Article

A Diagnostic Assessment for Introductory Molecular and Cell Biology

Jia Shi,*[†] William B. Wood,*[†] Jennifer M. Martin,* Nancy A. Guild,* Quentin Vicens,[§‖] and Jennifer K. Knight*[†]

Departments of *Molecular, Cellular, and Developmental Biology and [§]Chemistry and Biochemistry, and the [†]Science Education Initiative, University of Colorado, Boulder, CO 80309

We have developed and validated a tool for assessing understanding of a selection of fundamental concepts and basic knowledge in undergraduate introductory molecular and cell biology, focusing on areas in which students often have misconceptions. This multiple-choice Introductory Molecular and Cell Biology Assessment (IMCA) instrument is designed for use as a pre- and posttest to measure student learning gains. To develop the assessment, we first worked with faculty to create a set of learning goals that targeted important concepts in the field and seemed likely to be emphasized by most instructors teaching these subjects. We interviewed students using open-ended questions to identify commonly held misconceptions, formulated multiple-choice questions that included these ideas as distracters, and reinterviewed students to establish validity of the instrument. The assessment was then evaluated by 25 biology experts and modified based on their suggestions. The complete revised assessment was administered to more than 1300 students at three institutions. Analysis of statistical parameters including item difficulty, item discrimination, and reliability provides evidence that the IMCA is a valid and reliable instrument with several potential uses in gauging student learning of key concepts in molecular and cell biology.

INTRODUCTION

The use of multiple-choice tests designed to evaluate conceptual understanding and diagnose areas of difficulty in specific science disciplines (concept assessments or concept inventories) has expanded significantly in recent years. The Physics Force Concept Inventory (FCI; Hestenes, 1992) is generally acknowledged as the first of these assessments used to provide instructors with a measure of student conceptual understanding in introductory mechanics courses. Additional assessment instruments have since been devel-

oped for courses in other disciplines, including astronomy, chemistry, geological sciences, and life sciences (reviewed in Libarkin, 2008). Several groups have used these instruments to demonstrate that student conceptual understanding is enhanced in courses that emphasize interactive learning (e.g., Hake, 1998; Crouch and Mazur, 2001; Knight and Wood, 2005; Ding *et al.*, 2006). In the life sciences, a variety of assessments have been developed in areas such as natural selection (Anderson *et al.*, 2002), animal development (Knight and Wood, 2005), energy and matter (Wilson *et al.*, 2007), introductory biology (Garvin-Doxas *et al.*, 2007), genetics (Bowling *et al.*, 2008; Smith *et al.*, 2008), molecular life sciences (Howitt *et al.*, 2008), and host–pathogen interactions (Marbach-Ad *et al.*, 2009), but more instruments are needed that both address student misconceptions and align well with content normally taught in undergraduate biology courses.

Many educators have explored concepts for which students have either an incomplete understanding or more strongly held misconceptions (sometimes referred to as alternative conceptions; Tanner and Allen, 2005). For example, when some students describe biological examples of energy

flow (including photosynthesis and cellular respiration), they can give simple descriptions of these processes but can neither elaborate on their functions nor relate them to each other (Barak *et al.*, 1999). Students also often hold naïve beliefs, such as thinking that respiration is possible in plant leaves only because of special pores for gas exchange (Haslam and Treagust, 1987). Some of these ideas change as students take additional biology courses, but many, such as the role of carbon dioxide as a raw material for plant growth, and the movement of carbon through a cycle, remain poorly understood (Wandersee, 1985; Ebert-May *et al.*, 2003). On the topic of diffusion and osmosis, students often believe that molecules cease to move at equilibrium (Odom, 1995) and that molecules experience directed movement toward lower concentrations, rather than random movement (Meir *et al.*, 2005; Garvin-Doxas and Klymkowsky, 2008). Although implementing active-learning techniques such as computer simulations (Meir *et al.*, 2005) or the 5E learning cycle (Balci *et al.*, 2006; Tanner, 2010) help some students overcome these beliefs, their persistence indicates a continuing need to identify and create materials that help students change their ideas.

The misconceptions described above, and many others not mentioned, helped us to frame conversations with faculty as we developed the learning goals upon which the IMCA is based. In contrast to introductory physics, the topics emphasized in introductory biology courses can vary substantially depending on the differing needs and interests of departments, instructors, and student audiences. Therefore, targeted, area-specific assessments may be more useful than a single comprehensive test suitable for any introductory course. We chose to develop an assessment limited to concepts that are likely to be addressed in any introductory course in molecular and cell biology and that represent areas of common student misconceptions.

By following the guidelines outlined by Treagust (1988) for developing diagnostic tests to evaluate student misunderstanding, we created an instrument consisting of 24 multiple-choice questions that are as free of scientific jargon as possible and that include distracters reflecting common student misconceptions. The IMCA does not address noncontent learning goals and science process skills such as formulating hypotheses and interpreting data. Other assessments for assessing these skills have been developed (e.g., Brickman *et al.*, 2009).

When used as a pretest at the start of a course before any instruction and as a posttest at the end of the course, the IMCA can measure overall student learning gains, performance in specific content areas by learning goal, and areas in which misconceptions are held. We report here on the development process of the IMCA and describe validation of the assessment through student interviews, pilot testing, expert review, and statistical analysis. We also discuss the ways in which data from such an assessment can be used to understand student ideas and inform improvements in instruction.

METHODS

Development and Validation of the IMCA

There is still considerable variability in how science concept assessments are designed and validated (Lindell *et al.*, 2007; Adams and Wieman, 2010), as well as in their format and

intended use, even within biology (reviewed in D'Avanzo, 2008; Knight, 2010, among others). The IMCA was developed to be used as a pre- and posttest to measure change in student understanding of principles most likely to be taught in an introductory biology course that focuses on molecular and cell biology. It is also intended to be diagnostic of common student difficulties, as many questions are built around known student misconceptions. Development of the IMCA followed a multistep process (Table 1) similar to that used for the Genetics Concept Assessment (Smith *et al.*, 2008) and based on the guidelines of diagnostic assessment design (Treagust, 1988).

Development of the IMCA was begun by defining the content of the assessment through an iterative process involving discussion with faculty who teach introductory cell and molecular biology (including authors N.A.G., J.M.M., and Q.V.), as well as faculty who teach courses for which the introductory courses are prerequisites. The concepts addressed were chosen as those that students would be most likely to need for understanding more advanced topics in biology, and those about which students commonly have persistent incorrect ideas, based on our experience as well as a literature review of common misconceptions in cell and molecular biology (e.g., Odom, 1995; Marbach-Ad and Stavy, 2000; Anderson *et al.*, 2002; Gonzalez-Cruz *et al.*, 2003; Garvin-Doxas and Klymkowsky, 2008). Faculty also felt that students should be able to demonstrate knowledge of a few basic principles in addition to deeper conceptual understanding. Accordingly, a few questions on the assessment were written at a level below "application" in Bloom's taxonomy (Bloom *et al.*, 1956). We believe this range of questions is valuable for allowing instructors to identify students who have difficulty with application problems because they have not learned the basic content.

After agreement was reached on the fundamental knowledge and concepts that the assessment should address, they were reformulated as a set of specific learning objectives (referred to throughout this paper as learning goals), specifying what students should be able to do to demonstrate understanding of these concepts. A pilot set of learning

---

Table 1. Overview of the IMCA development process

1. Interview faculty who teach the introductory course and courses that follow, and develop a set of learning goals that most instructors consider essential to the understanding of basic cell and molecular biology.
2. Interview students to probe their understanding of these topics.
3. Review the literature concerning misconceptions in molecular and cell biology.
4. Develop and administer a pilot assessment based on learning goals and observed student difficulties and misconceptions.
5. Revise the assessment based on pilot results and further student feedback to eliminate jargon, modify distracters to include student-supplied incorrect answers, and rewrite questions answered correctly by >75% of students.
6. Validate and finalize the revised assessment through student interviews and input from introductory biology faculty at several institutions.
7. Administer the assessment to a large number of students.
8. Evaluate the assessment for item difficulty, item discrimination, and reliability.

goals was then used as the basis for open-ended student interviews and question construction, described below. Subsequently, the goals were further revised based on conversations with faculty at other institutions and on the results of further student interviews. The final set of nine core learning goals is shown in Table 2, each with a representative common student misconception or incorrect idea derived from student interviews and student answers on the final IMCA (see below).

Student Interviews

A diverse group of 41 students (15 males and 26 females) at the University of Colorado (CU) were interviewed. Thirty-six students had completed the Introduction to Cell and Molecular Biology course in the department of Molecular, Cellular, and Developmental Biology (MCDB) during the 2006–2007 academic year, earning grades ranging from A to D (14 A, 14 B, 6 C, and 2 D). The remaining five students had completed an introductory biology course in the CU Ecology and Evolutionary Biology Department; this course overlaps with the MCDB course in several but not all areas of content.

The initial goal of the student interviews was to probe student thinking, including the presence of misconceptions on these topics, and to then use student ideas to generate incorrect answer choices (distracters). The interviewer (J.S.) met one-on-one with students (n = 6) for these interviews using a "think aloud" protocol, in which students explained their answers to a series of open-ended questions that related to the nine identified learning goals. Emphasis was placed on letting the students explain the reasoning for their answers, so that the interviewer could explore student ideas.

Subsequent interviews were carried out not only to obtain evidence regarding the validity of the instrument (construct validity) but also to help modify the items of the assessment. In these interviews students were given multiple-choice questions that had been built using the main question stems initially asked as open-ended questions, with distracters derived primarily from student responses. Students were asked to select an answer to each question, and then, after answering all questions, to explain why they thought their choices were correct and the other choices incorrect. Student interview transcripts were again used to revise both correct and incorrect answer choices on the assessment.

Initially, multiple questions were designed to address each of the nine learning goals; however, some questions were eliminated during the validation process, so that some learning goals are addressed by only one question in the current version of the assessment (Table 2). For every question on the pilot version of the IMCA, at least five students chose the right answer using correct reasoning. However, for six questions, some students chose the right answer for incorrect or incomplete reasons. To decrease the likelihood of students guessing the right answer, we reworded these

Table 2. Learning goals, corresponding questions on the IMCA, and the most common incorrect student ideas relating to each

Learning goal	Question[a]	Incorrect student ideas or confusion[b]
1. Outline the theory of evolution, citing evidence that supports it and properties of organisms that it explains.	1	Mutations are directed, not random.
2. Contrast the features that distinguish viruses, bacteria, and eukaryotic cells.	2, 3	Bacteria can have RNA as genetic material, but viruses have only DNA.
3. Recognize structures of the four major classes of building-block molecules (monomers) that make up cellular macromolecules and membranes.	4–8	Students struggle to distinguish between the molecular structures of phospholipids and fatty acids and between monosaccharides, amino acids, and nucleotides.
4. Compare how the properties of water affect the three-dimensional structures and stabilities of macromolecules, macromolecular assemblies, and lipid membranes.	9, 10	Students do not understand the properties of polar molecules.
5. Given the thermodynamic and kinetic characteristics of a biochemical reaction, predict whether it will proceed spontaneously and the rate at which it will proceed.	11–14	Enzymes act by changing the equilibria of chemical reactions rather than by increasing their rates.
6. From their structures, predict which solutes will be able to diffuse spontaneously through a pure phospholipid bilayer membrane and which will require transport by membrane-associated proteins.	15	Ions, because of their small size, can diffuse through membranes.
7. Outline the flow of matter and energy in the processes by which organisms fuel growth and cellular activities, and explain how these processes conform to the laws of thermodynamics.	16–18	Oxygen is used in the formation of CO_2 during cellular respiration.
8. Using diagrams, demonstrate how the information in a gene is stored, replicated, and transmitted to daughter cells.	19–21	Individual chromosomes can contain genetic material from both parents.
9. Describe how the information in a gene directs expression of a specific protein.	22–24	Promoter regions are part of the coding region of a gene.

[a] The learning goals associated with each question are those intended by the authors and supported by biology faculty expert responses (see Table 3).

[b] Ideas listed are representative of the most commonly chosen wrong answers on the pretest (n > 700 students), as well as from answers students gave during the interview process.

Table 3. Summary of expert responses to three queries about the 24 IMCA questions

Subject of query	Agreement of experts		
	>90%	>80%	>70%
	No. of questions		
The question tests achievement of the specified learning goal	5	17	2
The information given in this question is scientifically accurate	17	7	0
The question is written clearly and precisely	9	14	1

questions and then interviewed additional students, obtaining an average total of 25 interviews for each question in the final version of the IMCA. For these questions, 90% (on average) of the interviewed students who chose the right answers explained their reasoning correctly. For the final version of the IMCA, each distracter was chosen by two or more students during interviews. Common incorrect student ideas are listed in Table 2.

Faculty Reviews

In addition to the internal review of the IMCA by faculty at CU, evidence for content validity was obtained by asking Ph.D. faculty experts who teach introductory biology at other institutions to take the IMCA online, respond to three queries about each question, and offer suggestions for improvement. Ten experts reviewed the pilot version of the IMCA; their feedback was used to modify some of the questions. An additional 15 experts took the final version; as summarized in Table 3, the majority (>80%) of this group confirmed that the questions tested achievement of the specified learning goals, were scientifically accurate, and were written clearly.

Administration and Measurement of Learning Gains

A pilot version of the IMCA was given in Fall 2007 to 376 students at the beginning and end of the Introduction to Cell and Molecular Biology course at CU. During the 2007–2008 academic year, the assessment was further modified on the basis of student interviews and input from outside expert consultants, as described above. Questions on which students scored >75% correct on the pretest were replaced or rewritten to make them more difficult and thereby more sensitive discriminators of conceptual understanding.

Subsequently, 1337 students took the current version of the IMCA, as both pre- and posttests. The actual number of students in each course was higher than reported here, but only students who took both pre- and posttests were included in our analysis. In Fall 2008, the IMCA was administered in a large (n = 284) and a small section (n = 24) of the CU Introduction to Cell and Molecular Biology course (taught by different instructors) and in an introductory biology course at a small liberal arts college (n = 15). In Fall 2009, the IMCA was administered again in the large (n = 328) and small sections (n = 20) of the introductory course at

CU. Students took the IMCA on paper during the first day of class before any instruction, for participation points. The pretest was not returned to the students, ensuring that these questions were not available for study during the course. At the end of each course, the identical questions were given to students as part of the graded final exam. Students were given 30 min to complete both the pre- and posttests. Because the pretest, posttest, and normalized learning gain results from both sections and from both years (e.g., 2008 and 2009) were statistically equivalent, the data were combined for analysis.

Students in another large introductory biology course at a different public research university (n = 666) also took both the pre- and posttests. However, due to time constraints and instructor preference, the questions were answered online in a timed format for participation points at the start of the course and then again before the final exam. Because of differences in how the posttest was administered and the potential impact on level of student effort (discussed in *Results* and *Discussion*), these data were used for comparison but were not included in the statistical analyses. At all institutions, the mean pretest scores, posttest scores, and normalized learning gains, defined as $<g> = 100 \times$ (posttest score $-$ pretest score)$/(100 -$ pretest score) (Hake, 1998), were calculated only for students who took both the pre- and posttests.

Statistical Characterization

At least three statistical tests are commonly used to evaluate assessment instruments: item difficulty and item discrimination (referred to as "item statistics") as well as reliability (Adams *et al.*, 2006; Lindell *et al.*, 2007). Item difficulty (P) for a question measures the percentage of students who answer the question correctly and is calculated as the total number of correct responses (N1) divided by the total number of responses (N). Thus, a low P value indicates a difficult question. Item discrimination (D) measures how well a question distinguishes between students whose total pre- or post-test scores identify them as generally high performing (top 1/3) or low performing (bottom 1/3). The students are divided into top, middle, and bottom groups based on their total scores for the assessment, and the D value is calculated using the formula $(NH - NL)/(N/3)$, where NH is number of correct responses by the top 33% of students, NL is number of correct responses in the bottom 33% of students, and N is total number of student responses (Doran, 1980). A high D value, therefore, indicates that on average, only strong students answered a question correctly. Together, these two parameters provide an informative comparison of performance on individual items from pre- to post-test, as discussed further below.

For assessments that test a variety of different concepts, the preferred measure for reliability is the coefficient of stability (Adams *et al.*, 2006). To obtain this measure, pretest scores were compared between two consecutive iterations of the course at CU, Fall 2008 (n = 371) and Fall 2009 (n = 403). In addition, χ^2 analysis was used to determine whether the spread of incorrect and correct answers on each item was similar between these two semesters.

Institutional Review Board Protocols

Permissions to use pre- and posttest data and student grades (exempt status: Protocol No. 0108.9) and to conduct student interviews (expedited status: Protocol No. 0603.08) were obtained from the University of Colorado Institutional Review Board.

RESULTS

Lessons Learned from the Development Process

One of the greatest challenges in developing a broad assessment tool such as the IMCA is deciding upon the content to be assessed. Although discussions with faculty were framed around known student misconceptions, there was still debate among faculty about the relative importance of certain topics and what concepts were most important for students to learn (and be assessed on) from an introductory biology course. The first version of the IMCA consisted of 43 questions addressing 15 learning goals. After discussions with the instructors for the introductory biology course as well as those teaching subsequent courses, several learning goals along with the corresponding questions were dropped. For example, faculty agreed that the learning goal stating "Distinguish the roles of the soma and the germ line in the life cycle of a typical multicellular organism" was better addressed in a genetics course. Another learning goal, "Explain how the process of scientific research leads to a 'working understanding' of the world around us," was valued by all faculty but was deemed more appropriate for a free-response assessment.

After selection of topics on which to interview students and write questions, the wording of the question stems for the assessment underwent several rounds of revisions. The interview process with students was vital to the rewording of question stems, and several questions and their answers were improved through direct student feedback. For example, some question stems contained "buzz" words (e.g., never, always) that gave away the answer or jargon that was confusing. In another example, Question 9, which addresses Learning Goal 4 (Compare how the properties of water affect different three-dimensional structures and stabilities of macromolecules, macromolecular assemblies, and lipid membranes) was originally asked without diagrams. Students had trouble interpreting the question stem, and the distracters used jargon like "hydrophobic" and "hydrophilic," which could be memorized but potentially not understood. When students identified these problems, we asked them to suggest alternative wording, as well as asked them for their answers and explanations. Incorporating experts' feedback was also important for further refining questions. For Question 9, one expert commented "The answer to the question depends on how the phospholipid is introduced into the water. Is it layered onto the water or is it injected into a water-filled vessel?" In response to student and expert comments, the question was modified to its current version: "The four diagrams A–D below represent cross sections of spherical structures composed of phospholipids. Which of these structures is most likely to form when a phospholipid is vigorously dispersed in water?" Both experts and students agreed that the revised question was clear and unambiguous.

Table 4. Mean pretest, posttest, and learning gain scores for students, TAs/LAs, and biology faculty experts

	n	Mean pretest (± SE), %	Mean posttest (± SE), %	Mean learning gain[a] (± SE), %
Students	671[b]	42.1 (± 0.6)	70.5 (± 0.6)	50.1 (± 1.0)
	666[c]	44.2 (± 0.6)	57.8 (± 0.7)	22.9 (± 0.9)
TAs/LAs	28	76.2 (± 1.9)[d]	86.7 (± 1.5)[d]	40.6 (± 4.7)
Biology experts	25	NA[e]	95.4 (± 1.4)[f]	NA

[a] Normalized learning gain; see *Methods* for calculation.
[b] Posttest administered in class, as part of final exam.
[c] Posttest administered online for participation points.
[d] TA, teaching assistants; LA, undergraduate learning assistants. The TA/LA group performed significantly better on both the pre- and posttests than students in the course (one-way ANOVA, Tukey's post hoc test, $p < 0.05$).
[e] NA, not applicable.
[f] The biology experts scored significantly higher on the assessment compared with the mean student and TA/LA posttest scores (one-way ANOVA, Tukey's post hoc test, $p < 0.05$).

Pretest and Posttest Scores and Learning Gains

The IMCA was administrated to several groups of students in 2008 and 2009 (see *Methods*). The mean pretest scores, posttest scores, and normalized learning gains for all groups are shown in Table 4. Students in all the groups had comparable pretest scores. Students who took the pre- and posttests in class with the posttest as a graded part of the final exam were given 30 min to take the pre-test, and a comparable time was allotted for these questions as part of the final exam. These students showed a significantly higher mean learning gain (one-way ANOVA, Tukey's post hoc test, $p < 0.05$) than students who took the pre- and posttests online (at a different university) for participation points only. For students who took the pre- and posttests online, mean scores increased with time spent on the assessment (Table 5). Students who spent a relatively short time (<20 min) online had the lowest scores on both pre- and posttests.

At CU, the pre- and posttests were also taken by graduate TAs and undergraduate learning assistants (LAs; Otero *et al.*, 2006), who facilitate small group work in weekly problem-

Table 5. Mean scores of students who took the pretest and posttest online

Time spent (minutes)	Mean pretest[a]	Mean posttest[a]
10	31.5 ± 2.1 (34)	42.3 ± 2.1 (47)
20	40.3 ± 2.7 (256)	57.8 ± 1.2 (227)
30	46.8 ± 2.8 (218)	60.9 ± 1.1 (252)
40	49.2 ± 2.8 (112)	60.6 ± 1.9 (106)
>40	51.0 ± 1.8 (46)	54.2 ± 2.7 (34)

[a] Entries are indicated as average percentage scores ± SE, with number of students in parentheses.
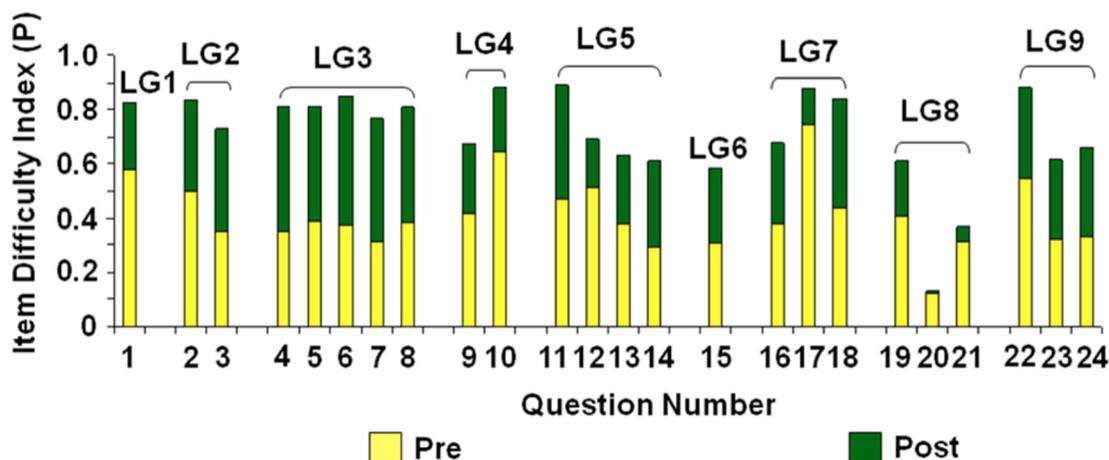
Figure 1. Item difficulty (P) values for each question on the IMCA for Fall 2008 and Fall 2009 pre- and posttests. P values represent percentages of correct answers. Results are based on the 671 students who took the pre- and posttests in class. Yellow bars show the correct answer percentages for each question on the pretest; green bars show the increases in correct answer percentages between pre- and posttest for each question. Questions are grouped according to learning goal (see Table 2).

solving sessions. The TAs (n = 18) and LAs (n = 10), all of whom had previously taken introductory biology courses, performed significantly better on both the pre- and posttests than students in the course (one-way ANOVA, Tukey's post hoc test, $p < 0.05$). The experts (introductory biology teaching faculty; n = 25) performed better than all these groups (one-way ANOVA, Tukey's post hoc test, $p < 0.05$; Table 4).

Descriptive Statistics of Individual Questions Can Be Used to Uncover Student Misconceptions

As described under *Methods*, item difficulty (P) is defined as the fraction of correct answers on a question while item discrimination (D) measures the ability of a question to distinguish between overall high- and low-performing students. As shown in Figure 1, the pre- and posttest P values of individual questions varied, but all questions showed higher values on the posttest than on the pretest.

The posttest P values can be used to identify topics that students continue to struggle with despite instruction. For example, for three of the four questions that address Learning Goal 5 (Given the thermodynamic and kinetic characteristics of a biochemical reaction, predict whether it will proceed spontaneously, and the rate at which it will proceed), student average posttest scores were all below 65%. These questions ask students to analyze the effects of an enzyme on the progress of a reaction (Question 12), the behavior of the reactants at equilibrium (Question 13), and the role of energy released by ATP hydrolysis in an enzyme-catalyzed reaction (Question 14). Students who answered these questions incorrectly generally showed a persistent misconception (e.g., enzymes are required to make chemical reactions happen, rather than simply affecting their rates; Table 2).

For Learning Goal 8 (Using diagrams, demonstrate how the information in a gene is stored, replicated, and transmitted to daughter cells), average posttest scores on the corresponding three questions ranged from 61% down to 13%.

These three questions address the DNA contents of a replicated chromosome in mitosis (Questions 19 and 20) and the mechanism of DNA synthesis (Question 21). For Question 20 in particular ($P = 13\%$), students have retained the misconception displayed in the pretest (individual chromosomes can contain genetic material from both parents; 70% of students answering incorrectly chose this answer on the posttest) suggesting that students are confusing the behavior of chromosomes in mitosis versus meiosis.

The item discrimination value, D, is useful for understanding how performance on an individual item relates to overall performance on the entire assessment. Figure 2 shows the D values for each question on both the pre- and posttests. For some questions, the D values are lower on the posttest than on the pretest, indicating that the question no longer discriminates as strongly between generally high- and low-performing students. For others, the D value stays the same or increases, indicating that the question continues to discriminate between students at different levels. In looking at both the P and D values for individual questions, several stand out as both having a relatively low P value and a high D value on the posttest (Questions 13, 14, 15, 19, 21, 23, and 24). These are relatively difficult questions (some already described above) that only overall high-performing students answer correctly. We discuss the interpretation of these results further in the *Discussion* section.

Reliability

Reliability of the IMCA was measured in two ways. Comparison of pretest scores from the CU Fall 2008 and Fall 2009 courses as described in *Methods* gave a mean coefficient of stability of 0.97, providing evidence that the IMCA is reliable. As a more rigorous test of reliability, the distributions of correct answers and individual incorrect choices on the pretests from both years were compared using χ^2 analysis, and no significant differences in student preferences for particular answer choices were found for any of the questions ($p > 0.05$).
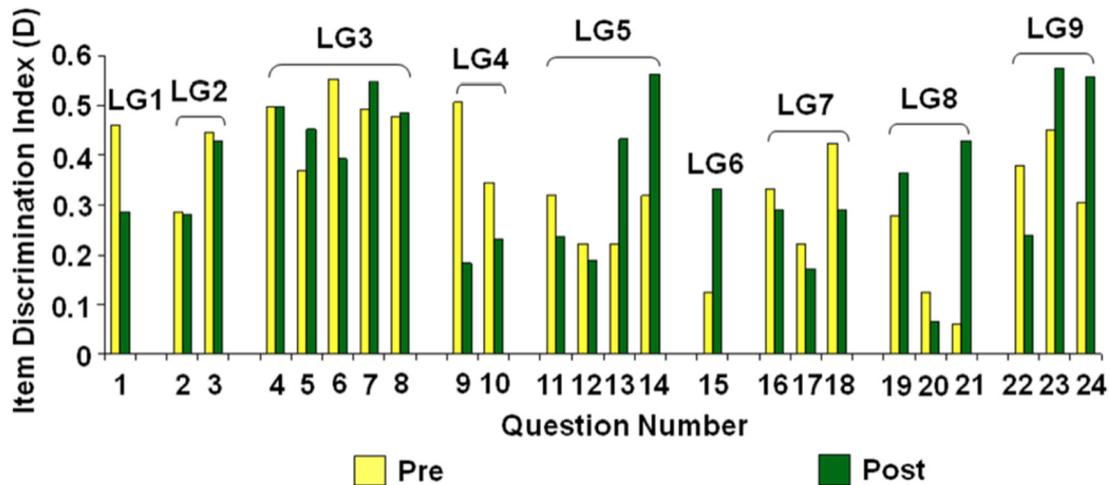
Figure 2. Item discrimination (D) values for questions on the IMCA for the 671 students who took the pre- and posttests in class. Questions with substantial increases in P value (Figure 1) and low D values on the posttest (most students answered correctly) correspond to concepts on which most students gained understanding during the course. Questions with high D values on both the pre- and posttests correspond to concepts that primarily only the stronger students understood at both the beginning and end of the course (see *Discussion* for further explanation). Questions are grouped according to learning goal (see Table 2).

DISCUSSION

We have developed an IMCA for gauging student understanding of these areas at the introductory undergraduate level. The assessment tests primarily conceptual understanding, but as explained in *Methods*, we have also included a few questions that test basic biological knowledge to increase its diagnostic utility. We collected evidence of validity for the IMCA through student interviews, expert reviews (Table 3), and statistical analyses (Figures 1 and 2). Student incorrect answers on the IMCA identify topics with which students typically struggle (Table 2), supporting its value as a diagnostic instrument.

Students in introductory molecular and cell biology courses at three different institutions began with similar distributions of pretest scores, indicating similar levels of knowledge and conceptual understanding at the outset. All students showed normalized learning gains at the end of their courses, but those who took the posttest as part of their final exam scored much higher, and thus showed higher normalized learning gains, than those who took the posttest online for participation credit only (Table 4). We suggest that the substantial difference in posttest scores between these populations of students is due primarily to differences in administration that affected both time spent on the assessment and motivation to answer the questions correctly. In an exam situation, students are motivated to answer questions to the best of their ability, while in an online, ungraded format, students may spend too little time, answer randomly, or put little effort into the assessment. We cannot separate out the effects of these two factors, but it is interesting to note that students who spent a very short time taking the posttest scored the lowest (Table 5). These findings do not preclude administering the IMCA in a nontesting situation but rather underscore the importance of comparing the results of

concept assessments only when they have been administered in an identical manner.

In addition to measuring overall learning gains, data from the IMCA can be used by instructors to provide information about concepts students learn well and those with which they continue to struggle. To make this comparison, one can analyze the combination of P and D values on a question-by-question basis (Figures 1 and 2). For example, Questions 23 and 24, which test understanding of concepts under Learning Goal 9 (gene expression; see Table 2), show some increase in P values but had high D values on both pre- and posttests. This result indicates that some of the stronger students understood these concepts coming into the course, and on average, it was still only the stronger students who understood them after instruction. Several other questions (e.g., 15 and 21 which address Learning Goals 6, membrane transport, and 8, DNA replication, respectively; see Table 2) had low pretest P and D values, indicating that few students understood the concepts coming in, and those who did were not necessarily overall high-performing students (Figures 1 and 2). For these questions, both P and D values increased, showing that although more students could answer these questions correctly on the posttest, only the overall stronger students were doing so. Still other questions had high pretest but low posttest D values, along with a substantial increase in P values, indicating that most students increased their understanding of the corresponding concepts during the course (e.g., Questions 9 and 10, which address Learning Goal 4, see Table 2).

Another way that instructors may extract valuable information from administration of the IMCA is by looking at the proportion of students who chose particular distracters for individual questions. The distracters on the IMCA were generated from student interviews to capture ideas likely to be prevalent in the student population. For example, Question 23, which addresses Learning Goal 9 (gene expression: see Table 2) asks:

The human hexokinase enzyme has the same function as a bacterial hexokinase enzyme but is somewhat different in its amino acid sequence. You have obtained a mutant bacterial strain in which the gene for hexokinase and its promoter are missing. If you introduce into your mutant strain a DNA plasmid engineered to contain the coding sequence of the human hexokinase gene, driven by the normal bacterial promoter, the resulting bacteria will now produce:

a) the bacterial form of hexokinase.
b) the human form of hexokinase.
c) a hybrid enzyme that is partly human, partly bacterial.
d) both forms of the enzyme.

The choice "c) a hybrid enzyme that is partly human, partly bacterial" was the most common wrong answer on the pretest (42% of those who answered incorrectly, and on the posttest 53% of those who answered incorrectly). In addition, almost half of the student answers to this question during the interview process also demonstrated that students believed the promoter and the gene would ultimately both be translated into a hybrid protein. For example, one student responded: "I don't know the process but a hybrid protein including the human part that is from the introduced human gene will be produced," and another: "If a bacterial promoter can bind to the human gene, a hybrid enzyme will be produced." These data suggest that even after instruction, some students still believe that the promoter region will become part of the transcript and be translated to produce part of the resulting protein.

Because the distracters on the IMCA are based on student responses in which their reasoning was explored (as described above), students who select a specific distracter on the pre- or posttest are likely thinking in similar ways to the students who were interviewed. Thus, student choices on the IMCA can reveal themes of student thinking for large numbers of students. The concepts with which students struggle, and the misconceptions they have, can then be used to help instructors design in-class activities, clicker questions, or homework questions that may help students achieve a change in their conceptual understanding.

Administration and Dissemination

From the results reported here, as well as elsewhere (Smith *et al.*, 2008; Adams and Wieman, 2010), we suggest that giving the IMCA as an in-class pretest at the start of the course, and as a posttest included on the final exam will maximize the number of participating students as well as the effort they put forth on the posttest. When students took the postassessment online for participation points only, the wide variability in time spent on the assessment and the overall low posttest scores suggest that many students did not take the assessment seriously. Alternative approaches, such as offering the posttest in class shortly before the final exam, informing students that their performance on the assessment will help shape the final exam review session, have proven reliable (W. K. Adams, personal communication). The most important consideration for administration seems to be that some motivation exists for students to do their best. Reliable comparisons can be made between courses or different populations of students only when assessments are administered in the same way.

Although individual questions from the IMCA could theoretically be used as in-class concept questions, or even homework questions, we encourage users to maintain the integrity of the instrument by administering it only in-class, in its entirety. The validation of the IMCA as a 24-item assessment holds only if it is used in this format. Instructors can choose to use only the questions that pertain to their own courses, or combine these questions with questions from other concept assessments, but the resulting modified assessment should be revalidated. The full set of IMCA questions is included in the Supplemental Material, and the key is available from the authors. We encourage instructors interested in using the IMCA to contact J.S. (jia.shi@colorado.edu) or J.K.K. (knight@colorado.edu) with questions or requests for additional information. We would also appreciate feedback on results obtained by users and will consider user suggestions for future versions of the assessment.

REFERENCES

Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., and Wieman, C. E. (2006). A new instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. Phys. Rev. Spec. Top. Phys. Educ. Res. 2, 010101

Adams, W. K., and Wieman, C. (2010). Development and validation of instruments to measure learning of expert-like thinking. Int. J. Sci. Educ. in press.

Anderson, D. L., Fisher, K. M., and Norman, G. J. (2002). Development and evaluation of the Conceptual Inventory of Natural Selection. J. Res. Sci. Teach. *39*, 952–978.

Balci, S., Cakiroglu, J., and Tekkaya, C. (2006). Engagement, exploration, explanation, extension, and evaluation (5E) learning cycle and conceptual change text as learning tools. Biochem. Mol. Biol. Educ. *34*, 199–203.

Barak, J., Sheva, B., Gorodetsky, M., and Gurion, B. (1999). As "process" as it can get: students' understanding of biological processes. Int. J. Sci. Educ. *21*, 1281–1292.

Bloom, B. S., Engelhart, M. D., Furst, F. J., Hill, W. H., and Krathwohl, D. R. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. New York, Toronto: Longmans, Green.

Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., Moskalik, C. L., and Huether, C. A. (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. Genetics *178*, 15–22.

Brickman, P., Gormally, C., Armstrong, N., and Hallar, B. (2009). Effects of inquiry-based learning on students' science literacy and confidence. Int. J. Schol. Teach. Learn. *3*, 1–22.

Crouch, C. H., and Mazur, E. (2001). Peer instruction: ten years of experience and results. Am. J. Phys. *69*, 970–977.

D'Avanzo, C. (2008). Biology concept inventories: overview, status, and next steps. Bioscience *58*, 1079–1085.

Ding, L., Chabay, R., Sherwood, B., and Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: brief electricity and magnetism assessment. Phys. Rev. Spec. Top. Phys. Educ. Res. *2*, 010105

Doran, R. (1980). Basic Measurement and Evaluation of Science Instruction, Washington, DC: National Science Teachers Association.

Ebert-May, D., Batzli, J., and Lim, H. (2003). Disciplinary research strategies for assessment of learning. Bioscience *53*, 1221–1228.

Garvin-Doxas, K., Klymkowsky, M., and Elrod, S. (2007). Building, using, and maximizing the impact of concept inventories in the biological sciences: report on a National Science Foundation sponsored conference on the construction of concept inventories in the biological sciences. CBE Life Sci. Educ. *6*, 277–282.

Garvin-Doxas, K., and Klymkowsky, M. W. (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). CBE Life Sci. Educ. *7*, 227–233.

Gonzalez-Cruz, J., Rodriguez-Sotres, R., and Rodriguez-Penagos, M. (2003). On the convenience of using a computer simulation to teach enzyme kinetics to undergraduate students with biological chemistry-related curricula. Biochem. Mol. Biol. Educ. *31*, 93–101.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. Am. J. Phys. *66*, 64–74.

Haslam, F., and Treagust, D. F. (1987). Diagnosing secondary students misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument. J. Biol. Educ. *21*, 203–211.

Hestenes, D. (1992). Force concept inventory. Phys. Teach. *30*, 141–158.

Howitt, S., Anderson, T., Costa, M., Hamilton, S., and Wright, T. (2008). A concept inventory for molecular life sciences: how will it help your teaching practice? Aust. Biochem. *39*, 14–17.

Knight, J. K. (2010). Biology concept assessment tools: design and use. Microbiol. Australia, March, 5–8.

Knight, J. K., and Wood, W. B. (2005). Teaching more by lecturing less. Cell Biol. Educ. *4*, 298–310.

Libarkin, J. (2008). Concept inventories in higher education science. Prepared for the National Research Council Promising Practices in Undergraduate STEM Education Workshop 2, (Washington, DC). www7.nationalacademies.org/bose/PP_Commissioned_Papers.html

Lindell, R., Peak, E., and Foster, T. M. (2007). Are they all created equal? A comparison of different concept inventory development methodologies. 2006 Physics Education Research Conference *883*, 14–17.

Marbach-Ad, G., and Stavy, R. (2000). Students' cellular and molecular explanations of genetic phenomena. J. Biol. Educ. *34*, 200–205.

Marbach-Ad, G., *et al*. (2009). Assessing student understanding of host pathogen interactions using a concept inventory. J. Microbiol. Biol. Educ. *10*, 43–50.

Meir, E., Perry, J., Stal, D., Maruca, S., and Klopfer, E. (2005). How effective are simulated molecular-level experiments for teaching diffusion and osmosis? Cell Biol. Educ. *4*, 235–248.

Odom, A. L. (1995). Secondary & college biology students' misconceptions about diffusion & osmosis. Am. Biol. Teach. *57*, 409–415.

Otero, V., Finkelstein, N., McCray, R., and Pollock, S. (2006). Who is responsible for preparing science teachers? Science *313*, 445–446.

Smith, M., K, Wood, W. B., and Knight, J. K. (2008). The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. CBE Life Sci. Educ. *7*, 422–430.

Tanner, K. D. (2010). Order matters: using the 5E model to align teaching with how people learn. CBE Life Sci. Educ. *9*, 159–164

Tanner, K., and Allen, D. (2005). Approaches to biology teaching and learning: understanding the wrong answers—teaching toward conceptual change. Cell Biol. Educ. *4*, 112–117.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. Int. J. Sci. Educ. *10*, 159–169.

Wandersee, J. H. (1985). Can the history of science help science educators anticipate students' misconceptions? J. Res. Sci. Teach. *23*, 581–597.

Wilson, C. D., Anderson, C. W., Heidemann, M., Merrill, J. E., Merritt, B. W., Richmond, G., Sibley, D. F., and Parker, J. M. (2007). Assessing students' ability to trace matter in dynamic systems in cell biology. CBE Life Sci. Educ. *5*, 323–331.