulation and mathematics performance: the influence of feedback and self-evaluative standards. *Metacogn Learn* 5, 173–194.
- Luckie D, Harrison SH, Ebert-May D (2011). Model-based reasoning: using visual tools to reveal student learning. *Adv Physiol Educ* 35, 59–67.
- Luckie DB, Aubry JR, Marengo BJ, Rivkin AM, Foss LA, Maleszewski JJ (2012). Less teaching, more learning: 10-yr study supports increasing student learning through less coverage and more inquiry. *Adv Physiol Educ* 36, 325–335.
- Luckie DB, Maleszewski JJ, Loznak SD, Krha M (2004). Infusion of collaborative inquiry throughout a biology curriculum increases student learning: a four-year study of "Teams and Streams." *Adv Physiol Educ* 28, 199–209.
- McKeough A, Lupart JL, Marini A (1995). *Teaching for Transfer: Fostering Generalization in Learning*, New York: Psychology Press.
- Mills PA, Sweeny WV, DeMeo S, Marino R, Clarkson S (2000). Using poster sessions as an alternative to written examination—the poster exam. *J Chem Educ* 77, 1158–1161.
- 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, 435–440.
- Morgan E (2003). *BIO2010: Transforming Undergraduate Education for Future Research Biologists*, Washington, DC: National Academy. [www.nap.edu/openbook.php?isbn=0309085357](http://www.nap.edu/openbook.php?isbn=0309085357) (accessed 15 July 2013).
- Nolen SB, Haladyna T (1990). Personal and environmental influences on students' beliefs about effective study strategies. *Contemp Educ Psychol* 15, 116–130.
- Paul R, Elder L (2006). *The Art of Socratic Questioning*, Dillon Beach, CA: Foundation for Critical Thinking.
- Perkins D (1999). The many faces of constructivism. *Educ Leadership* 57, 6–11.
- Peters E (2004). Maximize student time on task. *Science Scope* 28, 38–39.
- Pintrich PR (1998). A process-oriented view of student motivation and cognition. In: *Improving Teaching and Learning through Research: New Directions for Institutional Research*, ed. J Staker and L Mets, San Francisco: Jossey Bass, 57, 65–79.
- Posner GJ, Strike KA, Hewson PW, Gertzog WA (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Sci Educ* 66, 211–227.
- Rauschenberger MM, Sweeder RD (2010). Gender performance differences in biochemistry. *Biochem Mol Bio Ed* 38, 380–384.
- Schonborn KJ, Bogeholz S (2009). Knowledge transfer in biology and translation across external representations: experts' views and challenges for learning. *Int J Sci Math Educ* 7, 931–955.

- Schubert A, Tetzlaff JE, Tan M, Ryckman JV, Mascha E (1999). Consistency, inter-rater reliability, and validity of 441 consecutive mock oral examinations in anesthesiology: implications for use as a tool for assessment of residents. *Anesthesiology* 91, 288–298.
- Schunk DH (1986). Verbalization and children's self-regulated learning. *Contemp Educ Psychol* 11, 347–369.
- Schwartz MS, Sadler PM, Sonnert G, Tai RH (2009). Depth versus breadth: how content coverage in high school science courses relates to later success in college science coursework. *Sci Educ* 93, 798–826.
- Seymour E, Wiese D, Hunter A, Daffinrud SM (2000). Creating a better mousetrap: on-line student assessment of their learning gains. Paper presented at the National Meeting of the American Chemical Society, March 27, 2000, in San Francisco, CA.
- Silverthorn DU, Thorn PM, Svinicki MD (2006). It's difficult to change the way we teach: lessons from the Integrative Themes in Physiology curriculum module project. *Adv Physiol Educ* 30, 204–214.
- Smith J, Cheruvilil K (2009). Using inquiry and tree-thinking to "March through the animal phyla": teaching introductory comparative biology in an evolutionary context. *Evol Educ Outreach* 2, 429–444.
- Spangler G (1997). Psychological and physiological responses during an exam and their relation to personality characteristics. *Psychoneuroendocrinology* 22, 423–441.
- Spencer S, Steele CM, Quinn D (1999). Stereotype threat and women's math performance. *J Exp Soc Psych* 35, 4–28.
- Spicer J (2005). *Making Sense of Multivariate Data Analysis*, Thousand Oaks, CA: Sage.
- Steele CM, Aronson J (1995). Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol* 69, 797–811.
- Steele CM, Spencer SJ, Aronson J (2002). Contending with group image: the psychology of stereotype and social identity threat. *Adv Exp Soc Psych* 34, 379–440.
- Stumpf SE (1983). *Philosophy: History and Problems*, New York: McGraw-Hill.
- Sweeder RD, Jeffery KA (2012). A comprehensive general chemistry demonstration. *J Chem Educ* 90, 96–98.
- Sweeder RD, Jeffery KA, McCright AM (2012). Lyman Briggs College: an innovative living-learning community for STEM education. *Qual Approach High Educ* 3, 7–14.
- Thorburn M, Collins D (2006). Accuracy and authenticity of oral and written assessments in high-stakes school examinations. *Curriculum J* 17, 3–25.
- Tobias S (1992). *Revitalizing Undergraduate Science: Why Some Things Work and Most Don't*, Tucson, AZ: Research Corporation.
- Tobias S, Raphael J (1997). *The Hidden Curriculum—Faculty-Made Tests in Science, Part 1: Lower-Division Courses*, New York: Plenum.
- Traub RE, MacRury K (1990). Multiple-choice vs. free response in the testing of scholastic achievement. In: *Test und tends 8: Jahrbuch der pädagogischen diagnostik*, ed. K Ingenkamp and RS Jager, Weinheim, Germany: Beltz Verlag, 128–159.
- Vygotsky LS (1962). *Thought and Language*, Cambridge, MA: MIT Press.
- Wilder L, Harvey DJ (1971). Overt and covert verbalization in problem solving. *Commun Monogr* 38, 171–176.
- Wright RL, Klymkowsky MW (2005). Points of view: content versus process: is this a fair choice?: undergraduate biology courses for nonscientists: toward a lived curriculum. *Cell Biol Educ* 4, 189–196.
- Zheng AY, Lawhorn JK, Lumley T, Freeman S (2008). Assessment. Application of Bloom's taxonomy debunks the "MCAT myth." *Science* 319, 414–415.