

# Resources for Assessing Educational Interventions in Biology at the Collegiate Level

Youngeun Choi<sup>†</sup>, Genevieve C. Saphier<sup>†</sup>, and William J. Anderson\*

Harvard University, Department of Stem Cell and Regenerative Biology, Cambridge, MA, 02138

<sup>†</sup>These authors contributed equally to this work

## Abstract

Most scientific research is judged based on the quality of controlled experiments and carefully analyzed results. In addition, proper levels of regulation in terms of biosafety and animal usage are a routine part of the scientific research process for laboratories. For many biologists, educational research is much more of a black box. While faculty have many great ideas on how to improve education, they struggle with the best way to evaluate whether their ideas lead to greater student outcomes. Here we provide a review of compliance issues related to educational research, as well as describe ways in which pedagogical innovations in biology can be assessed. We also describe some of the challenges related to educational research and how these could be addressed.

**Citation:** Choi, Y., Saphier, G.C., and Anderson, W.J. 2016. Resources for Assessing Educational Interventions in Biology at the Collegiate Level. *CourseSource*. <https://doi.org/10.24918/cs.2016.1>

**Editor:** Robin Wright, University of Minnesota, Minneapolis MN

**Received:** 5/4/2016; **Accepted:** 6/20/2016; **Published:** 9/19/2016

**Copyright:** © 2016 Choi, Saphier, and Anderson. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Conflict of Interest and Funding Statement:** None of the authors have a financial, personal, or professional conflict of interest related to this work.

\*Correspondence to: 7 Divinity Ave, Bauer 204, Cambridge, MA 02138.

E-mail: [william\\_anderson@harvard.edu](mailto:william_anderson@harvard.edu)

## Compliance and Assessment are Critical for Successful Innovation in STEM Education

A series of national reports and commentaries calling for science, technology, engineering, and mathematics (STEM) education reform has gradually brought about changes in classroom environment and teaching methods (1-5). Terms such as “active learning,” “flipped classroom,” and “clicker questions” are no longer obscure to many science educators (6). Prestigious science journals have published science education research papers, and scientific communities have organized workshops where instructors can learn and discuss effective teaching strategies (1,7-9).

Equipped with growing support and resources, science educators are encouraged to implement and assess novel teaching methods more than ever. However, many faculty members in higher education are apprehensive about performing any systematic evaluation of their pedagogical strategies (10) beyond conventional student course evaluations. For instance, it is not hard to find education research presentations at conferences that lack formal assessment. In many cases, researchers use affective assessments on how students view the intervention on their own outcome without using a more objective quantification, such as normalized learning gain. Sometimes, the assessment tools are inconsistent

(and sometimes inappropriate) based upon the question asked. Furthermore, many of these studies are presented without obtaining IRB determination or approval prior to engaging in the research.

To foster and disseminate research-based pedagogical strategies in STEM education, it is necessary to provide instructors, who were trained as scientists and thus are generally unaccustomed to the setting of educational research, with the tools necessary to assess their classroom innovations. To aid science educators in entering possible uncharted territory surrounding educational research, we address in this essay two issues that instructors should consider while planning for educational research: regulatory compliance and assessment tools for pedagogical interventions.

## What Regulatory Compliance Issues Arise from Education Research?

### *Family Education Rights and Privacy Act*

In general, the compliance issues faced by faculty looking to conduct educational research are regulated by the Family Education Rights and Privacy Act (FERPA) and by the Department of Health and Human Services regulations for the Protection of Human Subjects (45CFR46), colloquially referred to as “the

Common Rule.” Understanding of and compliance with these rules and regulations are important for the responsible conduct of education research in the higher education setting.

FERPA is a federal law intended to protect the privacy of parents and students. It regulates the disclosure of personally identifiable information from the student record (e.g., grades, GPA, etc.). Certain educational research projects may wish to use some of this information (e.g., to correlate learning gains in a class with students’ performance in other biology classes or their science GPA). The use in research of information from student records is permitted by FERPA under the following circumstances:

1. The student has explicitly consented to the use of this information in the research, OR
2. The information has been de-identified, OR
3. The information will be used to either
  - Develop, validate, or administer predictive tests, OR
  - Improve instruction
 AND
  - The researcher enters into a data use agreement that includes specific language regarding data confidentiality (34CFR99.31(a)(6)) (11).

The aforementioned data use agreements vary by institution, as they typically include elements required by FERPA as well as institution-specific requirements. FERPA requirements for usage can be found here: <http://ptac.ed.gov/Guidance-Reasonable-Methods-Written-Agreements>. Consultation with the registrar’s office should help faculty to navigate specific institutional policy and procedures regarding compliance with FERPA.

### *Institutional Review Board on Human Subjects Research*

Educational research in the context of the Common Rule, which requires Institutional Review Board (IRB) review and approval of most human subjects research, is more nuanced and therefore often quite difficult to figure out. The Common Rule generally only applies to either federally funded research or research at institutions that accept federal funding, the latter of which applies to many academic institutions. In general, to determine if a particular project is subject to the Common Rule, the Office of Human Research Protection (OHRP) provides the guidance that one should begin by asking, “Is it research?” (12) In the context of educational research, this can be particularly difficult to decipher, as Hammack (1997) points out, “Good teaching practice has always required close observation and experimentation” (13). While initially in the context of medical education, Roberts et al. (2001) addresses something relevant to all teacher-led education research: “Taken together medical education research and medical education practice both involve being methodical, innovative, self-observing, forward-looking, and open to peer review” (14).

So when does good teaching practice cross over into educational research? In the regulation, “research” is defined as, “a systematic investigation designed to develop or contribute to generalizable knowledge.” Discussions of this definition generally focus on the words systematic, designed, and generalizable knowledge (15). Much of good teaching and even just academic record keeping may include the systematic

collection of data. But if the design of the project, commonly understood as “intent” (15) is not generalizable knowledge, but instead program evaluation or quality improvement, then the project is not research (OHRP; (16)). However, if part of the intent of the project is quality improvement and part is research, or if the intent of the project changes along the way from quality improvement to research, then human subjects protection applies and review is necessary (17). It is critical to note then that the distinction is not made based on what will be done by the teacher or researcher, but rather based on what the teacher or researcher intends (18).

Researchers should check with their local IRB to determine their exact requirements. Many educational research projects will fit into the regulatory category of “exempt.” Though exempt research does not technically have to undergo IRB review (15), OHRP recommends that investigators not make this determination on their own (OHRP; (12)), and many institutions have policies requiring IRB determination of exemption. The Common Rule lists six categories of research that are exempt from IRB review. The most germane rule states that if research is conducted in a typical education setting and involves only normal educational practices, then it is exempt from IRB review. Projects that have been determined to be exempt do not generally require annual renewal by the IRB. However, researchers are often required to submit any modifications to the study’s design or protocol to the IRB for review so that the IRB can determine that the modification does not alter the exempt status.

It is relevant to note that the US Department of Health and Human Services issued a Notice of Proposed Rulemaking (NPRM) in September 2015 that proposes substantial changes to the Common Rule. The proposal includes changes to how determinations of exempt research can be made and creating a new category of research called “excluded” that would not require any IRB review. As of now, it is unclear who would determine whether or not a particular project qualifies for exclusion. The comment period on these proposed changes ended in January 2016, and it is not clear what the final rules will be, when they will be published, and/or when they will take effect.

### **Methods to Assess the Effectiveness of the Pedagogical Intervention**

In addition to addressing compliance issues, educational researchers should select appropriate and effective methods to evaluate their pedagogical intervention. Depending on the goal of the pedagogical intervention, educational research may address one or more of the following questions:

- How much did the intervention contribute to student learning?
- How much did a particular skill/competency improve due to the intervention?
- How much did student attitudes change after the intervention?

Below, we first briefly discuss some unique attributes of the educational research that should be considered in selecting the pedagogical tool. We then categorize assessment tools that can be used to determine the impact of the pedagogical strategy on learning, development of critical thinking skills, or student attitudes.

### *Quasi-Experimental Research*

Trained as scientists, many biology educators may feel uncomfortable with quasi-experimental design, which is widely used in educational research. Two conditions that set apart education research from scientific research are the lack of identical study subjects (i.e., no two students are the same) and an ethical concern over students in a control group who cannot benefit from the educational innovation. To address the first issue, researchers have used analysis of covariance (ANCOVA) to control for strong indicators of student performance, such as grade point averages (19). Although diversity of student background needs to be controlled, it can inform researchers whether a group of students with a specific background benefits from the tool more than other groups. A recent study addressing the effect of increased course structure revealed the intervention disproportionately influenced particular subpopulations of students (20). This finding suggests that organizing student data by their backgrounds may uncover otherwise hidden or overlooked effects of the pedagogical technique.

The equity issue is complicated by the necessity of having a control group in the research. To circumvent this concern, researchers can perform their studies over the course of several years, during which the intervention is introduced, and use test scores from the year(s) before the pedagogical change as a control (19,8,10,21). In this case, the instructor should ensure that the difficulties of assessment tools stay the same throughout the years of the study. Alternatively, instructors can minimize the degree of inequity by forming a control group and an experimental group temporarily (e.g., one week) during the term, implementing the pedagogical intervention, assessing and comparing student performance (22,23). If different student groups are taught by different instructors, researchers should try to make the teaching quality and instructor background consistent between groups. A semester-long, randomized trial of a teaching technique is also feasible with student understanding of the study and their consent to participation, and it enables the instructor to rigorously examine the causal effect of the intervention (24,25,9).

### *Formal Learning Outcome Assessment*

#### **Content Specific Assessment**

If the intervention was implemented to enhance content knowledge and understanding, the instructor can measure how much students learned by administering the same test before and after the intervention (i.e., pre- and post-test). The difference between the pre-test score and the post-test score can be an intuitive measurement of the learning gain, but this simple calculation has some caveats (26,27). For instance, students with high pre-test scores have only small room for improvement, thus bigger differences between pre- and post-test scores are correlated with lower pre-test scores (28). Therefore, the education community has used the normalized learning gain instead (29):

Normalized Learning Gain = (Post Score - Pre Score) / (100 - Pre Score)

This in turn eliminates any bias caused by different pre-treatment scores (27). Moreover, researchers have developed other analysis methods [e.g., Rasch modeling (27,30)] to

extract additional information from the assessment data.

If student answers should be qualitatively evaluated (e.g., essay questions), experts in the field who are not involved in the course can be appropriate graders (31). In this case, it is important that these graders should be provided with a comprehensive rubric so that any discrepancies between graders can be minimized.

#### **Concept Inventory**

Concept inventory is a multiple-choice test validated to assess student understanding of core concepts in a specific subject (32). Inspired by the Force Concept Inventory, which was a crucial tool to evaluate different teaching methods in physics education (29,33-35), biology faculty and educators have collaborated to develop various biological concept inventories (reviewed in (32); Table 1 on page 4). These inventories are publicly available, making researchers less burdened with devising their own questions. Another advantage of using the concept inventory is that its widespread use can benefit the education community beyond the individual study by enabling the comparison between different pedagogical techniques.

#### **Self-Reported Assessment**

Students can self-report their learning gains using an instrument such as the Student Assessment of Learning Gains survey (Table 1). Although previous studies found a correlation between self-reported gains and student academic performance (36-38), the accuracy of self-reported learning gain is still debatable (39,40). Thus, we suggest that educators use one of the aforementioned formal assessments to measure the efficacy of the teaching strategy.

### *Critical Thinking/Scientific Thinking Assessment*

Improving student analytical skills and scientific literacy is one of the core objectives of undergraduate science education reform (1,41-47). As a response, educators have created various teaching methods to help students develop critical thinking skills (31,48,49) as well as diagnostic tests to measure student progress in scientific thinking (Table 1). Among them is the Test of Scientific Literacy Skills (TOSLS), which contains 28 multiple-choice questions pertaining to two major scientific literacy skills--scientific method validation and data interpretation (50). Furthermore, assessment tools, such as the Rubric for Experimental Design (51) and the Experimental Design Ability Test (52), are available for an instructor who wishes to evaluate student knowledge in experimental design.

### *Affective Assessment*

Student behavior is one of the critical factors linked to student learning and achievement (53-56). To determine any changes in student behavior and perception after the introduction of a teaching technique, questionnaires with multiple-choice and/or five- or seven-point Likert-scale questions have been developed (Table 1).

### **Disseminating the Results with the Community**

A variety of venues exist to share one's results with the greater community. For example, the American Society for Cell Biology, the Society for Developmental Biology, and many others welcome educational posters at their annual meetings.

**Table 1. Useful assessment tools for educational research.**

Concept Inventory Databases
<ul style="list-style-type: none"> <li>• DBER-Concept Inventories, Society for the Advancement of Biology Education Research (<a href="https://saber-biologyeducationresearch.wikispaces.com/DBER-Concept+Inventories">https://saber-biologyeducationresearch.wikispaces.com/DBER-Concept+Inventories</a>)</li> <li>• Questions for Biology Concept Inventories, The University of British Columbia (<a href="http://q4b.biology.ubc.ca/concept-inventories">http://q4b.biology.ubc.ca/concept-inventories</a>)</li> <li>• Concept Inventories/Conceptual Assessments in Biology, San Diego State University (<a href="https://go.sdsu.edu/dus/ctl/cabs.aspx">https://go.sdsu.edu/dus/ctl/cabs.aspx</a>)</li> </ul>
Critical Thinking Tests
<ul style="list-style-type: none"> <li>• Lawson’s Classroom Test of Scientific Reasoning (58–60)</li> <li>• California Critical Thinking Skills Test (<a href="http://www.insightassessment.com">http://www.insightassessment.com</a>; (61))</li> <li>• Critical Thinking Test (<a href="http://www.flaguide.org">http://www.flaguide.org</a>; (62, 63))</li> <li>• The Experimental Design Ability Test (52)</li> <li>• Critical Thinking Assessment Test (<a href="https://www.tntech.edu/cat">https://www.tntech.edu/cat</a>; (64))</li> <li>• The Rubric for Science Writing (65)</li> <li>• Test of Scientific Literacy Skills (50)</li> <li>• Rubric for Experimental Design (51)</li> <li>• Biological Experimental Design Concept Inventory (66)</li> </ul>
Affective Assessment
<ul style="list-style-type: none"> <li>• Motivated Strategies for Learning Questionnaire (67, 68)</li> <li>• The Scientific Attitude Inventory (69)</li> <li>• Measure of Acceptance of the Theory of Evolution (70, 71)</li> <li>• Student Assessment of Their Learning Gains (<a href="http://www.salgsite.org">www.salgsite.org</a>; (72))</li> <li>• The Classroom Undergraduate Research Experience and Summer Undergraduate Research Experience surveys (<a href="http://www.grinnell.edu/node/25703">www.grinnell.edu/node/25703</a>; (73))</li> <li>• An Extensive Review on Student Attitude Measures and Data Analysis (74)</li> </ul>

Additionally, many annual meetings also have educational symposia. The Society for the Advancement of Biology Education Research (<https://saber-biologyeducationresearch.wikispaces.com>) runs an annual meeting devoted to educational research and assessment. There are a variety of online resources (including *CourseSource*; <http://www.coursesource.org>) that publish peer-reviewed teaching materials. Journals like *CBE - Life Science Education* (<http://www.lifescied.org>) also provide a great vehicle to describe studies.

One barrier to sharing results from educational research is a worry that this public dissemination exposes the effectiveness of an instructor's classroom skills. Junior faculty in particular may be very self-conscious of their classroom performance, especially if teaching is a considerable part of the tenure process. Additionally, most biology instructors do not have enough time, support, or incentives from their schools to document their learning innovations and share them with other intramural or extramural instructors. This challenge underscores the importance of a departmental culture where teaching is highly valued. Educational research, especially involving studies that evaluate teaching effectiveness, should be included as part of the tenure review and teaching awards, as other colleagues have argued (57). This inclusion will allow educational research studies to be viewed as they should be - a positive sign that the faculty cares about their role as educators.

## Acknowledgments

We thank Doug Melton and Willy Lensch for advice and support of the work. We also are grateful to those readers who have provided constructive feedback.

## References

- American Association for the Advancement of Science. 2011. Vision and Change in Undergraduate Biology Education: A Call to Action. Washington, DC.
- National Research Council. 1997. Science Teaching Reconsidered: A Handbook. National Academies Press, Washington, DC.
- National Research Council. 2003. BIO2010: Transforming Undergraduate Education for Future Research Biologists. National Academies Press., Washington, DC.
- President's Council of Advisors on Science and Technology. 2012. Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Washington, DC.
- Slater SJ, Slater TF, Bailey JM. 2011. Discipline-Based Science Education Research: A Scientist's Guide. W. H. Freeman, New York, NY.
- 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 USA* 111:8410-5.
- Wood WB, Handelsman J. 2004. Meeting report: the 2004 National Academies Summer Institute on Undergraduate Education in Biology. *Cell Biol Educ* 3:215-7.
- Haak DC, Hillerislambers J, Pitre E, Freeman S. 2011. Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology. *Science* 332:1213-1217.
- Stockwell BR, Stockwell MS, Cennamo M, Jiang E. 2015. Blended Learning Improves Science Education. *Cell* 162:933-936.
- Knight JK, Wood WB. 2005. Teaching More by Lecturing Less. *CBE Life Sci Educ* 4:298-310.
- U.S. Department of Health and Human Services. Code of Federal Regulations - Title 45 Public Welfare CFR 46.
- Office of Human Research Protections, US Department of Health and Human Services. Exempt Research Determination FAQs.
- Hammack FM. 1997. Ethical Issues in Teacher Research. *Teach Coll Rec* 99:247-65.
- Roberts LW, Geppert C, Connor R, Nguyen K, Warner TD. 2001. An invitation for medical educators to focus on ethical and policy issues in research and scholarly practice. *Acad Med* 76:876-85.
- Pritchard IA. 2001. Searching for "research involving human subjects"-What is examined? What is exempt? What is exasperating? *IRB Ethics Hum Res* 23:5-13.
- Office of Human Research Protections, US Department of Health and Human Services. Quality Improvement Activities FAQs.
- Report B. 1979. The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research.
- Roberts LW, Geppert CMA, Coverdale J, Louie A, Edenharder K. 2005. Ethical and Regulatory Considerations in Educational Research. *Acad Psychiatry* 29:1-5.
- Gross D, Pietri ES, Anderson G, Moyano-Camihort K, Graham MJ. 2015. Increased Preclass Preparation Underlies Student Outcome Improvement in the Flipped Classroom. *Cell Biol Educ* 14:1-8.
- Eddy SL, Hogan KA. 2014. Getting Under the Hood: How and for Whom Does Increasing Course Structure Work? *CBE Life Sci Educ* 13:453-468.
- Labov J, Reid A, Yamamoto K. 2010. Integrated Biology and Undergraduate Science Education: A New Biology Education for the Twenty-First Century? *CBE Life Sci Educ* 9:10-16.
- Deslauriers L, Schelew E, Wieman C. 2011. Improved learning in a large-enrollment physics class. *Science* 332:862-864.
- Linton DL, Pangle WM, Wyatt KH, Powell KN, Sherwood RE. 2014. Identifying key features of effective active learning: the effects of writing and peer discussion. *CBE Life Sci Educ* 13:469-77.
- Drits-esser D, Bass KM, Stark LA. 2014. Using Small-Scale Randomized Controlled Trials to Evaluate the Efficacy of New Curricular Materials. *CBE Life Sci Educ* 13:593-601.
- Jensen JL, Kummer TA, Godoy PDM. 2015. Improvements from a Flipped Classroom May Simply Be the Fruits of Active Learning. *CBE Life Sci Educ* 14:1-12.
- Cronbach LJ, Furby L. 1970. How we should measure "change": Or should we? *Psychol Bull* 74:68-80.
- Pentecost TC, Barbera J. 2013. Measuring learning gains in chemical education: A comparison of two methods. *J Chem Educ* 90:839-845.
- Bereiter C. Some persisting dilemmas in the measurement of change, p. . In Harris, C (ed.), *Problems in Measuring Change*. The University of Wisconsin Press, Madison, WI.
- Hake RR. 1998. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am J Phys* 66:64.
- Bacon DR, Stewart KA. 2006. How Fast Do Students Forget What They Learn in Consumer Behavior? A Longitudinal Study. *J Mark Educ* 28:181-192.
- Rowe MP, Gillespie BM, Harris KR, Koether SD, Shannon L-JY, Rose LA. 2015. Redesigning a General Education Science Course to Promote Critical Thinking. *CBE Life Sci Educ* 14:1-12.
- D'Avanzo C. 2008. Biology Concept Inventories: Overview, Status, and Next Steps. *Bioscience* 58:1-7.
- Halloun IA, Hestenes D. 1985. The initial knowledge state of college physics students. *Am J Phys* 53:1043-1048.
- Hestenes D. 1998. Who needs physics education research!? *Am J Phys* 66:465-167.
- Klymkowsky MW, Garvin-Doxas K, Zeilik M. 2003. Bioliteracy and Teaching Efficacy: What Biologists Can Learn from Physicists. *Cell Biol Educ* 2:155-161.
- Anaya G. 1999. College Impact on Student Learning: Comparing the Use of Self-Reported Gains, Standardized Test Scores, and College Grades. *Res High Educ* 40:499-526.
- Benton SL, Duchon D, Pallett WH. 2013. Validity of student self-reported ratings of learning. *Assess Eval High Educ* 38:377-388.
- McNaught C, Ng SSW, Chow H. 2012. Literacies in the Humanities: the Student Voice. *High Educ Res Dev* 31:139-154.
- Bowman NA. 2011. Validity of College Self-Reported Gains at Diverse Institutions. *Educ Res* 40:22-24.
- Herzog S. 2011. Gauging Academic Growth of Bachelor Degree Recipients: Longitudinal vs. Self-Reported Gains in General Education. *New Dir Institutional Res* 150:21-39.
- Bok D. 2006. *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More*. Princeton University Press, Princeton, NJ.
- Brickman P, Gormally C, Francom G, Jardelezz SE, Schutte VGW, Jordan C, Kanizay L. 2012. *Media-Savvy Scientific Literacy: Developing Critical*

- Evaluation Skills by Investigating Scientific Claims. *Am Biol Teach* 74:374-379.
43. Coil D, Wenderoth MP, Cunningham M, Dirks C. 2010. Teaching the Process of Science: Faculty Perceptions and an Effective Methodology. *CBE Life Sci Educ* 9:524-535.
  44. Facione PA. 1990. *Critical Thinking : A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*The California Academic Press. California Academic Press, Millbrae, CA.
  45. Hodgson Y, Varsavsky C, Matthews KE. 2014. Assessment and teaching of science skills: whole of programme perceptions of graduating students. *Assess Eval High Educ* 39:515-530.
  46. National Research Council. 1996. *National Science Education Standards*1996. National Academies Press, Washington, DC.
  47. van Gelder T. 2005. Teaching Critical Thinking: Some Lessons From Cognitive Science. *Coll Teach* 53:41-48.
  48. Iyengar R, Diverse-pierluissi MA, Jenkins SL, Chan AM, Lakshmi A, Sobie EA, Ting AT, Weinstein DC. 2008. Integrating Content Detail and Critical Reasoning by Peer Review. *Science* 319:1189-1190.
  49. Quitadamo IJ, Faiola CL, Johnson JE, Kurtz MJ. 2008. Community-based Inquiry Improves Critical Thinking in General Education Biology. *CBE Life Sci Educ* 7:327-337.
  50. Gormally C, Brickman P, Lutz M. 2012. Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE Life Sci Educ* 11:364-77.
  51. Dasgupta AP, Anderson TR, Pelaez N. 2014. Development and Validation of a Rubric for Diagnosing Students' Experimental Design Knowledge and Difficulties. *CBE Life Sci Educ* 13:265-284.
  52. Sirum K, Humburg J. 2011. The Experimental Design Ability Test (EDAT). *Bioscene* 37:8-16.
  53. Donald JG. 2002. *Learning to Think: Disciplinary Perspectives*. Jossey-Bass, San Francisco.
  54. Dehaan RL. 2005. The impending revolution in undergraduate science education. *J Sci Educ Technol* 14:253-269.
  55. House JD. 1994. Student Motivation and Achievement in College Chemistry. *Int J Instr Media* 21:1-11.
  56. Oliver SJ, Simpson RD. 1988. Influences of Attitude Toward Science, Achievement Motivation, and Science Self Concept on Achievement in Science: A Longitudinal Study. *Sci Educ* 72:143-155.
  57. Anderson W a, Banerjee U, Drennan CL, Elgin SCR, Epstein IR, Handelsman J, Hatfull GF, Losick R, O'Dowd DK, Olivera BM, Strobel S a, Walker GC, Warner IM. 2011. Changing the culture of science education at research universities. *Science* 331:152-153.
  58. Johnson MA, Lawson AE. 1998. What Are the Relative Effects of Reasoning Ability and Prior Knowledge on Biology Achievement in Expository and Inquiry Classes? *J Res Sci Teach* 35:89-103.
  59. Lawson AE. 1978. The Development and Validation of a Classroom Test of Formal Reasoning. *J Res Sci Teach* 15:11-24.
  60. Lawson AE, Alkhoury S, Benford R, Clark BR, Falconer KA. 2000. What kinds of scientific concepts exist? Concept construction and intellectual development in college biology. *J Res Sci Teach* 37:996-1018.
  61. Facione P, Facione N, Giancarlo C. 1996. *The California Critical Thinking Disposition Inventory: Test manual*. California Academic Press, Millbrae, CA.
  62. Hoskins SG, Stevens LM, Nehm RH. 2007. Selective use of the primary literature transforms the classroom into a virtual laboratory. *Genetics* 176:1381-1389.
  63. Stevens LM, Hoskins SG. 2014. The CREATE Strategy for Intensive Analysis of Primary Literature Can Be Used Effectively by Newly Trained Faculty to Produce Multiple Gains in Diverse Students. *CBE Life Sci Educ* 13:224-242.
  64. Stein B, Haynes A. 2011. Engaging Faculty in the Assessment and Improvement of Students' Critical Thinking Using the Critical Thinking Assessment Test. *Chang Mag High Learn* 43:44-49.
  65. Timmerman BEC, Strickland DC, Johnson RL, Payne JR. 2011. Development of a "universal" rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assess Eval High Educ* 36:509-547.
  66. Deane T, Nomme K, Jeffery E, Pollock C, Birol G. 2014. Development of the Biological Experimental Design Concept Inventory (BEDCI). *CBE Life Sci Educ* 13:540-51.
  67. Crede M, Phillips LA. 2011. A meta-analytic review of the Motivated Strategies for Learning Questionnaire. *Learn Individ Differ* 21:337-346.
  68. Pintrich PRR, Smith D, Garcia T, McKeachie W. 1991. *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. University of Michigan, Ann Arbor, MI.
  69. Moore RW, Foy RLH. 1997. The Scientific Attitude Inventory: A Revision (SAI II) Richard. *J Res Sci Teach* 34:327-336.
  70. Rutledge ML, Sadler KC. 2007. Reliability of the Measure of Acceptance of the Theory of Evolution (MATE) Instrument with University Students. *Am Biol Teach* 69:332-335.
  71. Rutledge ML, Warden MA. 1999. The Development and Validation of the Measure of Acceptance of the Theory of Evolution Instrument. *Sch Sci Math* 99:13-18.
  72. Seymour E, Wiese DJ, Hunter A, Daffinrud SM. 2000. *Creating a Better Mousetrap: On-line Student Assessment of their Learning Gains*Presented to the National Meetings of the American Chemical Society Symposium. San Francisco, CA.
  73. Denofrio LA, Russell B, Lopatto D, Lu Y, Curricula S, Lu Y. 2007. Linking Student Interests to Science Curricula. *Science* 318:1872-1873.
  74. Lovelace M, Brickman P. 2013. Best practices for measuring students' attitudes toward learning science. *CBE Life Sci Educ* 12:606-617.