

## Article

# Development of a Meiosis Concept Inventory

Pamela Kalas,<sup>\*,†</sup> Angie O'Neill,<sup>\*</sup> Carol Pollock,<sup>\*</sup> and Gülnur Birol<sup>‡</sup>

<sup>\*</sup>Department of Zoology and <sup>†</sup>Department of Botany, <sup>‡</sup>Science Centre for Learning and Teaching, Faculty of Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Submitted October 17, 2012; Revised July 30, 2013; Accepted August 12, 2013

Monitoring Editor: Clarissa Ann Dirks

We have designed, developed, and validated a 17-question Meiosis Concept Inventory (Meiosis CI) to diagnose student misconceptions on meiosis, which is a fundamental concept in genetics. We targeted large introductory biology and genetics courses and used published methodology for question development, which included the validation of questions by student interviews ( $n = 28$ ), in-class testing of the questions by students ( $n = 193$ ), and expert ( $n = 8$ ) consensus on the correct answers. Our item analysis showed that the questions' difficulty and discrimination indices were in agreement with published recommended standards and discriminated effectively between high- and low-scoring students. We foresee other institutions using the Meiosis CI as both a diagnostic tool and an instrument to assess teaching effectiveness and student progress, and invite instructors to visit <http://q4b.biology.ubc.ca> for more information.

## INTRODUCTION

As educators, we strive to help our students work toward truly mastering a subject, facilitating their transition from novice to expert-like thinking (Adams and Wieman, 2010). To do this, we must first be able to assess student understanding, to identify and address student misconceptions. One tool commonly used for assessing student understanding is a concept inventory (Howitt *et al.*, 2008; Libarkin, 2008; Knight, 2010). Concept inventories are multiple-choice (MC) assessment tools in which each response option represents an alternate conception held by students (Garvin-Doxas *et al.*, 2007; D'Avanzo, 2008). Therefore, they can be used to identify not only how many students in a class have mastered a concept, but also what common misconceptions students may hold (Garvin-Doxas *et al.*, 2007).

Concept inventories are not traditional MC tests—they are designed for formative rather than summative assessment, and the incorrect response options (distracters) for each question are carefully designed and validated (Garvin-Doxas *et al.*,

2007; D'Avanzo, 2008; Adams and Wieman, 2010). The wording of concept inventory questions must be written in language that students will understand (Libarkin, 2008), preferably using wording that students have provided in responses to open-ended questions (Adams and Wieman, 2010). Concept inventories require extensive validation such that we can discern what students are thinking when they select each response option (Howitt *et al.*, 2008; Adams and Wieman, 2010). The validation process is time consuming but essential to confirm that students are selecting a given response option based on their understanding of the concept and not due to a misinterpretation of the question (Adams and Wieman, 2010). Adams and Wieman (2010) recommend that, to ensure consistency with statistical confidence, concept inventories be validated with at least 20–40 student interviews and six to 10 expert interviews. In addition, concept inventories differ from many traditional MC tests in that concept inventories are designed specifically to assess conceptual understanding, rather than memorization of content (Garvin-Doxas *et al.*, 2007).

One application of concept inventories is to rapidly diagnose common misconceptions that students may hold before instruction on a given topic; instructors can then specifically target any identified difficulties while teaching (Libarkin, 2008). Because of their MC format, concept inventory questions are perfectly suited for use with personal response systems (clickers) during class time to provide immediate formative assessment (reviewed in Libarkin, 2008). Many studies have confirmed that having students use clickers to answer MC questions in undergraduate biology classes provides both

DOI: 10.1187/cbe.12-10-0174

Address correspondence to: Pamela Kalas ([kalas@zoology.ubc.ca](mailto:kalas@zoology.ubc.ca)).

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

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

students and instructors with immediate formative feedback on student understanding of key course concepts (Brewer, 2004; Preszler *et al.*, 2007; Crossgrove and Curran, 2008). This feedback is highly beneficial in that it helps instructors to modify the pace of their teaching as appropriate to address areas of difficulty for students, and allows students the opportunity for metacognition and discussion of difficult concepts with their peers (Brewer, 2004).

Another application of concept inventories is to assess the effectiveness of an activity or teaching strategy (Steif and Dantzler, 2005; Howitt *et al.*, 2008; Libarkin, 2008). When compared with traditional lectures, classes that promote active investigation of concepts have been shown to increase student learning (reviewed in Prince, 2004). Courses structured with frequent activities, including in-class exercises that promote active learning, have been shown to be particularly effective in reducing failure rates (Freeman *et al.*, 2011). In light of these results, we should endeavor to improve our teaching by designing activities that we hope will promote increased conceptual understanding of a topic. However, as responsible scientists practicing scientific teaching, we must objectively test any new activity that we develop to ensure that it is actually increasing student learning (Handelsman *et al.*, 2004). Concept inventories can be used in a pre- and posttest format to assess the effectiveness of a given activity or teaching approach in improving student understanding or dispelling a given misconception (Steif and Dantzler, 2005; Howitt *et al.*, 2008; Libarkin, 2008; Knight, 2010).

Concept inventories have already been developed for several areas of biology, including natural selection (Anderson *et al.*, 2002), energy and matter (Wilson *et al.*, 2006), introductory biology (Garvin-Doxas *et al.*, 2007), genetics (Bowling *et al.*, 2008; Smith *et al.*, 2008), molecular and cell biology (Shi *et al.*, 2010), molecular life sciences (Howitt *et al.*, 2008), host-pathogen interactions (Marbach-Ad *et al.*, 2009), and diffusion and osmosis (Odom and Barrow, 1995). Currently, to our knowledge, only two of these published concept inventories in biology have been validated using the exacting methodology recommended by Adams and Wieman (2010): molecular and cell biology (Shi *et al.*, 2010) and genetics (Genetics Concept Assessment [GCA]; Smith *et al.*, 2008). There are currently no published concept inventories for meiosis; the GCA (Smith *et al.*, 2008) does include some questions on meiosis, but the focus of these questions is on the outcomes, rather than on the process of meiosis. Our goal was to use the stringent methodology recommended by Adams and Wieman (2010) to develop a meiosis concept inventory (Meiosis CI) to investigate foundational misconceptions regarding the underlying concepts and actual process of meiosis.

Many students struggle with learning meiosis and hold a wide variety of misconceptions about this process (Brown, 1990; Kindfield, 1991; Dikmenli, 2010). For example, when Dikmenli (2010) asked 124 biology student teachers (who had studied meiosis in previous terms) to draw meiosis, only 13% were able to produce a complete and accurate representation of this process, and 54% produced drawings containing at least one major misconception. Misconceptions about meiosis can stem from an inability to differentiate between sister chromatids and homologous chromosomes, from confusion about the timing of DNA replication during the cell cycle (Dikmenli, 2010), or from an inability to differentiate between replicated chromosomes and the number of DNA molecules present

in a chromosome (Kindfield, 1991). Students also have difficulty calculating probabilities of genotypes that result from particular meiosis or fertilization events, recognizing the arrangement of alleles on chromosomes, and inferring events of nondisjunction (Smith and Knight, 2012). With these misconceptions and difficulties in mind, we designed the Meiosis CI to investigate students' understanding of the subconcepts of ploidy; relationships among amount of DNA, chromosome number, and ploidy; timing of major events during meiosis; and pictorial representation of chromosomes (e.g., sister chromatids vs. homologous chromosomes, general arrangement of chromosomes at metaphase of meiosis I vs. meiosis II).

In our experience, first-year undergraduate students often believe that they already understand meiosis, because they have been introduced to the stages of cell division in grade 9 science (Bourget *et al.*, 2006); however, based on their exam performance, it is apparent that they often lack a deeper understanding of this complicated process. This lack of understanding can negatively impact these students' performance in other areas of biology; students lacking a solid understanding of meiosis often have difficulties understanding genetics and inheritance (Longden, 1982; Brown, 1990; Smith and Kindfield, 1999), as well as the role of meiosis in life cycles.

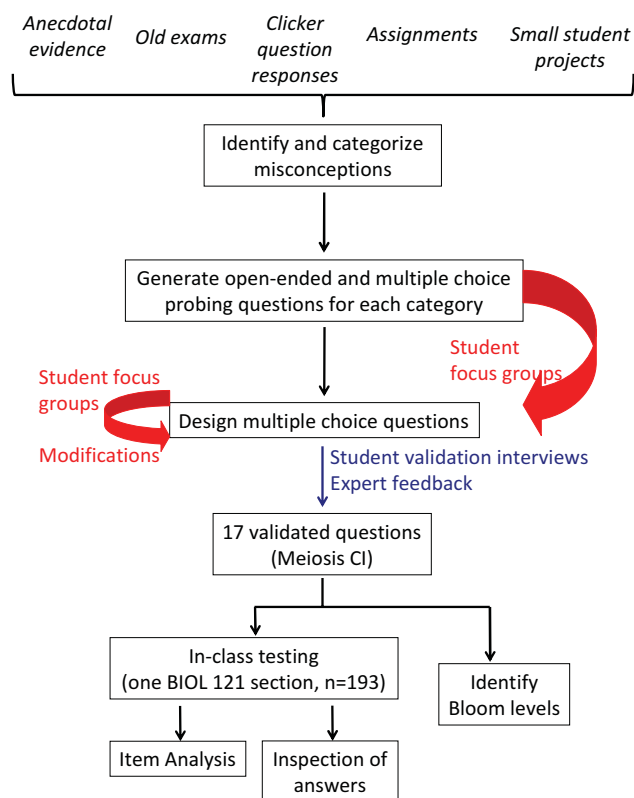
In the University of British Columbia (UBC) Biology Program, students are taught meiosis as an introduction to a unit on genetics in a first-year biology course (BIOL 121). This course serves as a prerequisite for subsequent biology courses, including genetics. A solid understanding of meiosis is an important foundation for genetics throughout the Biology Program. Therefore, we are highly motivated to develop activities that will improve student understanding of meiosis, as well as to evaluate whether these activities actually help increase student understanding. Several strategies already exist to assess student understanding of meiosis, including interviewing individual students orally (Kindfield, 1991; Dikmenli, 2010), analyzing student drawings (Dikmenli, 2010), and analyzing pipe-cleaner models constructed by students (Brown, 1990). However, none of these approaches are feasible for immediate formative assessment in large classes; the Meiosis CI will fill this gap.

This work is part of a larger project at UBC, which is to develop concept inventories for several fields of biology, including prokaryotic cell biology, eukaryotic cell biology, ecology, speciation, and experimental design (<http://q4b.biology.ubc.ca>).

## METHODOLOGY

### *Study Participants*

This study was conducted between January 2010 and November 2011 at UBC, a large research university in Canada. Unless otherwise stated, participants were students registered in BIOL 121, a first-year, multisection lecture-only biology course that introduces ecology, evolution, and genetics; it is a required course in the Biology Program and is also required in programs in other faculties (e.g., forestry), and has an enrollment of ~2000 students per year. This study was conducted using an ethics protocol for human subjects approved by the Behavioural Research Ethics Board at UBC (BREB # H09-03080). Only those students who gave consent



**Figure 1.** Flowchart of the methodology used in the development of the Meiosis CI.

were included in the study. Figure 1 displays a summary of the methodology used in this study.

### Development of the Meiosis CI

**Identification of Misconceptions.** To identify the aspects of meiosis for which first and second-year student thinking tends to be different from expert thinking, we used a collection of student artifacts, such as old examinations, homework, records of student responses to relevant clicker questions in previous terms, and our own experience with students' difficulties and typical mistakes related to meiosis. We categorized such misconceptions and kept track of the number of students demonstrating misconceptions in each category. Misconceptions that occurred only rarely and did not fit into any of the categories were not considered representative and therefore were not included in the concept inventory. The representative misconceptions were grouped into six conceptually distinct categories: ploidy; relationships among DNA, chromosomes, and chromatids; what "counts" as a chromosome; timing of the main events; chromosome segregation; and chromosome movement during gamete formation.

**Design of MC Questions.** We prepared a series of MC and open-ended questions to further probe students' thinking and also to see how the language normally used to describe chromosome movement during meiosis would be interpreted by novices. Whenever possible, the first version of the questions used symbols (e.g., specific letters to represent genes and alleles, particular shapes to represent cells and chromosomes)

and language similar to those used in other genetics concept inventories, such as the GCA (Smith *et al.*, 2008) and the Genetics Literacy Assessment Instrument (Bowling *et al.*, 2008), as we reasoned that these symbols and language, having already gone through rigorous validation, would be particularly well suited for our students. Initially, we administered these questions to small groups of three to five student volunteers (42 students in total) in focus groups. Students were encouraged to verbalize their thinking about both the content and the language used in the questions and were allowed to share ideas with one another. We used information and feedback gathered during these focus groups to craft and refine 20 MC questions for the inventory. Students also provided valuable feedback on how to represent chromosomes and cells so as to be as clear as possible, as well as what organisms to "feature" in our questions so as to minimize confusion.

The inventory contains conventional MC questions, as well as multiple true-false (MTF) questions, for which the participant has to select all the correct answers from a choice of four. MTF questions have received favorable reviews in the literature (e.g., Frisbie, 1992; Haladyna *et al.*, 2002; Libarkin *et al.*, 2011). MTFs have been shown to have higher reliability and to be less susceptible to the effects of student guessing than conventional MC questions in a variety of contexts, including biology examinations (Frisbie and Sweeney, 1982; Kreiter and Frisbie, 1989; Albanese and Sabers, 1988; Downing *et al.*, 1995; Kubinger and Gottschall, 2007). Including MTF questions in a concept inventory provides more information about students' understanding of the concepts under investigation, because the answers a student does not choose are as informative as the answers he or she does choose (Kubinger and Gottschall, 2007; Libarkin *et al.*, 2011; anecdotal observations from our validation interviews). If one wanted to use exclusively single-answer questions, it would be possible to convert the MTF format into a single-answer, complex multiple-choice (CMC) question. We do, however, caution against this practice, as the Meiosis CI questions were not validated in this format. Moreover, CMC-type questions have appeared to be generally less reliable than other multiple-choice formats and more prone to include unintended clues that may help one select the correct answer (Albanese 1993; Haladyna *et al.*, 2002). (See the Supplemental Material for sample Meiosis CI questions.)

**Validation of Multiple-Choice Questions.** After compiling the multiple-choice questions, we individually interviewed 28 students at UBC (11 first-year science students, five non-science students, and nine second-year and three upper-level biology majors) following the protocol described in detail by Adams and Wieman (2010). Each interview lasted approximately 1 h and was audio-recorded. Students were asked to think aloud as they completed the concept inventory, and the interviewer only intervened to encourage students to keep talking. Throughout the process, 13 questions underwent minor modifications in response to students' feedback, and two questions were discarded, because, even after multiple rounds of revision, they were subject to varied interpretations by students.

We also validated the inventory in terms of content with eight experts from UBC with research and teaching experience in genetics and molecular and cell biology. Experts answered Meiosis CI questions and were encouraged to

provide feedback on the clarity and accuracy of both question stems and what we considered the “expert” answers. As a result of this feedback, one question was discarded, as it raised some disagreement among the experts, and the wording of two questions was modified to eliminate inaccuracies. The expert agreement in the final version, which was also validated by student interviews, was 100%.

The final Meiosis CI contains 17 questions, two to four questions per conceptual category. Seven of the questions require the participant to select the single best answer, while 10 require the selection of all the options that are considered correct.

**In-Class Delivery of Meiosis CI.** Validated inventory questions were presented one at a time as PowerPoint slides to students during their second-to-last lecture of the term; students recorded their answers on multiple-choice answer sheets. Administering questions as PowerPoint preserves the integrity of the instrument, decreases use of paper, and ensures that all students see the questions in the same order and proceed through the inventory at the same pace. Each question was presented only once and included instructions about how it should be answered (i.e., “select the best answer” or “select all the answers that apply”), and this information was also provided orally at the start of each question to avoid any possible confusion. Time provided per question (1–2 min) was determined by our validation interviews, and students had the option to request more time if needed. Students received course credit (0.5% bonus mark) for completing the Meiosis CI and, as an additional incentive, had the opportunity to request personalized feedback that would direct them to the subtopics to which they should pay extra attention in preparation for their final examination.

**Item Analysis.** Student responses to each inventory question were scored dichotomously (1 for a completely correct response and 0 otherwise). Opinions on how to best score MTF questions differ, as at least six different methods exist to evaluate these types of items (Tsai and Suen, 1993). The method that we used in our item analysis, the multiple-response (MR) method, whereby no partial credit is given for “partially correct” answers, has been found to lead to the lowest scores (Albanese and Sabers, 1988; Tsai and Suen, 1993). The fact that no partial credit for “partially correct” answers was given could be an issue if concept inventory scores were used to assign grades to students. However, since the inventory is intended to inform teaching and identify student misconceptions, we believe that a more stringent scoring method is appropriate, as it would, if anything, lead the instructor to provide more support or remedial activities than necessary, while the opposite could be detrimental. Moreover, rather than individual student performances, we are primarily interested in the overall class score and, most of all, in the distribution of answers for each question. This can inform us about common misconceptions, their persistence, and gradual shifts toward more expert-like (albeit not necessarily “expert”) combinations of answers. Finally, because each question in the Meiosis CI is designed to test for misconceptions around a specific subconcept, “partial knowledge” indicates that a deep understanding of this subconcept (which is what we are looking for) has not yet been achieved.

Item analysis was used to examine student responses to individual inventory questions in order to assess the qual-

ity of the questions and the quality of the inventory as a whole. Because the inventory was originally developed with the needs of first-year students and instructors in mind, we used results from a section of BIOL 121, in which students ( $n = 193$ ) took the entire 17-question inventory at the end of the course. We decided to conduct the item analysis on posttest results, because, in our experience, students enter first year with extremely limited understanding of meiosis, resulting in very low overall scores for most students and higher frequency of random guessing. Problems may then arise when comparing item analysis results with the recommended standards, which emerged from instruments used as posttests.

For each question, we calculated the index of difficulty (proportion of students who selected an incorrect response), the discrimination index ( $D$ ), the item discrimination efficiency ( $D.E.$ ), and the point-biserial correlation ( $r_{pb}$ ). In educational literature, the index of difficulty is commonly used as an indicator of whether the level of a given question is appropriate for the group of students tested and determines the maximum discrimination potential of the question. For example, a low index of difficulty for a given question implies that most students answer this question as experts, leaving little potential to distinguish between students who perform well (high-performing) and poorly (low-performing) on the entire instrument.  $D$  and  $r_{pb}$  both measure (although in different ways) the correlation between a student’s performance on a given question and his or her performance on the entire test (minus the question of interest), indicating whether this question differentiates well between high-performing and low-performing students. Questions that are part of an instrument assessing competency in a small number of intimately related subtopics are expected to have higher  $D$  and  $r_{pb}$  values than questions in a test that requires multiple disparate skills and competencies. The  $D.E.$  of a question is a measure of its ability to differentiate between high-performing and low-performing students and takes into account the maximum possible discrimination ( $\max(d)$ ) achievable for that question with a given group of students.

To calculate and evaluate the  $D$  of each item we ranked students based on their total score on the concept inventory minus the score of the question of interest, and compared the answers of students in the top 27% with those in the bottom 27% of the ranking for the question of interest.  $D$  was calculated as  $U_{\text{correct}} - L_{\text{correct}}$ , where  $U_{\text{correct}}$  represents the proportion of students in the top 27% of the class who selected the correct answer, and  $L_{\text{correct}}$  is the proportion of students in the bottom 27% of the class who selected the correct answer (Findley, 1956). In addition, for each question, we calculated the 95% Clopper-Pearson’s intervals of the proportions of students who answered correctly in the top and bottom 27% groups.

The  $D.E.$  is calculated as the  $D$  observed in the group of students tested divided by  $\max(d)$ .  $\max(d)$  depends on item difficulty; it is calculated as the  $D$  that would have been obtained for each question if all the students who answered the question correctly were from the top 27% of the class and/or all those who answered the item incorrectly were in the bottom 27%. Point-biserial correlation coefficients ( $r_{pb}$ ) were calculated between the score on each question (0 for incorrect, 1 for correct) and the total score on the remaining 16 inventory questions.

**Table 1.** Summary of the characteristics of the Meiosis CI questions based on a large, first-year biology majors course in which the Meiosis CI was completed at the end of term ( $n = 193$ )<sup>a</sup>

	Concepts tested	Question	Bloom level	Difficulty index	<i>D</i>	<i>D.E.</i>	Point-biserial- ( <i>df</i> = 191)
A	Ploidy; differences between chromosomes, chromatids, and homologous pairs; + indicates question concerning what “counts” as a chromosome (see also concept C); ^ indicates question concerning chromosomal representation of genotypes (see also concept F)	1 <sup>b</sup>	II	0.37	0.38	0.50	0.17*
		2 <sup>+,b</sup>	III	0.30	0.52	0.87	0.41**
		3 <sup>b</sup>	III	0.35	0.75	0.95	0.49**
		4 <sup>c</sup>	III	0.54	0.77	0.83	0.57**
		5 <sup>c</sup>	III	0.48	0.81	0.84	0.59**
		6 <sup>^,c</sup>	IV	0.63	0.73	0.86	0.48**
B	Relationships between chromosomes, DNA, and chromatids, and relation to DNA replication	7 <sup>c</sup>	III	0.45	0.90	0.96	0.61**
		8 <sup>c</sup>	II	0.66	0.79	0.91	0.57**
		9 <sup>c</sup>	III	0.38	0.44	0.56	0.25**
		10 <sup>c</sup>	III	0.23	0.35	0.81	0.23*
C	What “counts” as a chromosome	11 <sup>c</sup>	III	0.25	0.44	0.79	0.42**
		12 <sup>b</sup>	III	0.09	0.21	1	0.23*
D	Timing of events during meiosis	13 <sup>b</sup>	II	0.53	0.38	0.42	0.21*
		14 <sup>c</sup>	I	0.71	0.62	1	0.46**
E	Segregation, chromosome arrangements, and consequences of crossing over	15 <sup>b</sup>	II	0.30	0.27	0.54	0.18*
		16 <sup>b</sup>	II	0.22	0.33	0.81	0.23*
F	Gamete formation; chromosomal representation of genotypes	17 <sup>c</sup>	IV	0.39	0.60	0.69	0.36**

\* $p < 0.05$ .\*\* $p \leq 0.01$ .<sup>a</sup>Difficulty index is the proportion of students who answered the question incorrectly in a posttest *D*, *D.E.*, and  $r_{pb}$  were calculated as described in *Methodology*.<sup>b</sup>Single-answer questions that have a total of four possible choices.<sup>c</sup>Questions for which students are required to check all the answers that apply out of the four that are provided.

**Blooming the Meiosis CI Questions.** To identify the cognitive skill levels of questions, we used the Blooming Biology Tool (Crowe *et al.*, 2008) as a rubric to rate each question based on the six levels of the cognitive domain of Bloom’s taxonomy (Bloom, 1956). The cognitive skill level of each inventory question was determined by consensus of two independent, experienced raters using methodology described elsewhere (Crowe *et al.*, 2008; O’Neill *et al.*, 2010).

## RESULTS

### Meiosis CI

We developed and validated a 17-question concept inventory focused on aspects of meiosis that are a source of confusion and misconceptions for first-year students. The difficulty index and discriminating efficiency of each of these questions, along with the cognitive skill levels, are summarized in Table 1.

The difficulty index of four of the questions was below 0.30, eight ranged from 0.30 to 0.50, and four were above 0.50 but below 0.70, while one question had a difficulty index of 0.71 (Table 1). This suggests that the inventory covers an appropriate spectrum of difficulty and that most questions have a good discrimination potential (items with a difficulty index of 0.50 have the highest discrimination potential). The low-difficulty items can serve as useful “internal controls” to identify students who may be guessing and/or using the

answer sheets incorrectly and students who may be having very serious difficulties with the topic of meiosis. The high-difficulty questions, on the other hand, ensure that there is no ceiling effect (i.e., the tool can capture learning gain even among high-performing students).

The *D* and *D.E.* achieved with our student sample were satisfactory. The mean *D* was 0.55 (minimum 0.21, maximum 0.90) and the *D.E.* was above 0.40 for all questions (minimum 0.42, maximum 1.0). In all cases, the 95% Clopper-Pearson’s intervals of the proportion of students who selected the correct answer in the top 27% and bottom 27% of the class did not overlap, further confirming that the inventory questions discriminate effectively between high- and low-performing students. The mean point-biserial correlation across the 17 questions was 0.38 (minimum 0.17, maximum 0.61) and above 0.20 for all but two questions. In addition, the internal reliability as estimated by Cronbach’s alpha was 0.78.

Nine of the Meiosis CI questions had a cognitive skill level of application (Bloom’s level III), while five were comprehension (II) questions, and one was a knowledge (I) question (Table 1). The highest cognitive skill level tested by the Meiosis CI is analysis (IV), with two questions. Both analysis and application questions are often considered to require higher-order cognitive skills such as problem solving (Fuller, 1997; O’Neill *et al.*, 2010). Therefore, the majority (65%) of the Meiosis CI questions require some element of critical thinking, while only 35% of these questions rely solely on lower-order cognitive skills, such as recalling or explaining memorized information.

**Table 2.** Grouping of Meiosis CI questions into three different categories based on the distribution of answers obtained in a large, first-year biology majors course in which the Meiosis CI was completed at the end of term ( $n = 193$ )

Category	Frequent misconceptions (question, frequent incorrect answer, % of students) <sup>a</sup>
1. Questions with high proportion of correct answers: 2 <sup>b</sup> , 9 <sup>c</sup> , 10 <sup>c</sup> , 11 <sup>c</sup> , 12 <sup>b</sup> , 15 <sup>b</sup> , 16 <sup>b</sup> , 17 <sup>c</sup>	No frequent incorrect answers; these questions had a low difficulty index (posttest).
2. Questions with one common incorrect answer: 1 <sup>b</sup> , 3 <sup>b</sup> , 4 <sup>c</sup> , 5 <sup>c</sup> , 6 <sup>c</sup> , 13 <sup>b</sup>	The absolute number of a cell's chromosomes determines ploidy (Q.1, 31%). Haploid cells are characterized by what an expert would call "unreplicated chromosomes" (Q.4, 23%), diploid cells by what an expert would call "replicated chromosomes" (Q.5, 20%; Q.6, 34%). Normal/real chromosomes are composed of sister chromatids (Q.6, 25%). DNA replication occurs in prophase I (Q.13, 31%).
3. Questions with broader distribution of incorrect answers: 7 <sup>c</sup> , 8 <sup>c</sup> , 14 <sup>c</sup>	Overall confusion in relation to DNA molecules, homologous chromosomes, and sister chromatids (Q.7, various answers). DNA replication results in doubling of chromosome number (Q.8, 40%). DNA replication occurs at the start of meiosis, in prophase I (Q.14, 33%).

<sup>a</sup>Compiled from individual validation interviews.

<sup>b</sup>MC-type questions with four response items from which only one may be selected.

<sup>c</sup>MTF-type questions with four response items, from which any number could be selected.

We organized the inventory questions into six groups (A–F) based on the specific concepts they test (Table 1). Most groups contain questions of different cognitive skill levels that, when used in combination, can differentiate among students who have achieved different levels of understanding of that particular concept. For example, all questions in group A explore student understanding of the difference between haploid and diploid cells. While question 1 requires students to demonstrate comprehension of the definition of haploid and diploid cells in relation to chromosome number, questions 4 and 5 require students to apply their understanding of haploid and diploid to diagrammatic representations of chromosomes in cells. As expected, because questions 4 and 5 are very similar to each other in content and format, the difficulty and discrimination indices of these two MTF-type questions are comparable (Table 1). However question 1, which has a lower cognitive skill level (comprehension) than questions 4 and 5 (application) also had lower difficulty and discrimination indices.

Student responses to questions 1, 3, 4, and 5 revealed misconceptions commonly held by students at the end of the course, namely, that a cell's ploidy is defined by its number of chromosomes (question 1, 31% of students) and that the presence or absence of sister chromatids is what differentiates diploid from haploid cells on a diagram (question 3, 34% of students; question 4, 23% of students; question 5, 20% of students; see Table 2). Student responses to questions in group D revealed another common misconception held by the students at the end of the course, namely, that DNA replication occurs during prophase of meiosis I (Table 2).

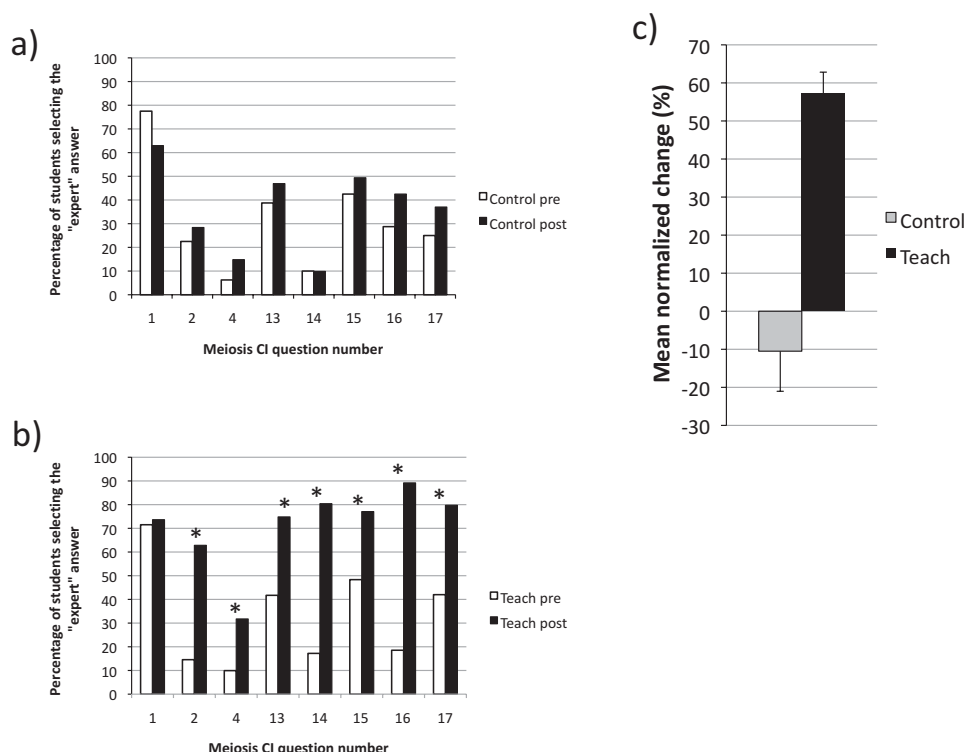
Based on general patterns of student responses to each question of the inventory, three broad categories of questions emerged. The first category includes questions with a high proportion of expert answers (i.e., questions 2, 9–12, and 15–17). The second category, which includes questions 1, 3–6, and 13, is characterized by a high proportion of students who selected a single incorrect answer. Finally, the third category (questions 7, 8, and 14) is composed

of questions that showed broader distributions of answers (Table 2).

### *Adaptations of the Meiosis CI to Inform Teaching*

While we recommend that the entire inventory be administered whenever evaluating teaching effectiveness or assessing students' misconceptions, it is possible to use subsets of questions to monitor student progress or to assess classes for specific misconceptions in a time-effective way and still gain valuable information (Knight, 2010). For example, at our institution, two BIOL 121 instructors used eight Meiosis CI questions to monitor student learning. The instructors administered these questions to their classes before and after the topic of meiosis was taught. These eight questions were specifically selected, because they aligned particularly well with the learning objectives for the unit on meiosis in this course. In the cohort of the 193 students that we used for item analysis, the scores on this subset of eight questions were highly predictive of the scores on the entire inventory ( $r = 0.88$ ).

Inadvertently, one of the BIOL 121 instructors did not teach the unit on meiosis during the time period between the pre- and the posttest. This afforded an opportunity for comparison of the respective performances of students in this section (Figure 2a, Control,  $n = 80$ ) to that of students in a section in which meiosis was taught (Figure 2b, Teach,  $n = 149$ ). There was no statistical difference in the proportion of correct answers on the pre-test between the two sections (Control and Teach), suggesting that students in both sections entered the course with similar levels of understanding of meiosis. Students in the Control section did not demonstrate any significant improvement from the pre- to the posttest, and their mean normalized change,  $\langle c \rangle$  (Marx and Cummings, 2007) was not statistically different from zero (Figure 2c, Control). In contrast, students in the Teach section showed significant improvement in seven of the eight questions (Figure 2b) and a mean normalized change of  $\sim 57\%$  (Figure 2c, Teach).



**Figure 2.** Comparisons between a section of a large first-year biology course in which meiosis was not taught during the time interval between the pre- and the posttest (Control:  $n = 80$ ) and a section in which meiosis was taught (Teach:  $n = 149$ ). In both cases, the pre- and posttests consisted of eight questions from the Meiosis CI, specifically selected because they aligned well with the instructors' objectives for the unit on meiosis. (a) Percentage of students who answered correctly in the pre- and in the posttest respectively, in the Control section. Differences were not statistically significant (chi-square test;  $p > 0.05$ ; Clopper-Pearson: 95% interval of a proportion). (b) Percentage of students who answered correctly in the pre- and in the posttest respectively, in the Teach section. Statistically significant differences (chi-square test;  $p < 0.05$ ; Clopper-Pearson: 95% interval of a proportion) are indicated with an asterisk. (c) Mean normalized change in the Control and Teach sections, calculated as described by Marx and Cummings (2007). The error bar represents the 95% confidence interval of the mean.

In addition, the same subset of questions from the Meiosis CI was used at another local institution (Simon Fraser University in Burnaby, BC). When compared, the mean score on these eight questions on the pretest was in the same range at both institutions (29–36%), and the common misconceptions found at Simon Fraser University were very similar to those found at UBC (J. Sharp, personal communication). This suggests that high school preparation in this topic is quite homogenous among students entering these two institutions.

Another BIOL 121 instructor used the same subset of eight questions to identify frequent misconceptions, such as those related to the concept of ploidy (Figure 3a, Pre). Knowing that students who answer option (a + b) in question 4 think that the absence of sister chromatids identifies haploid cells (see question 4 in the Supplemental Material), the instructor prioritized and specifically targeted the issue of what it means for a cell to be haploid or diploid in terms of its chromosomes' structure. At the end of the unit on genetics, which included meiosis, students took the posttest, and their answers to question 4 revealed that, in spite of the targeted teaching, this misconception related to ploidy was persistent (Figure 3a, Post). This result informed the instructor that this is a persistent misconception and the concept needed to be revisited. In fact, this misconception persists even in third year, with 49% of the students in BIOL 334 (Third-year Basic

Genetics) ( $n = 89$ ) displaying the same misconception as their first-year counterparts regarding what it means for a cell to be haploid. This result prompted the instructor of BIOL 334 to bring the issue to the attention of the students and make some adjustments to the course. In the posttest, this pernicious misconception was still present, although at a lower frequency (Figure 3b, Post), and students still unclear on the concept of ploidy were directed to additional remedial resources.

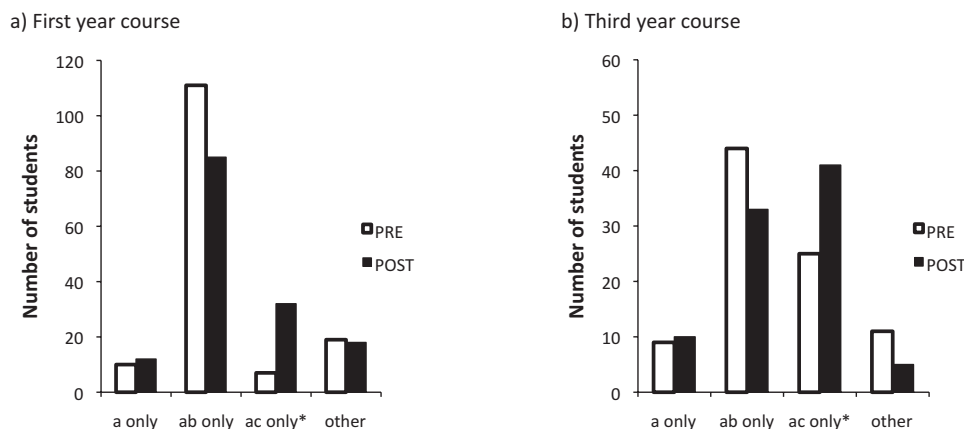
## DISCUSSION

### *Characteristics of the Meiosis CI*

The Meiosis CI was developed and validated using the methods of Garvin-Doxas *et al.* (2007), D'Avanzo (2008), and Adams and Wieman (2010). The 28 student interviews we conducted ensured the students understood the wording of the questions and were clear on the concept being tested in each question. We also had eight expert interviews to confirm the expert response.

The Meiosis CI questions' index of difficulty ranged from 0.09 to 0.71, with the majority between 0.30 and 0.50, which is within the recommended range from 0.30 to 0.90 (Ding *et al.*, 2006) or 0.30–0.70 (Craighead and Nemeroff, 2001). The average index of difficulty was 0.55, close to the optimum





**Figure 3.** (a) Distribution of answers to question 4 of the Meiosis CI in a first-year biology class ( $n = 148$ ) before meiosis was introduced (Pre) and 3 wk after the concept was taught in class (Post). The difference between the two distributions is highly significant (chi-square = 27.9;  $p < 0.001$ ). (b) Distribution of answers to item number 4 in a third-year genetics class (Biol 334) ( $n = 89$ ) on the first day of class (Pre) and 1 wk after meiosis was revisited (Post). The difference between the two distributions is highly significant (chi-square = 17.2;  $p = 0.001$ ). The expert answer is marked with an asterisk.

of 0.50 (Craighead and Nemeroff, 2001). Questions with low index of difficulty were included, as they allowed us to confirm that students were reading the questions carefully and answering to the best of their ability. In addition, 15 of the questions had a  $D$  above the minimum recommended value of 0.30 (Ding *et al.*, 2006). The point-biserial correlation ( $r_{pb}$ ) was above 0.20 for 15 of the 17 questions, and the mean  $r_{pb}$  was 0.38, suggesting that the inventory provides adequate discrimination power (Ding *et al.*, 2006).

Concept inventory questions, instead of testing simple recall of memorized information, should ideally assess conceptual understanding, which requires students to use higher-order cognitive skills, such as critical thinking (Garvin-Doxas *et al.*, 2007). Therefore, most concept inventory questions should have a cognitive skill level of application (III) or higher. The majority of the Meiosis CI questions (65%) are application (III) or analysis (IV) questions, suggesting that the inventory does indeed assess the understanding of concepts. The knowledge (I) and comprehension (II) questions are included, because they can be paired with higher-order cognitive skill level questions that address the same concept. For example, question 1 (lower-order cognitive skill level) and question 4 (higher-order cognitive skill level) address the same concept; students perform better on question 1 (see *Results*), indicating that many students can choose a correct definition but cannot apply that definition to a new situation.

### Misconceptions Identified

A number of prominent misconceptions became evident during the student interviews associated with the development of this inventory; these are consistent with the literature (Brown, 1990; Kindfield, 1991; Dikmenli, 2010). A major misconception centered around differentiating between haploid and diploid cells and between sister chromatids and homologous pairs of chromosomes. Both in the initial and in the validation interviews, most students were able to provide a textbook-type definition of “haploid” and “diploid” but could not utilize this definition to accurately identify haploid

and diploid cells when shown diagrams of chromosomes in cells. This misconception was also evident in our cohort of 193 students who took the Meiosis CI at the end of term, and even among students in a third-year genetics course.

A general trend observed during student interviews was that many students who had difficulty with the concept of meiosis demonstrated a general deficit in their ability to connect meiosis to important events in the cell cycle, for example, the process of DNA replication. For many of these students, the common confusion between sister chromatids and homologous chromosomes, and what “counts as a chromosome” seemed to be tightly connected to a deeper confusion about the relationship between a cell’s ploidy and its chromosome number, the relationships among the DNA double helix and chromosome structure, and/or confusion about what happens to chromosomes throughout the various stages of the cell cycle. The prevalence of these student difficulties is the rationale for the large number of inventory questions dealing with ploidy, chromosome structure and number, and DNA. This also suggests that students may be struggling with these concepts at a more basic level and that we are observing students’ difficulty visualizing and internalizing three-dimensional microscopic structures and events in a meaningful manner. If this is the case, more practice with activities and material targeting these difficulties may be helpful in improving student learning in meiosis.

### Further Insights Gained

We found that the focus group interviews during the initial phase of the Meiosis CI development revealed how small differences in the wording of a question can have a large impact on students’ interpretation of that question. For instance, the stem of question 14 was initially: “Which of the following events occur during prophase I?” and several students selected answers that included “DNA replication.” When explaining their reasoning, some of these students gave a perfect description of all the events taking place during interphase of the cell cycle, which does in fact include DNA replication. When probed, they also incorrectly stated that prophase



I occurs before meiosis (instead of it being the first stage of meiosis). However, if we pointed out to them that another name for “prophase I” is “prophase of meiosis I,” almost all of them would accurately interpret the question (i.e., they realized that prophase is part of meiosis) and correct their previous mistake regarding DNA replication. Once the wording was revised, students had no difficulty correctly interpreting the final version of question 14 (“Which of the following events occur during prophase of meiosis I?”) during the validation interviews.

### *Meiosis CI as a Teaching Tool*

At this time, the use of the Meiosis CI is still in its infancy. In our case, instructors used it as a diagnostic tool; this allowed them to target learning activities to suit the specific needs of a class. It is sobering to note that students coming into a third-year genetics course at UBC held on to their persistent misconceptions, especially about ploidy and that, although there was significant improvement between pre- and posttest, the misconception persisted for many students (Figure 3). Tutorial materials are currently being developed in this course to address these misconceptions. In BIOL 121, more time is now scheduled for introducing these concepts, such as ploidy, and the use of clicker questions targeting these concepts allows instructors to confirm that a majority of students understand key issues before moving forward. An unintended benefit of the Meiosis CI is that students taking the pretest may realize they are not sure about the answers (even though they have studied the topic in high school) and may become more attentive in class.

It is important to use scientific teaching in our classes, that is, to test objectively any new teaching activity that we develop to ensure that it is indeed increasing student learning (Handelsman *et al.*, 2004). Although developing and validating concept inventories is a time-consuming and costly undertaking, a validated concept inventory is extremely useful as a diagnostic tool for instructors, as well as for assessing any activity introduced into a course to enhance student learning.

We have described the development and validation of a Meiosis CI, and presented some adaptations of this inventory that were used by instructors to inform their teaching practices. At UBC, instructors are currently using concept inventories to assess the development and introduction of course material, and we foresee other institutions using the Meiosis CI both as a diagnostic and an assessment tool. We invite instructors to visit the UBC Questions for Biology (Q4B) website (<http://q4b.biology.ubc.ca>) or to contact the corresponding author for more information, including access to the entire inventory.

### ACKNOWLEDGMENTS

Funding for this study was provided by the UBC Teaching and Learning Enhancement Fund (TLEF) and a development grant from the UBC Science Center for Learning and Teaching of the Faculty of Science. The authors thank the UBC students who took part in the interviews and who consented to participate in the study, instructors in BIOL 121 and BIOL 334, and members of the Q4B group for useful feedback. The authors also thank the following for their contributions: Dr. Jennifer Klenz, UBC (assistance with the design of inventory questions); Joan Sharp, Simon Fraser University (for sharing the Meiosis CI data from her students); meiosis experts (assistance with

validation of inventory questions); and faculty, teaching assistants, and peer tutors who assisted with the administration of the concept inventory. The authors also thank the two anonymous reviewers for their invaluable feedback on the manuscript.

### REFERENCES

- Adams WK, Wieman CE (2010). Development and validation of instruments to measure learning of expert-like thinking. *Int J Sci Educ* 33, 1–24.
- Albanese MA (1993). Type K and other complex multiple-choice items: an analysis of research and other item properties. *Educ Meas* 12, 28–33.
- Albanese MA, Sabers DL (1988). Multiple true-false items: a study of interim correlations, scoring alternatives, and reliability estimations. *J Educ Meas* 25, 111–123.
- Anderson DL, Fisher KM, Norman GJ (2002). Development and evaluation of the conceptual inventory of natural selection. *J Res Sci Teach* 39, 952–978.
- Bloom BS (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*, New York: McKay.
- Bourget M, Chisholm A, Kermer D, McKinley B (2006). *Science Grade 9: Integrated Resource Package 2006*, Victoria, BC: British Columbia Ministry of Education. [www.bced.gov.bc.ca/irp/welcome.php](http://www.bced.gov.bc.ca/irp/welcome.php) (accessed 21 June 2011).
- Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, Moskalik CL, Huether CA (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics* 178, 15–22.
- Brewer CA (2004). Near real-time assessment of student learning and understanding in biology courses. *Bioscience* 54, 1034–1039.
- Brown CR (1990). Some misconceptions in meiosis shown by students responding to an advanced level practical examination question in biology. *J Biol Educ* 24, 182–186.
- Craighead WE, Nemeroff CB (eds.) (2001). *The Concise Corsini Encyclopedia of Psychology and Behavioral Science*, 3rd ed., New York: Wiley, 498–499.
- Crossgrove K, Curran KL (2008). Using clickers in nonmajors- and majors-level biology courses: student opinion, learning, and long-term retention of course material. *CBE Life Sci Educ* 7, 146–154.
- Crowe A, Dirks C, Wenderoth MP (2008). Biology in Bloom: implementing Bloom’s taxonomy to enhance student learning in biology. *CBE Life Sci Educ* 7, 368–381.
- D’Avanzo C (2008). Biology concept inventories: overview, status, and next steps. *BioScience* 58, 1079–1085.
- Dikmenli M (2010). Misconceptions of cell division held by student teachers in biology: a drawing analysis. *Sci Res Essays* 5, 235–247.
- Ding L, Chabay R, Sherwood B, Beichner R (2006). Evaluating an electricity and magnetism assessment tool: brief electricity and magnetism assessment. *Phys Rev ST Phys Educ Res* 2, 010105.
- Downing SM, Baranowski RA, Grosso LJ, Norcini JJ (1995). Item type and cognitive ability measured: the validity evidence for multiple true-false items in medical specialty certification. *Appl Meas Educ* 8, 187–197.
- Findley WG (1956). A rationale for evaluation of item discrimination statistics. *Educ Psychol Meas* 16, 175–180.
- Freeman S, Haak D, Wenderoth MP (2011). Increased course structure improves performance in introductory biology. *CBE Life Sci Educ* 10, 175–186.
- Frisbie DA (1992). The multiple true-false item format: a status review. *Educ Meas* 11, 21–26.
- Frisbie DA, Sweeney DC (1982). The relative merits of multiple true-false achievement tests. *J Educ Meas* 19, 29–35.

- Fuller D (1997). Critical thinking in undergraduate athletic training education. *J Athl Train* 32, 242–247.
- Garvin-Doxas K, Klymkowsky M, 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.
- Haladyna TM, Downing SM, Rodriguez MC (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Appl Meas Educ* 15, 309–333.
- Handelsman J *et al.* (2004). Scientific teaching. *Science* 304, 521–522.
- Howitt S, Anderson T, Costa M, Hamilton S, Wright T (2008). A concept inventory for molecular life sciences: how will it help your teaching practice? *Austral Biochem* 39, 14–17.
- Kindfield ACH (1991). Confusing chromosome number and structure: a common student error. *J Biol Educ* 25, 193–200.
- Knight JK (2010). Biology concept assessment tools: design and use. *Microbiol Aust* 31, 5–8.
- Kreiter CD, Frisbie DA (1989). Effectiveness of multiple true-false items. *Appl Meas Educ* 2, 207–216.
- Kubinger KD, Gottschall CH (2007). Item difficulty of multiple choice tests dependent on different item response formats—an experiment in fundamental research on psychological assessment. *Psychol Sci* 49, 361–364.
- Libarkin JC (2008). Concept inventories in higher education science. Manuscript prepared for the National Research Council Promising Practices in Undergraduate STEM Education Workshop 2, held 13–14 October 2008, in Washington, DC.
- Libarkin JC, Ward EMG, Anderson SW, Kortemeyer G, Raeburn SP (2011). Revising the geoscience concept inventory: a call to the community. *GSA Today* 21, 26–28.
- Longden B (1982). Genetics—are there inherent learning difficulties? *J Biol Educ* 16, 135–140.
- Marbach-Ad G *et al.* (2009). Assessing student understanding of host pathogen interactions using a concept inventory. *J Microbiol Biol Educ* 10, 43–50.
- Marx JD, Cummings K (2007). Normalized change. *Am J Phys* 75, 87–91.
- Odom AL, Barrow LH (1995). Development and application of a two-tier diagnostic test measuring college biology students’ understanding of diffusion and osmosis after a course of instruction. *J Res Sci Teach* 32, 45–61.
- O’Neill A, Birol G, Pollock C (2010). A report on the implementation of the Blooming Biology Tool: aligning course learning outcomes with assessments and promoting consistency in a large multi-section first-year biology course. *CJSoTL* 1, article 8.
- Preszler RW, Dawe A, Shuster CB, Shuster M (2007). Assessment of the effects of student response systems on student learning and attitudes over a broad range of biology courses. *CBE Life Sci Educ* 6, 29–41.
- Prince M (2004). Does active learning work? A review of the research. *J Eng Educ* 93, 223–231.
- Shi J, Wood WB, Martin JM, Guild NA, Vicens Q, Knight JK (2010). A diagnostic assessment for introductory molecular and cell biology. *CBE Life Sci Educ* 9, 453–461.
- Smith MK, Kindfield AC (1999). Teaching cell division: basics and recommendations. *Am Biol Teach* 61, 366–371.
- Smith MK, Knight JK (2012). Using the Genetics Concept Assessment to document persistent conceptual difficulties in undergraduate genetics courses. *Genetics* 181, 21–32.
- Smith MK, Wood WB, Knight JK (2008). The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. *CBE Life Sci Educ* 7, 422–430.
- Steif PS, Dantzler J (2005). A statics concept inventory: development and psychometric analysis. *J Eng Educ* 33, 363–371.
- Tsai F-J, Suen HK (1993). A brief report on a comparison of six scoring methods for multiple true-false items. *Educ Psychol Meas* 53, 399–404.
- Wilson CD, Anderson CW, Heidemann M, Merrill JE, Merritt BW, Richmond G, Sibley D, Parker JM (2006). Assessing students’ ability to trace matter in dynamic systems in cell biology. *CBE Life Sci Educ* 5, 323–331.