Research and Teaching

Two-Dimensional, Implicit Confidence Tests as a Tool for Recognizing Student Misconceptions

By Michael W. Klymkowsky, Linda B. Taylor, Shana R. Spindler, and R. Kathy Garvin-Doxas

The misconceptions that students bring with them, or that arise during instruction, are a critical barrier to learning. Implicit-confidence tests, a simple modification of the multiple-choice test, can be used as a strategy for recognizing student misconceptions. An important issue, however, is whether such tests are gender-neutral. We analyzed the results of exams administered to students (both majors and nonmajors) in Molecular, Cellular and Developmental Biology (MCDB) 1111: Biofundamentals at the University of Colorado at Boulder. At a statistically significant level (> 95%), there was no difference between women and men regardless of whether their answers were confidently correct or incorrect, suggesting that such two-dimensional tests are a gender-neutral tool.

I t is well established that a major barrier to learning are the misconceptions that students bring with them into the classroom or that develop during the course of instruction (Committee on Undergraduate Science Education 1997). Conventional teaching strategies, even those that directly address identified misconceptions, often fail to change students' deeply held ideas (see as an example Gregg et al. 2001, which describes a study

Michael W. Klymkowsky (klymkowsky@bioliteracy.net) *is director of the Bioliteracy Group, which is based in the Molecular, Cellular and Developmental Biology Department at the University of Colorado at Boulder.* **Linda B. Taylor,** *graduated from the University of Colorado at Boulder with degrees in molecular, cellular, and developmental biology and psychology and is currently doing graduate work in cognitive science.* **Shana R. Spindler** *is a graduate student in molecular, cellular, and developmental biology at the University of California, Los Angeles.* **R. Kathy Garvin-Doxas** *is a research associate in the Center for Integrated Plasma Studies at the University of Colorado at Boulder.*

on students' views on vision and the effects of various interventions). While critical to recognize and address, the exact nature of student misconceptions can be difficult to identify through standard classroom interactions, particularly in larger classes. In a study on the prevalence among prospective elementary teachers of misconceptions about blood circulation and gas exchange, it was noted that different ways of asking students to display their understanding revealed distinct types of misconceptions (Palaez et al. 2005). For example, student drawing appears to be a particularly effective strategy in uncovering misconceptions. A similar insight is illustrated in the film *A Private Universe* (Harvard-Smithsonian Center for Astrophysics 1989) with respect to the relative movements of the Earth, Moon, and Sun.

Generally, student misconceptions persist until students recognize that their understanding does not work and that following it leads to incorrect answers or illogical conclusions. For

a question or task to lead students to recognize and address their mistaken assumptions, it must force students to actively use their conceptual understanding. A particularly effective strategy to lead students to such a "eureka" moment involves the use of tutorials, conceptual problems that require students to apply information and ideas they have learned to solving novel problems. Tutorials have been developed in physics (see Redish 2003 and references therein) and astronomy (e.g., Adams et al. 2005), and we have begun to develop, test, and revise similar tutorials for introductory biology (*www.colorado.edu/MCDB/MCDB1111/Tutorial%20TOC.htm*).

Without identifying and addressing the misconceptions students harbor, robust and resilient understanding is nearly impossible to achieve. Instructors commonly assume that conventional testing strategies, whether multiple-choice or essay, evaluate students' conceptual understanding, but this has been shown not to be the case (Pelaez et al. 2005). More often than not, testing instruments require that students simply recognize or remember the correct answer rather than use their understanding to construct it. In this context, there is little impetus for real learning; misconceptions, if present, remain firmly in place. Exams of this type, whether multiple-choice or essay, have been labeled "inauthentic" by McClymer and Knoles (1992). Good grades on such tests often lead

both students and instructors to the mistaken belief that learning has occurred, when in fact students may still harbor significant misconceptions that actively interfere with true understanding. A critical ramification of such ersatz learning is that teachers from the elementary through the college level can transmit their own unresolved misconceptions to their students (see *www.ems.psu.edu/~fraser/BadScience.html*). Given the high percentage of high school teachers, as well as members of the general public, who favor teaching various forms of creationism, often to the exclusion of biology (Scott 1999; Pew Research Center Report 2005), it is probably no exaggeration to suggest that both high school and college biology courses are failing to address serious misconceptions about biology in general and evolutionary biology in particular (Visit *http://bioliteracy.net/Comment-Alberts.htm* for more background on this topic.)

One attempt to identify and address student misconceptions is the use of concept inventories. The first and most influential of these is the Force Concept Inventory (FCI) (Hestenes, Wells, and Swackhamer 1992). Designed to probe conceptual understanding of Newton's laws of motion, rather than the ability to memorize terms or manipulate equations, the use of the FCI revealed that high student grades often did not correlate with a robust conceptual understanding (Hake 1998; Powell 2003). Since the introduction of the FCI, other concept inventories have begun to appear. Among these are the Force and Motion Concept Evaluation (FMCE) (Sokoloff and Thorton 1998), the Brief Electricity and Magnetism Assessment (Ding et al. 2006), the Quantum Mechanics Concept Survey (McKagan and Wieman 2005), the Natural Selection Concept Inventory (Anderson, Fisher, and Norman 2002), and our own Biology Concept Inventory (available at *bioliteracy.net*). However, because they are gen-erated through research to uncover student misconceptions and validated through student interviews, concept inventories are both expensive and time consuming. More importantly, they are not meant for the assessment of individual students, but rather as a pre/postinstructional instrument to evaluate learning gains associated with specific teaching strategies.

Given their role as a teaching assessment tool, concept inventories do not address an instructor's immediate and ongoing need to identify their own students' misconceptions. This raises the practical question: Are there simple and efficient ways for instructors to recognize at least some of their students' misconceptions in time to address them? Clearly, the most direct method, engaging in Socratic interactions during which students are encouraged to explicitly define their terms and enunciate their assumptions and approach to a particular problem, is generally impractical. An increasingly popular alternative is the use of electronic student-response systems (Ward 2003; Wood 2004; Dye 2005), by which instructors can replace the generally rhetorical, "Are there any questions?" and, "Does everyone get it?" with an immediate, and often enlightening, assessment of student understanding. A complementary approach involves a simple modification of the standard multiple-choice test.

Information reference or two-dimensional multiple-choice testing

Two-dimensional tests (TDTs) aim to determine students' certainty, or confidence, in their answers (Bruno 1993). Certainty in answers does not necessarily translate into student self-confidence in their abilities, as is usually defined in the literature. However, the confidence nomenclature is well established in the context of TDTs and we will follow this convention here. There are two versions of a confidence-based TDT: self-assessed, or explicit and implicit (Figure 1). In a self-assessed, confidence-based TDT, each question consists of two parts: a conventional multiple-choice question paired with a second part that asks the student to indicate their level of confidence in their answer. For example, they can indicate whether they are "very confident," "semi-confident," or "guessing." In an implicit confidence test, the level of student confidence is reflected by a single response. Students are assured that one and only one response is correct, but they are given the option to

FIGURE 1

Examples of explicit- (top) and implicit- (bottom) confidence test questions.

FIGURE 2

Grading scheme for two-dimensional test questions.
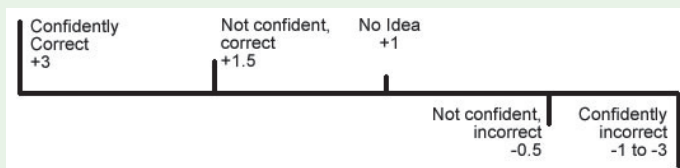
waffle—if they are not sure whether the correct answer is "A" or "B," they can pick "A or B." If they have no idea as to which is the correct choice, they can indicate that as well. While both types of TDTs assess student confidence, implicit tests have the advantage in that students do not have to explicitly assess their confidence level—their choice of answer contains that information.

A critical component in both types of TDTs is the way that points are assigned to particular responses (Figure 2). Students receive maximum credit for confidently correct answers and the maximum penalty, in the form of negative points, for confidently incorrect answers. As students reveal their uncertainty, the number of points they receive or lose is reduced. One consequence of this grading strategy is that, because an incorrect answer is penalized more than admitting one's ignorance, students' optimal test-taking strategy is dramatically altered (Bruno 1993). For this reason, the structure of the exam may initially upset many students until they adapt to it, which usually occurs quite quickly.

While there are important benefits associated with changing the strategies students use in answering questions, the pedagogical value of using TDTs involves formative assessment—it does not take much time or effort to identify those questions where a large percentage of students have answered confidently wrong. These questions serve as signposts for student misconceptions; the more closely the incorrect answers (distracters) reflect student misconceptions, as they are designed to do in a well-researched and validated concept inventory, the more student responses mirror their actual misconceptions. Even with poorly designed distracters, however, TDT responses provide quick and often valuable feedback to the instructor. Assuming that the questions themselves are not seriously flawed, if the majority of students taking the test are confidently correct on a particular question, then it is likely that the instructor has effectively conveyed the relevant concepts. If, however, the majority of students are confidently wrong on a particular question, it is likely that students are misinformed or that the instructor has failed to lead students to a robust understanding, perhaps because of interfering (and unrecognized) student misconceptions. If many students answer "semi-confident" or "guessing/no idea," instructors may need to evaluate their approach to presenting the materials or encouraging student engagement with it.

To use the information garnered from TDTs in learning interventions, it is important that students have an incentive to revisit and correct their understanding. A number of strategies are possible. In MCDB 1111: Biofundamentals, we use what we call "I know it now" (IKIN) exams. These exams are in the implicit TDT format, but without the "no idea/guessing" option. IKIN exams are given during the final exam, and the points gained (or possibly lost) are added to the original midterm exam scores. In this scenario, the original exam score is seen as provisional and can be improved through further study.

Confidence-based testing and gender concerns

Beginning in the spring 2003 semester, we implemented confidence-based TDTs in MCDB 1111: Biofundamentals. During the spring of 2004, we conducted an informal survey (Figure 3) to assess student attitudes toward multiple-choice tests (42 students responding out of 75). Approximately 42% of the responding students liked multiple-choice exams because they could guess and use partial knowledge to get the answer "right," and/or they did not feel that they had to study

FIGURE 3

MCDB 1111 multiple-choice test survey

We are conducting a survey of student opinions about testing. We would appreciate your responses to these questions.

1. Please finish the following statement: I like multiple-choice questions because
   a. I can use partial knowledge to select answers, and no one will know that I am not confident in my answer.
   b. I do not have to study as hard or commit as much to memory because I only need to recognize the answer
   c. Without knowing anything at all about the question, there is a still the chance that I can get it right
   d. None of the above capture my feelings about multiple-choice questions
   e. I do not like multiple-choice questions
2. If you do not like multiple-choice questions, tell us why.
3. Do you think that traditional multiple-choice questions accurately evaluate your knowledge of the subject matter? yes or no
4. What do you think is the best way to evaluate student learning?

as much for multiple-choice exams. That said, 51% indicated that simple multiple-choice exams did not assess their true knowledge of the subject. A number of students had concerns that confidence testing would lead to a bias against females, on the assumption that females would be less likely to mark "confident" on explicit confidence tests and so would be less likely to score full points. The literature echoes students' concerns; it is generally accepted that female students in science disciplines are less confident than males (Seymour and Hewitt 1997; Light et al. 2000). That said, there is a significant difference between self-confidence and certainty in test answers. To address this concern and determine whether female students were less likely to answer questions confidently, we performed a statistical analysis of implicit (embedded-confidence) TDTs to determine whether females answer "semi-confident" rather than "confident" more often than males, and whether males score better on average than females when implicit TDT's were used.

Statistical methods

The MCDB 1111: Biofundamentals course (spring 2003/2005) at the University of Colorado at Boulder was used for all data collection. The 2003 class consisted of 39 females and 36 males, and the 2005 class consisted of 25 females and 26 males; approximately 30% of all students in both years were MCDB majors. All data were collected after the drop/add deadline for the course. In 2003, three midterm TDTs were analyzed. In 2005, the final TDT (but not the IKIN exams) was analyzed. Each midterm exam had 20 to 25 implicit TDT questions. Two-tailed t-test analysis ($> 95\%$ confidence level) (Harris 1998) was used to identify any significant difference between the number of times females and males chose confident versus semi-confident answers. In addition, two-tailed t-tests were performed to determine if there was a significant difference in overall exam scores between males and females. We also reexamined the data using the less rigorous Chi square test and again found no significant difference between the numbers of confidently answered questions (whether correct or incorrect), between males and females.

For all three exams in 2003, males and females did equally well (Figure 4, left panel), and there was no statistically significant difference in the frequency of confident/semi-confident/guess choices between the two groups. Out of a total of 3,601 "confident" male answers and 3,960 "confident" female answers, males were correct 82% of the time, while females were correct 80% of the time. Similar results were found in 2005 (Figure 4, right panel), with males and females displaying similar levels of confidently correct and confidently incorrect responses.

Summary

Even the simplest, most rigorously designed multiple-choice tests consistently overestimate student learning and provide little feedback to the instructor about the nature or source of student mistakes (Bruno 1993). If a student guesses on an exam question and does so correctly, the student rarely revisits the information. Misconceptions, if present, go unrecognized and unaddressed. These misconceptions are not benign; they can lead to a cascade of misunderstandings that result in intellectual confusion and an inability to build on and apply past knowledge to new situations. In contrast, implicit confidence TDTs change the testing landscape and provide instructors with valuable feedback about student confusions, as well as identify material that students are uncertain about or areas in which they are actively misinformed. If linked to other student-response systems and targeted interventions (e.g., tutorials), implicit-confidence TDTs can become an important tool in increasing student understand-

FIGURE 4

Result of response analysis.
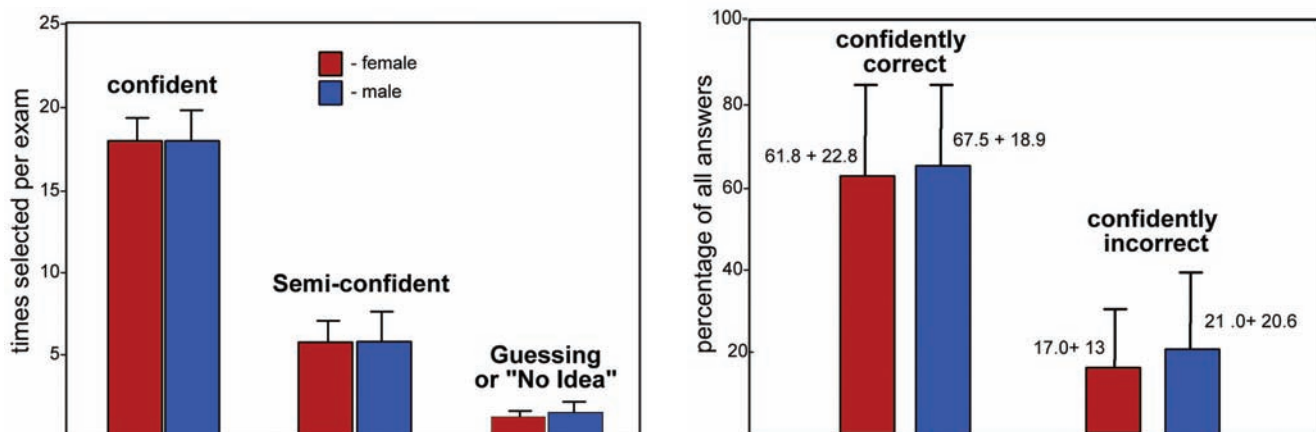
ing and serve as a reality check on runaway course content, which we refer to as syllabus bloat (Klymkowsky 2005). It is easy to transform standard multiple-choice tests into implicit TDTs, and straightforward to identify questions that provoke confidently incorrect or confused student responses. Moreover, it appears that implicit-confidence TDTs are not gender-biased and so may be implemented without placing females at an instructional disadvantage.

References

Adams, J., E.E. Prather, T. Slater, and J. Dostal. 2005. *Lecture tutorials for introductory astronomy*. Upper Saddle River, NJ: Prentice-Hall.

Anderson, D.L., K.M. Fisher, and G.J. Norman. 2002. Development and validation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching* 39: 952–78.

Bruno, J.E. 1993. Using testing to provide feedback to support instruction: A reexamination of the role of assessment in educational organizations. In *Item banking: Interactive testing and self-assessment,* eds. D.A. Leclercq and J.E. Bruno, 190–209. New York: Springer Verlage. Available at *www.p-mmm.com/founders/UsingTestingBody.htm*.

Committee on Undergraduate Science Education. 1997. *Science teaching reconsidered : A handbook*. Washington, DC: National Academy Press.

Ding, L., R. Chabay, B. Sherwood, and R. Beichner. 2006. Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics: Physics Education Research* 2 (1): 7.

Duit, R. 2006. Students' and teachers' conceptions and science education: A bibliography. Available at *www.ipn.uni-kiel.de/aktuell/stcse/stcse.html*.

Dye, L. 2005. Students use clickers to help guide college lectures. Available at *http://abcnews.go.com/Technology/DyeHard/story?id=727409andpage=1*.

Gregg, V.R., G.A. Winer, J.E. Cottrell, K.E. Hedman, J.S. Fourneir. 2001. The persistence of a misconception about visions after educational interventions. *Psychonomic Bulletin and Review* 8: 622–26.

Hake, R.R. 1998. Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics* 66 (1): 64–74.

Harris, M.B. 1998. *Basic statistics for behavioral science research.* Needham Heights, MA: Allyn and Bacon.

Harvard-Smithsonian Center for Astrophysics. 1989. A private universe. Available at *www.learner.org/resources/series28.html*.

Hestenes, D., M. Wells, and G. Swackhamer. 1992. Force concept inventory. *Physics Teacher* 30 (3): 141–58.

Klymkowsky, M.W. 2005. Can non-majors courses lead to biological literacy? Do major's courses do any better? *Cell Biology Education* 4: 196–98.

Light, P., K. Littleton, S. Bale, R. Joiner, and D. Messer. 2000. Gender and social comparison effects in computer-based problem solving. *Learning and Instruction* 34 (6): 483–96.

McClymer, J.F., and L.Z. Knoles. 1992. Ersatz learning, inauthentic testing. *Journal of Excellence in College Teaching* 3: 33–50. Available at *http://bioliteracy.net/Readings/ErsatzLearning.pdf*.

McKagan, S.B., and C.E. Wieman. 2005. Exploring student understanding of energy through the quantum mechanics conceptual survey. *AIP Conference Proceedings* 818: 65–68. Available at *http://cosmos.colorado.edu/phet/survey/QMCS*.

Pelaez, N.J., D.D. Boyd, J.B. Rojas, and M.A. Hoover. 2005. Prevalence of blood circulation misconceptions among prospective elementary teachers. *Advances in Physiological Education* 29: 172–81.

Pew Research Center Report. 2005. Public divided on origins of life. Available at *http://pewforum.org/publications/surveys/religion-politics-05.pdf*.

Powell, K. 2003. Science education: Spare me the lecture. *Nature* 425 (6955): 234–36.

Redish, E.F. 2003. *Teaching physics with the physics suite*. Hoboken, NJ: John Wiley and Sons.

Scott, E.G. 1999, The creation/evolution continuum. *NCSE Reports* 19 (4): 16–17, 23–25.

Seymour, E., and N.M. Hewitt. 1997. *Talking about leaving: Why undergraduates leave the sciences.* Boulder, CO: Westview Press Boulder.

Sokoloff, D.R., and R.K. Thornton. 1998. Assessing student learning of Newton's laws: The force and motion conceptual evaluation of active learning laboratory and lecture curricula, *American Journal of Physics* 66 (4): 338–52.

Ward, D.L. 2003. The classroom performance system: The overwhelming research results supporting this teacher tool and methodology. Available at *www.einstruction.com/index.cfm? fuseaction=news.displayandmenu=newsandcontent=showArticleandid=66*.

Wood, W.B. 2004. Clickers: A teaching gimmick that works. *Developmental Cell* 7 (6): 796–98.