Small Changes: Using Assessment to Direct Instructional Practices in Large-Enrollment Biochemistry Courses

Xiaoying Xu,[†] Jennifer E. Lewis,[†] Jennifer Loertscher,[†*] Vicky Minderhout,[‡] and Heather L. Tienson[§]

[†]Department of Chemistry and Center for Improvement of Teaching and Research on Undergraduate STEM, University of South Florida, Tampa, FL 33620; [‡]Department of Chemistry, Seattle University, Seattle, WA 98122; [§]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095

ABSTRACT

Multiple-choice assessments provide a straightforward way for instructors of large classes to collect data related to student understanding of key concepts at the beginning and end of a course. By tracking student performance over time, instructors receive formative feedback about their teaching and can assess the impact of instructional changes. The evidence of instructional effectiveness can in turn inform future instruction, and vice versa. In this study, we analyzed student responses on an optimized pretest and posttest administered during four different quarters in a large-enrollment biochemistry course. Student performance and the effect of instructional interventions related to three fundamental concepts—hydrogen bonding, bond energy, and pK_a—were analyzed. After instructional interventions, a larger proportion of students demonstrated knowledge of these concepts compared with data collected before instructional interventions. Student responses trended from inconsistent to consistent and from incorrect to correct. The instructional effect was particularly remarkable for the later three quarters related to hydrogen bonding and bond energy. This study supports the use of multiple-choice instruments to assess the effectiveness of instructional interventions, especially in large classes, by providing instructors with quick and reliable feedback on student knowledge of each specific fundamental concept.

INTRODUCTION

Administration of assessment instruments at the beginning and end of a course can provide instructors formative feedback on their students' understanding of specific concepts and how student thinking changes as a result of instruction. By analyzing formative assessment data, instructors learn about their students' thinking and can respond in real time or from term to term with instructional changes (Sadler, 1989; Haudek *et al.*, 2011; Brown *et al.*, 2013; Evans, 2013). Several studies describe use of diagnostic instruments to inform instruction in large-enrollment courses in the molecular life sciences (Smith *et al.*, 2008; Marbach-Ad *et al.*, 2010; Shi *et al.*, 2010; Loertscher *et al.*, 2014b). However, questions remain regarding the validity of inferences made using formative assessment and the impact on student learning, especially when methodological approaches are ill-defined (Dunn and Mulvenon, 2009; Bennett, 2011). Additionally, instruments are not static documents but rather should evolve as new data and analyses provide ongoing insight into how well the instruments function. Discipline-based education researchers can keep up with the best practices according to contemporary standards for educational and psychological testing by contributing evidence to improve existing instruments and thus lay a solid foundation to support effectiveness of ongoing educational innovations (Arjoon *et al.*, 2013; American Educational Research Association, American Psychological Association, and

Mary Lee Ledbetter, *Monitoring Editor*

Submitted June 15, 2016; Revised November 15, 2016; Accepted November 16, 2016

CBE Life Sci Educ March 1, 2017 16:ar7

DOI:10.1187/cbe.16-06-0191

*Address correspondence to: Jennifer Loertscher (loertscher@seattleu.edu).

National Council on Measurement in Education [AERA/APA/NCME], 2014). Responses to items can be influenced by many factors such as wording, distractors, and instruction (Haladyna *et al.*, 2002). The literature has many examples of instrument improvement over time to better measure intended variables in both the cognitive and affective domains (Pintrich *et al.*, 1993; Tan and Treagust, 1999; Tan *et al.*, 2002; Chandrasegaran *et al.*, 2007; Bauer, 2008; Xu and Lewis, 2011).

A 21-question multiple-choice instrument was previously developed to assess students' understanding of foundational concepts from chemistry and biology before and after completing a biochemistry course. The seven concepts included on the instrument were originally chosen in collaboration with a diverse community of biochemistry educators. A group of experts agreed that a firm grasp of these foundational concepts is essential to enable students to build a deep understanding of biochemical concepts (Villafañe *et al.*, 2011a). While this instrument was not designed to comprehensively cover the prerequisite knowledge required for success in biochemistry, it can provide targeted information for instructors wishing to assess student understanding in several important areas. Each concept has a set of items with a parallel structure of response options, meaning each item in the set contains one distractor corresponding to each of three common incorrect ideas about the concept. This structure allows not only for the capture of correct responses across the set of items related to the concept but also for the capture of responses that are incorrect. Patterns of incorrect responses across the set of items can then reveal either a confused understanding or consistent misunderstanding. Reliability and validity evidence has been collected according to the framework suggested by standards for educational and psychological testing (Arjoon *et al.*, 2013; AERA/APA/NCME, 2014). This study aims to continue the revision of the instrument to achieve better quality and capacity to identify student incorrect ideas and to gauge knowledge gains due to instruction from the pretest and posttest responses.

Improving instruction in large-enrollment classes can be challenging, because instructors often lack avenues to gain meaningful insights into student thinking. Classroom layout may further limit instructional options. Techniques such as clickers have been reported to be used by instructors to make necessary adjustment during instruction (Caldwell, 2007; Kay and LeSage, 2009). Open-ended questions and interviews have been used to examine student understanding but require great time and effort (Songer and Mintzes, 1994; Orgill and Sutherland, 2008). Diagnostic assessments can be more feasible and have been used in large college classrooms (Howitt *et al.*, 2008; Tsui and Treagust, 2010; Villafañe *et al.*, 2011a). Here, we describe how one instructor used data from a multiple-choice pre/posttest to implement and assess manageable changes in instructional practices to improve students' understanding of foundational concepts in biochemistry. The process described here could be used as a model by other instructors interested in improving teaching and learning in their classrooms.

As part of this study, we chose to track biochemistry students' understanding of three foundational concepts (hydrogen bonding, bond energy, and pK_a). Undergraduate students in the United States usually encounter these concepts in courses that precede upper-level biochemistry, such as general chemistry, organic chemistry, and/or introductory biology. Understanding of these foundational concepts is essential for students' success in biochemistry. For example, molecular interactions direct a wide range of important biochemical phenomena, including ligand binding (Sears *et al.*, 2007), enzyme–substrate interactions (Bretz and Linenberger, 2012), and macromolecular structure formation. Noncovalent interactions are so important in biochemistry that the physical basis of interactions has been identified as a threshold concept for biochemistry (Loertscher *et al.*, 2014a) and an understanding of macromolecular interactions was identified as a core concept in biochemistry and molecular biology (Tansey *et al.*, 2013). Hydrogen bonding plays an especially important role in biochemistry, because of the ubiquity of water in biological systems and the prevalence of oxygen, nitrogen, and hydrogen in biological molecules. Additionally, without the fundamental concept that energy is always required to break isolated bonds/interactions and is always released when bonds/interactions form, students are unable to consider the complex energy transfer and coupling events that are so prevalent in biochemistry. Finally, because biochemistry occurs in an aqueous environment and many important biological molecules such as amino acids and nucleotides act as acids or bases, it is vital that students understand how to use the pK_a of a molecule to determine the protonation state at a given pH. The charges that arise on these molecules as a result of protonation or deprotonation determine their structure, their interactions, and therefore their function in a biological context.

The goals of this study are twofold. First, we aim to revise an existing multiple-choice instrument to produce results that better identify students' incorrect ideas and knowledge gains. Second, we seek to understand how an instructor of a large-enrollment biochemistry course can use pretest and posttest data to inform instructional changes to better support student learning. We analyzed whether different kinds of instructional interventions, including targeted changes in lecture content, specially designed clicker questions, and in-class activities, affect student pretest and posttest performance. Specifically, we asked, 1) How do students respond to the parallel test items related to the concepts of hydrogen bonding, bond energy, and pK_a, and how can the response patterns be used to identify common incorrect ideas? 2) Can small, but targeted, instructional interventions fill gaps in student knowledge of these basic concepts?

METHODS

Data Collection: Course Characteristics and Student Population

The Instrument of Foundational Concepts for Biochemistry (IFCB) was administered as a pretest and posttest to students enrolled in a biochemistry course at a large public research university in the western United States during four different quarters (Table 1). The course is the first quarter of a three-quarter biochemistry sequence and covers topics related to macromolecular structure formation, enzyme function, and metabolism, including glycolysis, the citric acid cycle, and oxidative phosphorylation. Every quarter, two lecture sections are run identically, each with an enrollment of 200–230 students. The enforced prerequisite for the course is the second quarter of organic chemistry; thus, the students have also taken general chemistry. Approximately 50% of the students are life science majors, 30% are physical science majors, and 20% are

TABLE 1. Summary of data collection conditions (*N* = 1185)

Quarter	Pretest method	Posttest method		*N* (total)
Fall 2014	Paper	Paper	Before the final	198
Winter 2015	Paper	Paper	During the final	381
Spring 2015	Online	Paper	During the final	199
Fall 2015	Online	Paper	During the final	407

transfer students. There is no biology prerequisite. All life science majors are required to take the course, so the students' majors include biology, molecular cell and developmental biology, microbiology and immunology, psychobiology, physiology, and neuroscience, as well as chemical or biomedical engineering, chemistry, and biochemistry. The students are mainly juniors and seniors.

The pretest was administered to students at the beginning of the quarter before any material related to targeted concepts was taught. In the first two quarters (Fall 2014, Winter 2015), the instrument was administered as a paper-and-pencil quiz during class time, and in the latter two quarters (Spring 2015 and Fall 2015), it was administered online. For online administration, students were required to complete the assessment before lectures on targeted concepts started, and they could only open the quiz once. They were given 25 minutes to complete the quiz online, which was the same amount of time allowed for paper-and-pencil administration. The quiz questions and answers were set in the same order as on the paper quiz, and the students could not go back and forth between questions. Students were not able to access the online questions after they had taken the quiz. All data were collected in accordance with approved institutional review board policies and practices.

Characteristics of the Instrument before and after Revision

The IFCB had been previously developed and tested to uncover incorrect ideas that students bring to biochemistry courses from prior chemistry and biology courses (Villafañe *et al.*, 2011a,b). Instrument design and use are described in detail in the two papers by Villafañe and colleagues, and a summary is given here. The instrument is composed of 21 multiple-choice questions (henceforth called items) relating to seven concepts (hydrogen bonding, bond energy, pK_a, equilibrium, free energy, alpha-helical structure, protein function). Each concept is tested by a set of three items, all of which must be answered correctly for the student to demonstrate correct knowledge of that concept. Every item has four response options: one correct and three distractors corresponding to common incorrect ideas. The distractors were designed to follow a parallel structure across the set of items for a given concept. In other words, three common incorrect ideas were identified for each concept, and each item in the set has these three incorrect ideas as distractors, but with different wording appropriate to the context of the specific item and in different order. Statements describing each of the incorrect ideas are provided in Table 2. The items were generated and refined in collaboration with a community of biochemistry faculty members. Common incorrect ideas used as distractors were identified from faculty experience, existing literature, and student responses to pilot questions. Although the published version of the IFCB generated much useful information about student understanding, analysis of results over time revealed that some items were not functioning as expected. In an attempt to improve the usefulness of the IFCB, we revised the originally published version for use in this study. The total number of items and the use of parallel structure across sets of items were the same before and after revision. One concept (London dispersion forces) was dropped from the instrument and another concept (equilibrium) was added. Items about hydrogen bonding were substantially revised due to poor internal consistency when students answered these three items in the previous version (changes described in detail in the *Results* section). One of the items related to protein function was modified. To maintain the security and usefulness of the instrument in ongoing assessment efforts, it is not being included with this publication. Instructors interested in using the instrument are invited to contact the authors, who are committed to providing it in a timely manner.

TABLE 2. Items/codes and corresponding ideas for targeted concepts

Concept		Correct/incorrect ideas
Hydrogen bonding	A	Correct: A hydrogen-bonding interaction is a noncovalent interaction involving a small electronegative atom and a hydrogen covalently bonded to a small electronegative atom.
	B	Incorrect: A hydrogen covalently bonded to a carbon can participate in hydrogen-bonding interactions.
	C	Incorrect: All hydrogens are capable of participating in hydrogen bonds, regardless of their covalent bond participation.
	D	Incorrect: Hydrogen bonding is a covalent bond between a hydrogen and another atom.
Bond energy	A	Correct: When an isolated chemical bond forms, energy is released.
	B	Incorrect: The bond strength determines whether energy is released or absorbed when an isolated chemical bond is formed.
	C	Incorrect: Formation of an isolated chemical bond sometimes requires energy and sometimes releases energy.
	D	Incorrect: Formation of an isolated chemical bond requires energy.
pK_a	A	Correct: When the pH is above the pK_a of the carboxyl group, the predominant charge on the group is negative. When the pH is below the pK_a of the amine group, the predominant charge on the group is positive.
	B	Incorrect: The predominant charge on an ionizable group is unaffected by pH.
	C	Incorrect: When the pH is below the pK_a of the carboxyl group, the predominant charge on the group is negative. When the pH is above the pK_a of the amine group, the predominant charge on the group is positive.
	D	Incorrect: At pH = pK_a there is a predominant charge on the ionizable group in question.

Data Analysis: Knowledge Gain for Each Concept

Student response patterns related to the concepts of hydrogen bonding, bond energy, and pK_a were analyzed for knowledge gains by comparing individual response patterns on the pretest and posttest. These concepts were chosen for analysis because they were of particular interest to the instructor. Recall that each concept maps to three different items and that a student must answer all three items correctly to receive credit for the concept. Each of the three concepts was coded as 1 for correct answers to all three items. A zero was assigned if students answered one, two, or all three of the items related to that concept incorrectly. In the analysis, we found four groups of students: those who *always knew* the concept (answered all three items correctly on the pretest and posttest), *came to know* the concept (answered one or more items incorrectly on the pretest and answered all three items correctly on the posttest), *went backward* (answered all three items correctly on the pretest but answered at least one item incorrectly on the posttest), and *never knew* the concept (answered at least one item incorrectly on both the pretest and posttest). Histographs were used to illustrate the proportion of students in each group for each concept each quarter and whether student knowledge shifted as a result of instruction.

The IFCB's structure is such that one correct and three incorrect ideas for each concept can be tracked across a set of three items relating to a concept (see Table 2). Therefore, to determine common incorrect ideas, the combinations of the four parallel response options across the set of items for each concept were also analyzed. For this analysis, the correct option for each concept was coded as A and the incorrect options were coded as B, C, and D, respectively, as shown in Table 2. Missing responses were coded as N. Thus, students who consistently chose the correct response option across the set of items for a concept were labeled as having the combination AAA. Those who picked specific incorrect ideas consistently would therefore be combination BBB, CCC, or DDD. The combinations of student responses were ranked by occurrence to identify the most popular ones within each concept. The ribbon tool was used to visualize the flow of student responses from pretest to posttest and create Sankey diagrams (Pagliarulo and Molinaro, 2015), which we will call "ribbon graphs." As a point of reference, there are 64 possible combinations of four statements across three items ($4 \times 4 \times 4 = 64$). To show all 64 as separate combinations in a single figure would be a challenge, so the combinations were grouped into meaningful categories. The ribbon graphs therefore show six categories: consistently correct (AAA), at least one correct response in combination with incorrect response(s) (e.g., AAB, ACD, labeled in the graphs as AXX), consistently incorrect (BBB, CCC, or DDD), and inconsistently incorrect (e.g., DBC, CCB, NBN, labeled in the graphs as XXX). The last category includes a missing response (N) only if at least one of the items in the set did have a response from that student.

For pK_a, the molecule embedded in one of the items associated with that concept has both carboxyl and amine groups, so there is no response option reflecting incorrect statement D (at $pH = pK_a$ there is a predominant charge on the ionizable group in question) for this item. Instead, two response options for this item were mapped to statement C. Because of this more complex molecular structure, the category DDD for the pK_a concept is not applicable and does not appear in the ribbon graph.

Analysis of the Reliability and Characteristics of the Revised Instrument

A total of 1185 student responses with both pretest and posttest scores were used in the analysis. Descriptive statistics of mean score for each item were calculated using SAS version 9.3. Internal consistency reliability was calculated by Cronbach's alpha. Each test item was assigned 1 for a correct response and 0 for an incorrect response. A Cronbach's alpha greater than 0.7 is considered satisfactory for research purposes (Murphy and Davidshofer, 2005). Confirmatory factor analysis was performed to estimate the item structure using Mplus version 7.31. Because the instrument has 21 items to measure seven concepts, a seven-factor model was run to examine how this intended model fit the empirical data. To support the interpretation of a total score, we also ran a bifactor model to examine the general construct beyond the seven factors (Xu *et al.*, 2016). Because all measured variables (item scores) were categorical, a means and variance–adjusted weighted least-squares method was applied, using the tetrachoric correlation matrix for the 21 items (Brown, 2015). The model was identified by fixing the first item on the factor at 1. In general, chi-square values from a model based on a large sample size are likely to show a significant lack of model fit. The additional criteria of comparative fit index (CFI) value greater than 0.95, root-mean-square error of approximation (RMSEA) value <0.05, and weighted root-mean-square residual (WRMR) value <1 were used to indicate a good model fit (Bentler, 1990; Hu and Bentler, 1999).

RESULTS

Optimization of the IFCB

Revisions of Hydrogen-Bonding Items. Analysis of student response data for the set of hydrogen-bonding items on the previously published version of the IFCB revealed poor internal consistency reliability. Two of the items asked students to identify hydrogen-bonding interactions using visual cues from molecular structures, but the third asked about hydrogen bonding in methanol without a structure given. Most students correctly answered the methanol question, yet many of these students were unable to correctly answer the other two questions, leading to a relatively low internal consistency across the set of items. Interviews with students indicated that methanol was often memorized as an example of a molecule that could participate in hydrogen bonding. In the current version, the methanol item has been removed, and each hydrogen-bonding item provides students with the visual cue of a molecular structure. The electronegativity values for each element were also provided. In addition, the response options were revised to better reflect each of the corresponding incorrect ideas.

Characteristics of the Revised Instrument. The first goal of this study was to revise an existing multiple-choice instrument to produce results that better identify student incorrect ideas and knowledge gains. For determining whether this goal was met, 1185 student responses to the revised instrument collected over four quarters of the same course were analyzed. We examined Cronbach's alpha to establish the evidence for reliability for each concept. Table 3 shows the alpha range from 0.53 to 0.92 for the revised version of the IFCB. Most values of Cronbach's alpha are above the satisfactory level of 0.7 (Murphy and Davidshofer, 2005). Internal consistency reliability was

TABLE 3. Reliability by Cronbach's alpha

	Previous IFCB[a]	Fall 2014		Winter 2015		Spring 2015		Fall 2015	
Concept	Pre	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Hydrogen bonding	0.27	0.92	0.92	0.90	0.85	0.84	0.90	0.91	0.84
Bond energy	0.77	0.87	0.88	0.89	0.89	0.83	0.53	0.90	0.83
pK_a	0.48	0.71	0.69	0.61	0.75	0.53	0.79	0.76	0.58

[a]Value obtained from the original version results by Villafañe *et al.*, 2011a.

improved, especially for the hydrogen-bonding concept, as compared with the original version of IFCB (Villafañe *et al.*, 2011a). The revised set of pK_a items shows a Cronbach's alpha less than 0.7 half of the time and therefore has room for further improvement.

Fit to a predicted factor model was reported for the previous version of the IFCB. To determine whether these parameters had improved for the revised IFCB, we examined both the original seven-factor model and a new bifactor model for the pretest and posttest data ($N = 1185$). From confirmatory factor analysis (CFA) results (Table 4), both models fit the data well, as indicated by CFI > 0.95, RMSEA < 0.05, and WRMR < 1. For the pretest, the seven-factor model fits better than the bifactor model, with smaller χ^2 and higher CFI values. For the posttest, the bifactor model fits better than the seven-factor model. Both the seven-factor and the bifactor model allow interpretation of a score for each of the seven concepts in the instrument, but the bifactor model also allows interpretation of the overall score on the instrument. It may be that the bifactor model fits better after instruction because students have a better overall understanding of biochemistry prerequisite knowledge after the class, whereas their knowledge was more fragmented (they knew some prerequisite concepts but not the others) before the class. Ultimately, because both models support interpretation of seven factors aligned with the instrument design, it is safe to interpret the score for each concept on both the pretest and the posttest. Factor loadings for all three hydrogen-bonding items are consistently significant and greater than 0.7 (as compared with loadings of 0.3 in the previous version), demonstrating that the revision of the hydrogen-bonding items resulted in an improvement (Fornell and Larcker, 1981).

Student Performance on Targeted Concepts of Hydrogen Bonding, Bond Energy, and pK_a

Table 5 presents how well students responded to each of the three items within the targeted concepts. These concepts were chosen for analysis because they were of particular interest to the instructor and the instructor was interested in tracking student performance related to these concepts. On the pretest, item scores (i.e., the proportion of students getting an item correct) ranged from 0.20 to 0.31 on hydrogen-bonding items, from 0.29 to 0.43 on bond energy items, and from 0.26 to 0.43 on pK_a items. On the posttest, an overall increase in the percentage of students getting an item correct was observed. For example, in Fall 2015, 26% students answered the first hydrogen-bonding item correctly on the pretest, and this value rose to 69% for the posttest.

Student Knowledge Gains: Assessment and Instructional Changes

The second major goal of this study was to understand how an instructor of a large-enrollment biochemistry course can use pretest and posttest data to inform instructional changes to better support student learning. Below we describe one instructor's iterative process of analyzing learning gains for three concepts (hydrogen bonding, bond energy, pK_a) at the end of a term, making changes in instructional practices, and analyzing learning gains in subsequent terms. It is important to recognize that this process differs from typical formative assessment, in which data are used to make changes in real time that affect the students who generated the data. By analyzing trends in student data over a number of quarters, the instructor was engaged in formative assessment to improve instructional practices for successive groups of students.

The parallel structure of the IFCB allowed for pretest identification of specific incorrect ideas related to each concept and posttest determination of whether these problems were corrected by the end of the term. Although the incorrect ideas on the IFCB are known to be problematic for students in general, it was very useful for the instructor in this study to identify the most common incorrect ideas among students at the instructor's own institution. The detailed account of the relationship between assessment data and instruction given in the following sections is intended to provide not only insight into students' understanding of foundational concepts for biochemistry courses but also a model for how instructors of large-enrollment classes can use analysis of student learning gains to improve instruction over time.

Hydrogen Bonding

Many instructors are not familiar with documented misconceptions related to concepts typically covered in prerequisite

TABLE 4. Model fitness from seven-factor and bifactor CFA analysis[a]

Pre or post	Model	*N*	χ^2	*df*	*p* Value	CFI	RMSEA	WRMR
Pretest	Bifactor	1185	260	168	<0.001	0.995	0.02	0.99
	Seven factor	1185	248	168	<0.001	0.996	0.02	0.85
Posttest	Bifactor	1185	212	168	0.01	0.999	0.02	0.86
	Seven factor	1185	254	168	<0.001	0.998	0.02	0.85

[a]Satisfactory model fit can be indicated by CFI > 0.95, RMSEA < 0.05, WRMR < 1.

TABLE 5. Proportion of students providing the correct answer for each listed item

	Fall 2014 (n = 198)		Winter 2015 (n = 381)		Spring 2015 (n = 199)		Fall 2015 (n = 407)	
Concept: Item	**Pretest**	**Posttest**	**Pretest**	**Posttest**	**Pretest**	**Posttest**	**Pretest**	**Posttest**
H bonding: 1	0.30	0.35	0.28	0.82	0.20	0.76	0.26	0.69
H bonding: 2	0.31	0.38	0.29	0.81	0.24	0.79	0.29	0.69
H bonding: 3	0.29	0.38	0.31	0.79	0.22	0.78	0.26	0.69
Bond energy: 1	0.43	0.44	0.35	0.71	0.28	0.64	0.33	0.60
Bond energy: 2	0.40	0.38	0.29	0.66	0.29	0.59	0.30	0.60
Bond energy: 3	0.38	0.41	0.32	0.70	0.27	0.60	0.34	0.62
pK_a: 1	0.41	0.80	0.41	0.83	0.32	0.78	0.31	0.78
pK_a: 2	0.31	0.76	0.33	0.80	0.26	0.73	0.29	0.76
pK_a: 3	0.36	0.80	0.39	0.81	0.42	0.81	0.43	0.76

courses. Before first administration of the IFCB in Fall 2014, the instructor was unaware of common incorrect ideas related to hydrogen bonding and assumed that students entering biochemistry either had a firm understanding of the concept or would gain an understanding over the course of the term via regular instruction. Historically, in this biochemistry course, students were exposed to hydrogen bonding during lectures in the first week of class, in which the physical basis of noncovalent interactions is taught. Instruction included three slides defining hydrogen-bonding interactions followed by a clicker question (shown in the Supplemental Document). The clicker question asked students to identify how many hydrogen bonds could be donated or accepted by urea, and it was followed by a discussion of how students came to their conclusions. Later in the term, hydrogen bonding was also mentioned in the context of protein folding, ligand binding, and catalysis. Before analyzing student learning gains data from Fall 2014, the instructor believed that this instruction was sufficient to create student understanding.

In Fall 2014, the instructor administered the current version of the IFCB for the first time and was struck by students' persistent incorrect ideas about hydrogen bonding even after instruction (see Table 5 and Figure 1, Fall 2014). Figure 1 shows that 64% of students in Fall 2014 were unable to demonstrate understanding of the concept on both the pretest and posttest ("never knew" in blue). Twelve percent of students came to know and 6% went backward. This result signifies poor student understanding, frequent guessing, or simply forgetting basic knowledge. Without basic knowledge of hydrogen bonding, it is impossible for students to meet the expectation of applying the concept of hydrogen bonding to understand a more complex biochemical context.

In response to these data, the instructor introduced two new clicker questions in all subsequent quarters (shown in the Supplemental Document). The new questions were added after the previously described clicker question regarding urea and hydrogen bonding. In each new clicker question, students were shown two biologically relevant small molecules interacting and were asked, "Is this a hydrogen bond?" One question depicted a canonical hydrogen bond and the other depicted an interaction that was not a hydrogen bond. The molecules are different from those on the IFCB to avoid rote memorization. Salient characteristics of each interaction were briefly discussed after all students had responded. In the subsequent three quarters of instruction there was a large increase in the proportion of students in the came to know group when the learning gains data were analyzed (Figure 1, Winter 2015, Spring 2015, Fall 2015). This increase in the group of students who came to know the concept is stable over subsequent quarters, indicating the improvement is likely due to the introduction of the two clicker questions. Therefore, the hydrogen-bonding concept is a key example of how use of the IFCB enabled an instructor to obtain better insights into student understanding and the effectiveness of instruction, which resulted in meaningful instructional changes.

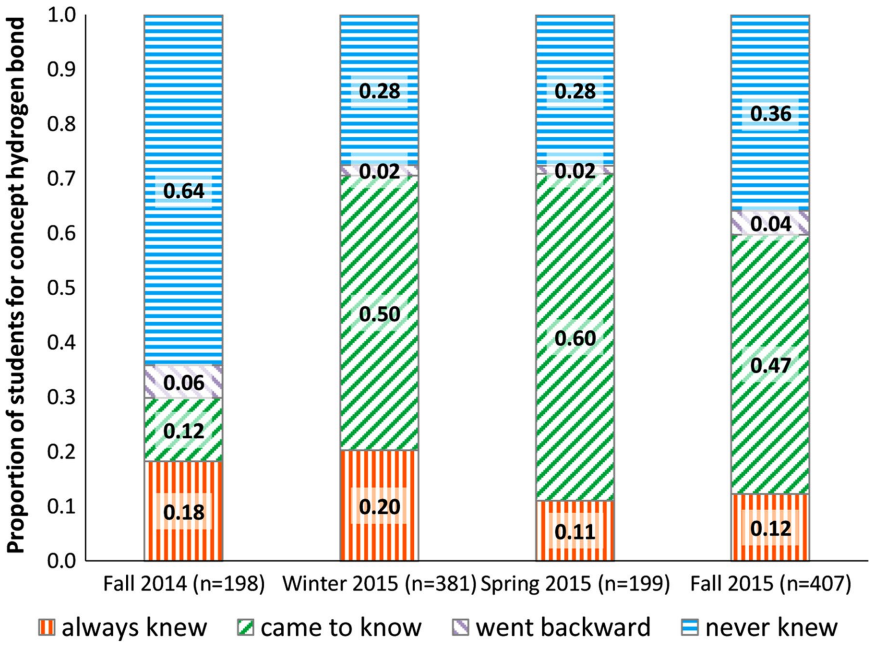

FIGURE 1. Proportion of students in each group for the hydrogen-bonding items.

In addition to tracking changes in learning gains over time, the structure of the IFCB allows for identification of common incorrect ideas. To correctly answer IFCB questions related to hydrogen bonding, students need to recognize that a

hydrogen-bonding interaction is a noncovalent interaction involving a small electronegative atom and a hydrogen covalently bound to a small electronegative atom (correct idea A). From Table 2, the incorrect ideas related to this concept are 1) that a hydrogen covalently bound to a carbon can participate in hydrogen-bonding interactions (incorrect idea B); 2) that all hydrogens are capable of participating in a hydrogen bond, regardless of their covalent bond participation (incorrect idea C); and 3) that hydrogen bonding is a covalent bond between a hydrogen and another atom (incorrect idea D). We analyzed student patterns of responses for the three quarters after introduction of the new clicker questions. Patterns for all quarters analyzed were similar (raw percentage data can be found in Supplemental Table 1s).

Figure 2 illustrates how student responses shifted over time by presenting a ribbon graph of representative results from Fall 2015, the most recent quarter with the largest sample size (N = 407). This figure tracks responses of individual students to the set of three items related to hydrogen bonding on the pretest (the left-hand side) and the posttest (the right-hand side). As discussed in the *Methods* section, we grouped the students into six catgories based on their response patterns. Category AAA is for consistently correct responses. Category AXX is for inconsistent responses, with at least one correct and one incorrect. Category BBB, CCC, or DDD is for consistently incorrect responses corresponding to the incorrect statement with the given letter (see Table 2). Finally, category XXX is for inconsistently incorrect responses, in other words, for students who selected a mixture of incorrect responses. In the ribbon graph, lines connect the responses of individual students on the pretest (left) to their responses on the posttest (right), creating "ribbons." The ribbons revealing movement toward dominant categories (categories with more than 50 students) on the posttest are highlighted. Analysis of data in this way allows the instructor to determine whether changes in instruction lead to correction of specific incorrect ideas and/or whether instruction leads to more consistent understanding of the concept.

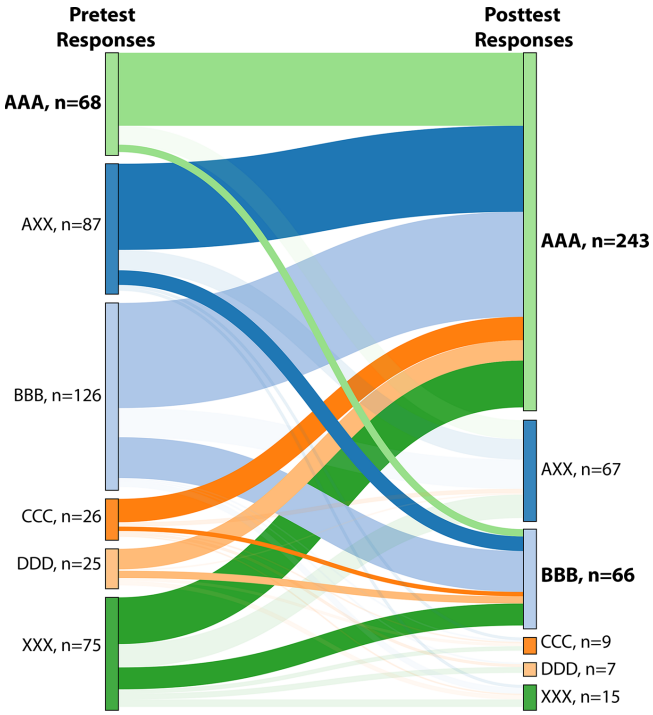

FIGURE 2. Categories of student responses on hydrogen-bonding items Fall 2015 (n = 407). AAA: consistently correct responses for three items; AXX: at least one correct response in combination with incorrect response(s); BBB: consistently incorrect responses corresponding to statement B; CCC: consistently incorrect responses corresponding to statement C; DDD: consistently incorrect responses corresponding to statement D; XXX: inconsistently incorrect responses, corresponding to a mixture of statements B, C, and D. The last category includes missing responses only if at least one of the items in the set did receive a response from that student. Categories corresponding to all answers correct (AAA) on either the pretest or the posttest, and the most common incorrect responses on the posttest are bolded. Statements can be found in Table 2.

On the pretest, the two most common categories reflect consistently and inconsistently incorrect ideas: BBB (126 students) and partially correct (87 students). These responses demonstrate that a majority of students enter the course with a specific misconception (hydrogen-bonding idea B, Table 2) or with confusion, susceptible to influence by item context or possibly guessing. On the posttest, the two most common combinations are AAA (243), and BBB (66). As shown by the width (top to bottom) of the ribbons on the graph (Figure 2), students tended to shift away from consistently incorrect ideas (BBB, and, in smaller numbers, CCC and DDD) and mixed incorrect ideas (XXX) to the consistently correct idea (AAA). Students starting with mixed incorrect ideas (XXX) on the pretest also show some shifting toward BBB and toward partially correct (AXX) on the posttest, though the majority moved to AAA. Although the data reveal that not all students abandoned their incorrect ideas, the increase in the number of students who could consistently identify a hydrogen bond after instruction was substantial. The data are, of course, insufficient to demonstrate a causal relationship between clicker questions and student recognition of hydrogen bonding, but the results were informative and intriguing for the instructor, who went on to consider whether similar small changes could better support student understanding in other areas.

Bond Energy

The bond energy items assess whether students know that breaking an isolated bond always requires energy. Before administering the IFCB in Fall 2014, the instructor was not aware of students' incorrect ideas related to bond energy and did not explicitly focus on the concept that formation of an isolated chemical bond releases energy. Historically, the concept of bond energy was referred to throughout the quarter in discussions of strength of ligand binding, protein folding, and cleaving "high-energy" bonds. Figure 3 shows how students' understanding of bond energy changed over the course of each quarter. In Fall 2014, the largest percent of students are in the never knew group. Only 14% students came to know, and 12% went backward, indicating that instruction did not play a major role in fixing the incorrect ideas observed on the pretest. After administering the instrument in Fall 2014 and observing poor performance on this concept, the instructor decided to include explicit instruction that had not previously been a part of the

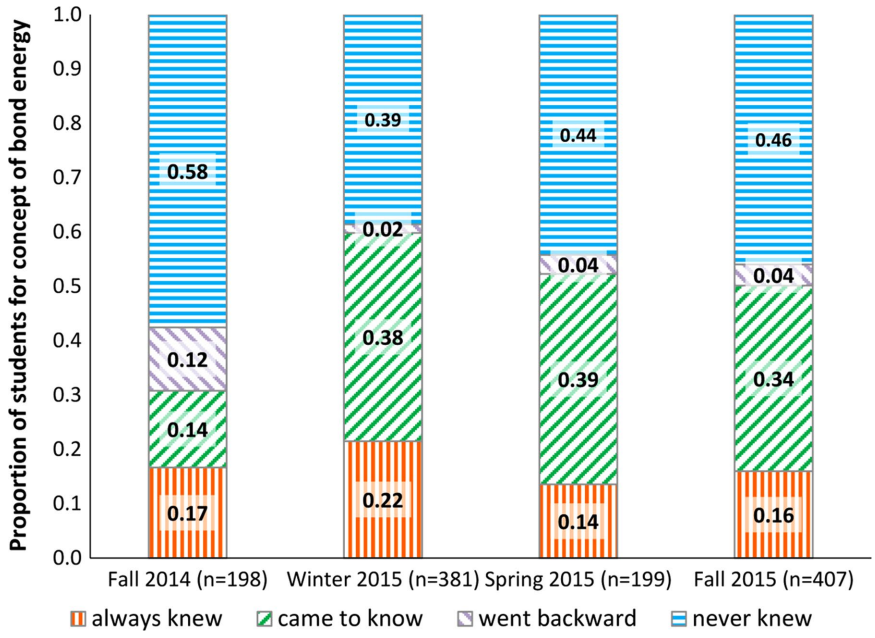

FIGURE 3. Proportion of students in each group for the bond energy items.

class. Beginning in Winter 2015, students were asked in discussion to consider the relative strengths of interactions that had to be broken and formed in the process of macromolecular structure formation. Each subsequent quarter, more emphasis was put on understanding this aspect of structure formation. In Spring 2015, a worksheet on the thermodynamics of protein folding was introduced in discussion during the second week of the quarter. The worksheet required students to draw different interactions and explicitly consider how much heat would be required to break the bonds (or interactions) along with how much heat would be released by the new interactions (or bonds) that would form. After these changes, the proportion of students in the came to know group doubled, and very few students went backward (Figure 3, Winter 2015, Spring 2015, Fall 2015).

Again, common incorrect ideas and consistency of responses were tracked using the parallel structure of the IFCB. To correctly answer questions related to bond energy, students needed to know that breaking an isolated chemical bond always requires energy and forming an isolated chemical bond always releases energy (correct idea A). As shown in Table 2, the incorrect ideas related to this concept are 1) whether energy is released or absorbed depends on the strength of the bond (incorrect idea B), 2) whether energy is released or absorbed is conditional in an unspecified way (incorrect idea C), and 3) breaking an isolated chemical bond always releases energy and forming an isolated chemical bond always requires energy (incorrect idea D). On the pretest for each quarter, roughly 50% of students were inconsistent in their responses, demonstrating some confusion (Supplemental Table 1s). Changes in patterns of students picking statements for Fall 2015 on the pretest and posttest are shown in a ribbon graph in Figure 4. On the pretest, the most common consistent combinations are DDD (109 students) and AAA (81 students). This demonstrates that more than 25% of students enter biochemistry with an idea about bond energy that is opposite to reality. Most of the remaining students who were not in the AAA category have mixed correct and incorrect ideas (AXX: 107 students; XXX: 108 students), revealing that this population of students is confused or guessing.

Our results suggest that instructional interventions used starting in Winter 2015 were effective, as evidenced by a major shift in student responses from inconsistent on the pretest (AXX and XXX) to AAA on the posttest (Supplemental Table 1s). To illustrate this shift, all the lines that converge to AAA or DDD on the posttest for Fall 2015 are highlighted in the ribbon graph (Figure 4). However, even after instruction, 68 students were in the DDD category on the posttest this quarter, suggesting that some students remember that it always either takes energy (A) or releases energy (D) to form an isolated chemical bond, but don't understand the concept well enough to decide which one

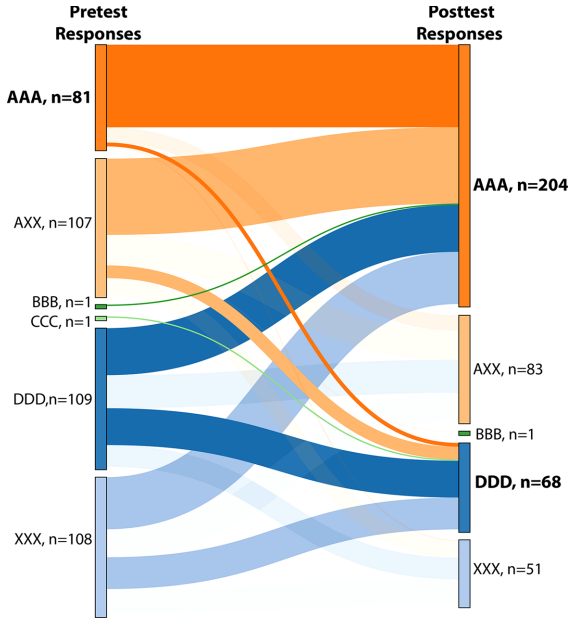

FIGURE 4. Categories of student responses on bond energy items Fall 2015 ($n = 407$). AAA: consistently correct responses for three items; AXX: at least one correct response in combination with incorrect response(s); BBB: consistently incorrect responses corresponding to statement B; CCC: consistently incorrect responses corresponding to statement C; DDD: consistently incorrect responses corresponding to statement D; XXX: inconsistently incorrect responses, corresponding to a mixture of statements B, C, and D. The last category includes missing responses only if at least one of the items in the set did receive a response from that student. Categories corresponding to all answers correct (AAA) on either the pretest or the posttest, and the most common incorrect responses on the posttest are bolded. Statements can be found in Table 2.

is correct. This suggestion is also supported by the fact that only 21 out of 407 students never picked A or D on the pretest, and only 13 students never picked A or D on the posttest. No students consistently endorse on both tests the conditional statements B and C that energy intake during bond formation depends on other factors such as bond strength.

pK_a

The third targeted concept, pK_a, was covered extensively in the course before use of the IFCB, and instruction changed little over the four quarters reported. This situation is in contrast to the instructor's prior actions with regard to hydrogen bonding and bond energy. For those two concepts, the instructor had originally assumed either that students understood the concepts well from their previous course work or that a minimal review would suffice for students to understand and use the concepts in biochemistry. For pK_a, however, the instructor emphasized this concept heavily, even before use of the IFCB, because the instructor had previously realized students enter biochemistry without a clear understanding of this concept. In the first week of each quarter, students were reintroduced to buffers and completed an in-class activity that asked them to use pK_a values to determine isoelectric point values for amino acids and peptides (from *Foundations of Biochemistry*, 3rd edition; Loertscher and Minderhout, 2011). The concept was further discussed in the context of ion-exchange chromatography. Finally, as part of the discussion of enzyme mechanisms, students were asked to examine protonation states of substrates and amino acids in enzyme active sites, predict protonation states, and relate these structural features to binding interactions.

Figure 5 shows student knowledge of pK_a over the course of the four quarters studied. The percentage of students in each group is very similar for all four quarters, with more than 50% of students in the came to know group each quarter. This consistency is not surprising, since no new specialized instruction related to this concept was introduced. These results illustrate that analysis of learning gains data can be used not only to identify problematic areas in which instructional changes are needed but also to identify those concepts that are well supported by instruction. Additionally, despite the fact that the instrument was administered slightly differently each quarter (Table 1), the similarity of response patterns suggests no evidence of bias associated with differences in administration, which lends support to the validity of instrument use with the studied population. It is also notable that changes in instruction related to hydrogen bonding and bond energy, discussed previously, did not disrupt the gains related to pK_a that were already part of normal instruction.

Again, common incorrect ideas and consistency of responses were tracked using the parallel structure of the IFCB. To answer questions related to pK_a correctly, students needed to determine the predominant charge of ionizable groups (-COOH and $-NH_2$) depending on the pH and pK_a. Table 2 presents the incorrect ideas probed by the IFCB. Changes in patterns of student responses between the pretest and posttest are shown in a ribbon graph in Figure 6. On the pretest, the incorrect ideas are often mixed and diversely distributed across the three items, indicating poor understanding or guessing for this concept. A total of 198 (AXX, 49%) students chose the correct response (A) at least once in combination with incorrect responses. The second most-prevalent set of responses on the pretest was mixed incorrect ideas (XXX, 124 students or 30%). Only 54 students (13%) answered all three questions correctly (AAA). Two consistently incorrect ideas (BBB and CCC) were selected by a few students (five and 26, respectively) on the pretest, but by the posttest, all of these students had shifted to other responses, and only a few (six) students had moved to CCC from XXX. The most frequently selected combination on the posttest was the correct answer (AAA), chosen by 244 students (60%). In general, students in each category on the pretest moved in the direction of better understanding on the posttest. Instruction related to pK_a is distinguished from the other two concepts in that the instructor did not need additional formative assessment data to help inform instruction. The instructor already had a strong sense, which was borne out by the IFCB pretest, that students coming into biochemistry do not understand pK_a very well and had proactively worked instruction about pK_a into the normal curriculum. It could be that the instructor recognized the need to work on student understanding of pK_a because having an incorrect idea related to pK_a is more visible in a biochemistry context than having an incorrect idea related to hydrogen bonding or bond energy.

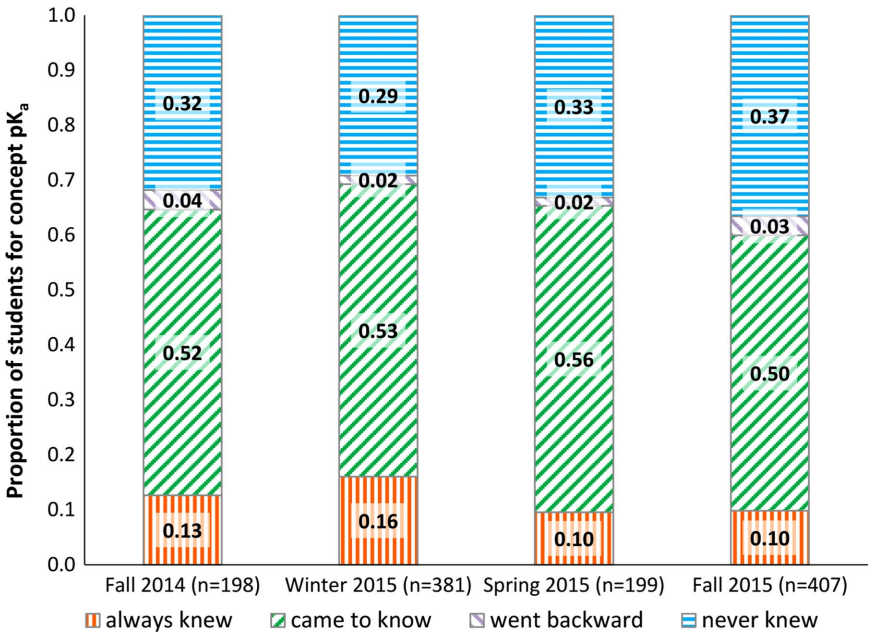

FIGURE 5. Proportion of students in each group for the pK_a items.

DISCUSSION

This study describes the revision and implementation of an instrument to measure students' knowledge of prerequisite concepts that are essential to success in

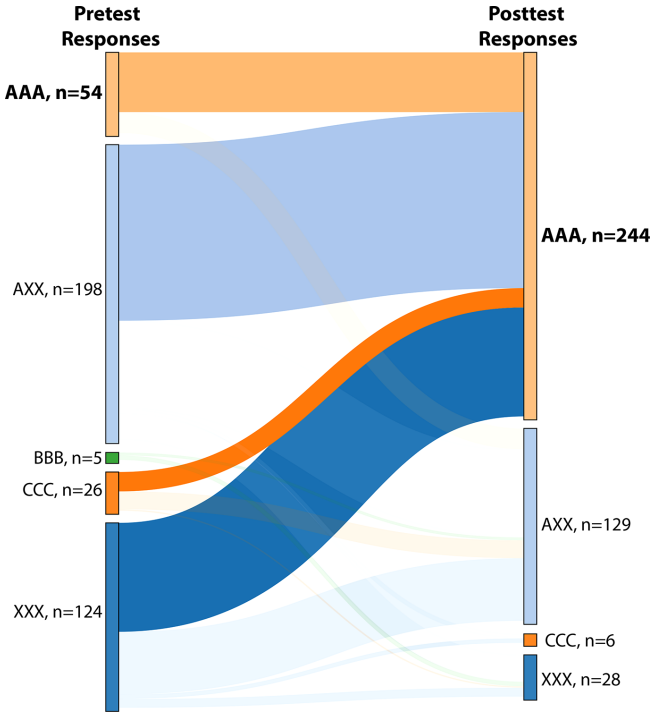

FIGURE 6. Categories of student responses on pK_a items Fall 2015 ($n = 407$). AAA: consistently correct responses for three items; AXX: at least one correct response in combination with incorrect response(s); BBB: consistently incorrect responses corresponding to statement B; CCC: consistently incorrect responses corresponding to statement C; DDD: consistently incorrect responses corresponding to statement D; XXX: inconsistently incorrect responses, corresponding to a mixture of statements B, C, and D. The last category includes missing responses only if at least one of the items in the set did receive a response from that student. Categories corresponding to all answers correct (AAA) on either the pretest or the posttest, and the most common incorrect responses on the posttest are bolded. Statements can be found in Table 2.

biochemistry. The effect of instructional changes in an undergraduate biochemistry classroom was assessed using the revised instrument. The findings show important interactions among assessment, instructional interventions, and student knowledge. Instrument development and validation is an ongoing and iterative process. With the described revisions, we have shown that the current version of IFCB functions better to identify student knowledge with evidence of improved reliability and validity and is useful in helping to understand the effects of instructional changes.

Targeted Concepts and Their Importance to Biochemistry

Three foundational concepts of this study were chosen because of their importance to learning in biochemistry and because of documented difficulties related to student understanding of these concepts, both of which will be discussed in detail for each concept. It is important to note that the targeted learning outcomes for each of these three concepts are relatively low level, asking students to identify hydrogen-bonding interactions, recall information about bond energy, and apply knowledge about pK_a to make predictions about charges on molecules. However, even at this level, these outcomes are relevant for learning in biochemistry, because immediate recall of these key concepts is necessary for students to engage with biochemistry in a meaningful way. Because most biochemistry instructors assume that students gained a deep understanding of these concepts in prerequisite courses, data from the IFCB can alert instructors to problems they had not anticipated so that changes in instruction can be made to best support student learning.

Our analysis shows similarities and areas of distinction in student response patterns related to the three concepts. For all three concepts, the posttest shows more correct and consistent student performance. However, each concept also exhibits unique trends in student response patterns. For example, for hydrogen bonding, student responses shifted from several different, yet consistent, incorrect ideas along with inconsistency to the consistently correct idea. For bond energy, only one incorrect idea was consistently attractive to students on both the pretest and posttest. Finally, student responses related to pK_a moved from very inconsistent to more consistently correct. It is possible that these patterns could be widely observed across diverse institutions, but the real power of the IFCB comes from the ability of individual instructors to identify the specific patterns in their own student populations. Implications of findings related to each of the concepts are discussed in detail in the following sections.

Findings and Implications

Hydrogen Bonding. Hydrogen bonding is important in biochemistry and is particularly noticeable in research articles describing protein structure. A simple search of PubMed Central using hydrogen bonding as a search term resulted in more than 58,000 citations (Kuster *et al.*, 2015). Yet many studies involving students enrolled in general chemistry, organic chemistry, and biochemistry show poor student understanding of noncovalent interactions in general and hydrogen bonding in particular. Interviews of students who had completed second-semester organic chemistry revealed that many relied on rote memorization of the definition of hydrogen bonding and lacked the ability to explain why the definition works (Rushton *et al.*, 2008; Williams *et al.*, 2015). Furthermore, some students identified a covalent bond between carbon and hydrogen as a hydrogen bond (Henderleiter *et al.*, 2001). In another study involving general chemistry students, few students were able to correctly identify hydrogen bonding as occurring between ethanol molecules, and fewer still drew the interactions correctly (Cooper *et al.*, 2015). Many depicted hydrogen bonds as covalent bonds within an ethanol molecule. These authors also found that students used appropriate words without deeply understanding their meaning. Another study showed that fewer than 20% of students could correctly identify a hydrogen bond even after a semester of biochemistry instruction (Villafañe *et al.*, 2011b).

Changes in general chemistry curriculum have been shown to improve students' understanding of hydrogen bonding and other noncovalent interactions. Most students enrolled in the CLUE curriculum, an innovative general chemistry curriculum that uses an integrated approach in which students are frequently asked to contrast and explain models, correctly identified hydrogen bonding as occurring between molecules and

correctly identified the atoms involved in the interaction (Williams *et al.*, 2015). In contrast, Williams and colleagues show that only ~30% of students in a traditional course exhibited this level of understanding. It was further shown that students tend to retain the understanding of hydrogen bonding that they had in general chemistry, whether correct or incorrect, through the end of instruction in organic chemistry (Cooper *et al.*, 2015). This suggests that biochemistry faculty should not expect students progressing through a traditional curriculum to bring a correct understanding of hydrogen bonding to their courses. The IFCB presented in this study can identify areas in which students are lacking understanding and can help colleagues within a department align prerequisite courses.

This investigation aligns with previously published studies showing that students do not know the basics of hydrogen bonding when entering biochemistry. However, most biochemistry faculty do not plan to spend time on this concept, as it is assumed prerequisite knowledge from previous chemistry courses. In exit evaluations administered as part of multiple faculty workshops related to teaching in biochemistry, many instructors state that they had assumed that students understood hydrogen bonding when they began biochemistry (unpublished data). It was similarly surprising to the instructor in this study that nearly 70% of students remained unclear on this topic after completing the biochemistry course in Fall 2014 (Figure 2).

On the basis of these data and the literature, we realized that we could not expect students to understand higher-order structure in biochemistry if we could not even rely on students looking at the correct part of the molecule when we discussed hydrogen bonding. Therefore, we decided as a first step that it was absolutely necessary for biochemistry students to learn to accurately identify hydrogen-bonding interactions in biologically relevant molecules. If biochemistry students are unable to identify whether hydrogen bonding between two molecules or parts of molecules is even possible, how could we expect them to make predictions about the formation of complex macromolecular structures? In response, the instructor in this study implemented clicker questions to address a simple hydrogen-bonding definition. This small change in instruction could be associated with the large increase in the group of students who came to know this concept at the end of course. In other words, once the instructor recognized the difficulty students had with this concept, even a small intervention could result in a substantial shift in student understanding. Here, we have shown that a set of multiple-choice items can provide a convenient way for instructors to visualize a range of student correct and incorrect ideas related to hydrogen bonding.

Bond Energy. Bond energy has been deemed important for multiple disciplines, including biochemistry (Cooper and Klymkowsky, 2013). The literature has previously shown that students are confused as to whether energy is absorbed or released in a simple bond formation in a variety of contexts (Ross, 1993; Boo, 1998; Teichert and Stacy, 2002; Özmen, 2004; Villafañe *et al.*, 2011b; Cooper and Klymkowsky, 2013). Boo suggests that this problem arises from a worldview that building a structure requires energy input, whereas destruction involves release of energy—that is, students believed that bond breaking releases energy and bond making involves energy input. Cooper and Klymkowsky attribute the source of this misconception to everyday use of language and a macroscopic approach in multiple science courses (Cooper and Klymkowsky, 2013). The misconception might also come from our simplified language around "high-energy" phosphoanhydride bonds in ATP. Instructors assume the students understand we are talking about the total chemical reaction, not just the physical breaking of an isolated bond. Interestingly, after the instructor in this study started emphasizing the fact that energy is absorbed when bonds are broken, students began asking, "But how can ATP be a high-energy bond if it always takes energy to break a bond?" Most processes that biochemists encounter involve both bond breaking and forming, and the important variable is overall change in energy. Yet an understanding of the direction of energy flow at the level of individual bonds is important and is a first step in enabling deep understanding of the thermodynamics of complex processes. Furthermore, even in the context of living organisms, there are cases in which bonds are broken without concomitant bond formation, such as in the generation of free radicals through ionizing radiation. The high reactivity of these and other free radicals (superoxide anion radical) is fundamentally understood due to the inherent free-energy change when a bond is formed.

Results from Fall 2014, before specialized instruction related to bond energy was introduced, mirror results previously reported in the literature (Figure 3). After incorrect ideas were revealed from assessment data, the instructor included more intentional instruction related to this concept, and student knowledge improved overall, although some students still retained an incorrect idea. While the change in instruction was not a specific intervention but rather a decision to discuss bond energy more explicity in several places in the existing curriculum, it is likely the cause of the improvement in student understanding of the concept.

pK_a. Meaningful understanding of pK_a is very important to learning biological concepts and in applying this knowledge to contexts such as buffers and electrophoresis (Moore, 1985; Curtright *et al.*, 2004; Orgill and Sutherland, 2008). The concept of pK_a is especially difficult, because it involves both conceptual and algorithmic manipulations. This study focused on how to use the pK_a of an ionizable group within a molecule to determine the protonation state at a given pH. The instructor for this biochemistry course had previously observed that incoming students were not well versed in using pK_a to make predictions and had already included specific instruction on pK_a in the course. As a result, 60% of students in Fall 2015 were able to answer these questions correctly at the end of the course. The remaining student responses were inconsistent, but a large fraction (32%) selected the correct response (A) in at least one context, indicating some achievement toward implementing pK_a in all circumstances.

Limitations and Future Directions

We used the revised IFCB with evidence of reliability and validity to identify student incorrect ideas and to probe the effectiveness of instructional changes in a quick and reliable way. This work is a first step in an iterative process to use assessment data to improve teaching and learning in biochemistry over time. Therefore, the study presents a number of limitations but also promising opportunities for future work.

One limitation is that students' knowledge of the three target concepts was probed using multiple-choice questions only. Other assessment techniques could provide rich information on student understanding in large-enrollment courses. For example, Haudek *et al.* (2012) used automated text analysis to examine responses of students enrolled in an introductory cellular and molecular biology class at a large public university. They probed students' understanding of a specific question: "Which functional groups, hydroxyl (-OH) or amino (-NH_2), are most likely to have an impact on cytoplasmic pH?" (Haudek *et al.*, 2012). Similarly, assessment of student responses to open-ended questions related to the target concepts in this study could be conducted in the future and could provide deeper insights into students' thinking. Regardless of the limitation of multiple-choice format, the IFCB had the potential for wide classroom use due to the advantage of easy administration, quick analysis, and utility to provide timely and valid information that instructors can readily use to make adjustments in their teaching.

Another limitation is the simplicity of the described instructional interventions. For example, our instructional strategy related to hydrogen bonding is to show examples illustrating what is and what is not hydrogen bonding. This "tell and practice" approach was effective to help students remedy their consistently incorrect ideas, but similar approaches have been criticized by Schwartz and coworkers because use of familiar routines can limit students' engagement with new ways of thinking about problems (Schwartz *et al.*, 2012). However, we would argue that this gap in basic knowledge must be fixed to prepare students for future learning. In the next step, more active-learning activities such as process-oriented guided-inquiry learning (Loertscher and Minderhout, 2011) can be used to help students to build more advanced and complicated conceptual links, especially when students have mixed rather than consistent incorrect ideas. Ultimately, rather than having memorized disconnected pieces of knowledge from being told an answer to a question or following a procedure of pattern matching, students should be encouraged to have freedom to process information, to explore their own best solutions to a complex problem, and to be told the accepted solution only afterward (Schwartz *et al.*, 2011).

This work provides an example of iterative process of gathering validity evidence associated with an instrument and observing improvement in student learning due to interventions. This study showed that a multiple-choice instrument can be improved over time and can be used to quickly identify areas for instructional attention, such as hydrogen bonding and bond energy, which are not emphasized in traditional curricula. The assumption that students have mastered a basic level of fundamental knowledge from prerequisite courses may not necessarily be true. In addition, this study demonstrates that IFCB can be used as a pretest and posttest to help inform instructional choices for a large-enrollment biochemistry course. Even slight changes in the curriculum can help students more fully explore and transfer knowledge from previous gateway chemistry courses to the more advanced biochemistry course. In the future, it would be beneficial to elevate the level of learning outcomes related to all of the target concepts and to develop instructional materials that use evidence-based practices to support students' learning and transfer of these concepts.

ACKNOWLEDGMENTS

The work described in this article was supported by National Science Foundation (NSF) DUE-0717392 and DUE-1224868. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). Standards for Educational and Psychological Testing, Washington, DC.

Arjoon JA, Xu X, Lewis JE (2013). Understanding the state of the art for measurement in chemical education research: examining the psychometric evidence. J Chem Educ 90, 536–545.

Bauer CF (2008). Attitude toward chemistry: a semantic differential instrument for assessing curriculum impacts. J Chem Educ 85, 1440.

Bennett RE (2011). Formative assessment: a critical review. Assess Educ Princ Pol Pract 18, 5–25.

Bentler PM (1990). Comparative fit indexes in structural models. Psychol Bull 107, 238–246.

Boo HK (1998). Students' understandings of chemical bonds and the energetics of chemical reactions. J Res Sci Teach 35, 569–581.

Bretz SL, Linenberger KJ (2012). Development of the enzyme–substrate interactions concept inventory. Biochem Mol Biol Educ 40, 229–233.

Brown GA, Bull J, Pendlebury M (2013). Assessing Student Learning in Higher Education, New York: Routledge.

Brown TA (2015). Confirmatory Factor Analysis for Applied Research, New York: The Guilford Press.

Caldwell JE (2007). Clickers in the large classroom: current research and best-practice tips. CBE Life Sci Educ 6, 9–20.

Chandrasegaran A, Treagust DF, Mocerino M (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. Chem Educ Res Pract 8, 293–307.

Cooper MM, Klymkowsky MW (2013). The trouble with chemical energy: why understanding bond energies requires an interdisciplinary systems approach. CBE Life Sci Educ 12, 306–312.

Cooper MM, Williams LC, Underwood SM (2015). Student understanding of intermolecular forces: a multimodal study. J Chem Educ 92, 1288–1298.

Curtright R, Emry R, Heaton RM, Markwell J (2004). Facilitating student understanding of buffering by an integration of mathematics and chemical concepts. Biochem Mol Biol Educ 32, 71–77.

Dunn KE, Mulvenon SW (2009). A critical review of research on formative assessment: the limited scientific evidence of the impact of formative assessment in education. Pract Assess Res Eval 14, 1–11.

Evans C (2013). Making sense of assessment feedback in higher education. Rev Educ Res 83, 70–120.

Fornell C, Larcker DF (1981). Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 18, 39–50.

Haladyna TM, Downing SM, Rodriguez MC (2002). A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ 15, 309–333.

Haudek KC, Kaplan JJ, Knight J, Long T, Merrill J, Munn A, Nehm R, Smith M, Urban-Lurain M (2011). Harnessing technology to improve formative assessment of student conceptions in STEM: forging a national network. CBE Life Sci Educ 10, 149–155.

Haudek KC, Prevost LB, Moscarella RA, Merrill J, Urban-Lurain M (2012). What are they thinking? Automated analysis of student writing about acid–base chemistry in introductory biology. CBE Life Sci Educ 11, 283–293.

Henderleiter J, Smart R, Anderson J, Elian O (2001). How do organic chemistry students understand and apply hydrogen bonding? J Chem Educ 78, 1126–1130.

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? Australian Biochemist 39, 14–17.

Hu L, Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 6, 1–55.

Kay RH, LeSage A (2009). Examining the benefits and challenges of using audience response systems: a review of the literature. Comput Educ 53, 819–827.

Kuster DJ, Liu C, Fang Z, Ponder JW, Marshall GR (2015). High-resolution crystal structures of protein helices reconciled with three-centered hydrogen bonds and multipole electrostatics. PLoS One 10, e0123146.

Loertscher J, Green D, Lewis JE, Lin S, Minderhout V (2014a). Identification of threshold concepts for biochemistry. CBE Life Sci Educ 13, 516–528.

Loertscher J, Minderhout V (2011). Foundations of Biochemistry, 3rd ed., Hampton, NH: Pacific Crest.

Loertscher J, Villafañe SM, Lewis JE, Minderhout V (2014b). Probing and improving student's understanding of protein α-helix structure using targeted assessment and classroom interventions in collaboration with a faculty community of practice. Biochem Mol Biol Educ 42, 213–223.

Marbach-Ad G, McAdams KC, Benson S, Briken V, Cathcart L, Chase M, El-Sayed NM, Frauwirth K, Fredericksen B, Joseph SW (2010). A model for using a concept inventory as a tool for students' assessment and faculty professional development. CBE Life Sci Educ 9, 408–416.

Moore DS (1985). Amino acid and peptide net charges: a simple calculational procedure. Biochem Educ 13, 10–11.

Murphy KR, Davidshofer CO (2005). Psychological Testing: Principles and Applications, 6th ed., Upper Saddle River, NJ: Prentice Hall.

Orgill M, Sutherland A (2008). Undergraduate chemistry students' perceptions of and misconceptions about buffers and buffer problems. Chem Educ Res Pract 9, 131–143.

Özmen H (2004). Some student misconceptions in chemistry: a literature review of chemical bonding. J Sci Educ Technol 13, 147–159.

Pagliarulo C, Molinaro M (2015). iAMSTEM: increasing student success at UC-Davis. Peer Rev 17, 13.

Pintrich PR, Smith DA, García T, McKeachie WJ (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). Educ Psychol Meas 53, 801–813.

Ross K (1993). There is no energy in food and fuels, but they do have fuel value. Sch Sci Rev 75, 39–47.

Rushton GT, Hardy RC, Gwaltney KP, Lewis SE (2008). Alternative conceptions of organic chemistry topics among fourth year chemistry students. Chem Educ Res Pract 9, 122–130.

Sadler DR (1989). Formative assessment and the design of instructional systems. Instr Sci 18, 119–144.

Schwartz DL, Chase CC, Bransford JD (2012). Resisting overzealous transfer: coordinating previously successful routines with needs for new learning. Educ Psychol 47, 204–214.

Schwartz DL, Chase CC, Oppezzo MA, Chin DB (2011). Practicing versus inventing with contrasting cases: the effects of telling first on learning and transfer. J Educ Psychol 103, 759–775.

Sears DW, Thompson SE, Saxon SR (2007). Reversible ligand binding reactions: why do biochemistry students have trouble connecting the dots? Biochem Mol Biol Educ 35, 105–118.

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, 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.

Songer CJ, Mintzes JJ (1994). Understanding cellular respiration: an analysis of conceptual change in college biology. J Res Sci Teach 31, 621–637.

Tan DK-C, Goh NK, Chia LS, Treagust DF (2002). Development and application of a two-tier multiple choice diagnostic instrument to assess high school students' understanding of inorganic chemistry qualitative analysis. J Res Sci Teach 39, 283–301.

Tan DK-C, Treagust DF (1999). Evaluating students' understanding of chemical bonding. Sch Sci Rev 81, 75–84.

Tansey JT, Baird T, Cox MM, Fox KM, Knight J, Sears D, Bell E (2013). Foundational concepts and underlying theories for majors in "biochemistry and molecular biology." Biochem Mol Biol Educ 41, 289–296.

Teichert MA, Stacy AM (2002). Promoting understanding of chemical bonding and spontaneity through student explanation and integration of ideas. J Res Sci Teach 39, 464–496.

Tsui CY, Treagust D (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. Int J Sci Educ 32, 1073–1098.

Villafañe SM, Bailey CP, Loertscher J, Minderhout V, Lewis JE (2011a). Development and analysis of an instrument to assess student understanding of foundational concepts before biochemistry coursework. Biochem Mol Biol Educ 39, 102–109.

Villafañe SM, Loertscher J, Minderhout V, Lewis JE (2011b). Uncovering students' incorrect ideas about foundational concepts for biochemistry. Chem Educ Res Pract 12, 210–218.

Williams LC, Underwood SM, Klymkowsky MW, Cooper MM (2015). Are noncovalent interactions an Achilles heel in chemistry education? A comparison of instructional approaches. J Chem Educ 92, 1979–1987.

Xu X, Kim ES, Lewis JE (2016). Sex difference in spatial ability for college students and exploration of measurement invariance. Learn Individ Differ 45, 176–184.

Xu X, Lewis JE (2011). Refinement of a chemistry attitude measure for college students. J Chem Educ 88, 561–568.