# Using the Intended–Enacted–Experienced Curriculum Model to Map the *Vision and Change* Core Competencies in Undergraduate Biology Programs and Courses

**Alexa W. Clemmons,[†]* Deborah A. Donovan,[‡§] Elli J. Theobald,[‖] and Alison J. Crowe[‖]**

[†]Codon Learning, Golden, CO 80401; [‡]Biology Department and [§]Science Education Group, Western Washington University, Bellingham, WA 98225; [‖]Department of Biology, University of Washington, Seattle, WA 98195

## ABSTRACT

**One critical step in the challenging process of curricular reform is determining how closely a curriculum aligns with national recommendations. Here, we examine the alignment of teaching, assessment, and student experience in undergraduate biology courses with the *Vision and Change* core competency recommendations. We applied the intended–enacted–experienced curriculum model to obtain a more complete, multiperspective view of the curriculum. First, we developed and piloted the BioSkills Curriculum Survey with more than 100 biology instructors across five institutions. Using multilevel logistic regression modeling of the survey data, we found that instructors were equally likely to report teaching all competencies; however, they reported assessing some competencies more than others. After adding course characteristics to our model, we found that the likelihood of teaching certain competencies depended on course type. Next, we analyzed class materials and student perceptions of instruction in 10 biology courses in one department. Within this smaller sample, we found that instructors messaged a narrower range of competency learning outcomes on their syllabi than they reported teaching on the survey. Finally, modeling revealed that inclusion of an outcome on assessments, but not syllabi, increased the likelihood that students and their instructor agreed whether it was taught.**

## INTRODUCTION

### Undergraduate Biology Curricular Reform

*Vision and Change in Undergraduate Biology Education* was written to be the guiding document for pedagogy and curricula in U.S. undergraduate biology (American Association for the Advancement of Science [AAAS], 2011). The recommendations resulted from several iterative meetings of more than 500 stakeholders. The meetings were motivated by a combination of advances in science education research and the changing nature of biology research and associated career paths. The stakeholders, who included biologists, biology educators, college administrators, and discipline-based education researchers, combined their varied expertise to inform a vision for the future of biology education. The curricular recommendations of *Vision and Change* encourage instructors to structure biology courses around five core concepts and six core competencies. In other words, courses should de-emphasize facts and content in favor of overarching themes and transferable competencies that will prepare students for a variety of careers and life as scientifically literate citizens. Considering the vast and growing landscape of biological fields and knowledge, a core concept– and competency-based curriculum

Tammy Long, *Monitoring Editor*
Submitted Mar 15, 2021; Revised Nov 1, 2021; Accepted Nov 17, 2021
CBE Life Sci Educ March 1, 2022 21:ar6
DOI:10.1187/cbe.21-02-0054

*Address correspondence to: Alexa W. Clemmons (alexaclemmons@gmail.com).

helps faculty and administrators make difficult decisions about what should stay and, importantly, what can be cut from college courses. Similar curricular reform efforts are underway in subdisciplines of biology, other undergraduate disciplines, and K–12 science (Michael *et al.*, 2009; Merkel and ASM Task Force on Curriculum Guidelines for Undergraduate Microbiology, 2012; National Research Council [NRC], 2012; Tansey *et al.*, 2013; Grunspan *et al.*, 2018; Talanquer *et al.*, 2020).

### Toward a Deeper Understanding of the Undergraduate Biology Curriculum

Using the principles of backward design, curricular development should involve the intentional alignment of intended student learning goals, assessment, and instruction (Wiggins and McTighe, 1998). This requires first establishing learning outcomes. Program-level learning outcomes are generally defined by departmental committees of faculty based on their shared vision for what graduating students should know and be able to do. Course-level learning outcomes are generally the purview of the instructor(s) of that course; however, they should have clear links to the program-level outcomes of that department (Allen, 2004).

After establishing clear and measurable learning goals, many departments embarking on curricular reform next examine current course offerings. When asked to document "the curriculum," many may cite course catalogues and established programmatic and course-level goals. However, course offerings and course-related learning goals are subject to constant flux as faculty change and courses evolve. Furthermore, the term "curriculum" can be defined at many different levels depending on the intent, from the program level, as a set of established learning goals or national standards, all the way down to what is actually being taught and learned in individual courses. Work to create effective curricula is intended to benefit students, and so it follows that review of any curriculum is not complete without student reports and/or learning data (Erickson and Shultz, 1992). Thus, curricular development and program evaluation experts argue that the curriculum can only truly be understood after reconceptualizing it from three (or more) perspectives (Gehrke *et al.*, 1992; Ewell and Jones, 1996; Matthews and Mercer-Mapstone, 2018).

In this study, we adapt and employ a three-part curricular model to better understand the current state of the core competency curriculum in undergraduate biology courses (Erickson and Shultz, 1992; Porter and Smithson, 2001; Matthews *et al.*, 2013; Matthews and Mercer-Mapstone, 2018). The first curriculum (called the *intended*, planned, written, or designed curriculum) establishes the curricular goals, learning outcomes, or national standards explaining what students should be able to know and do after completing the curriculum. The second curriculum (called the *enacted*, taught, tested, or delivered curriculum) is how instructors translate the intended curriculum into teaching and assessment in actual courses. The third curriculum (called the *experienced* or learned curriculum) is the curriculum perceived by students in response to instruction. The most complete and accurate view of the curriculum requires examination from all three perspectives, which we refer to here as the "intended–enacted–experienced" curriculum model.

Studies comparing the multiple curricula often reveal misalignment. A comparison of syllabi and exams in undergraduate biology courses found that the level of cognitive complexity of learning outcomes stated on syllabi was generally much higher than the level measured by assessments (intended vs. enacted curricula; Momsen *et al.*, 2010). In another study, medical students perceived less teaching of cultural competencies relative to instructors, which authors propose may be explained by their finding that cultural competency, although included in course instruction, was largely absent from assessments (enacted vs. experienced curricula; Wachtler and Troein, 2003). A multidepartment survey of science, technology, engineering, and mathematics (STEM) faculty and students found that, while students' and instructors' perceived value of transferable skills aligned on many points, students were more likely to rate lower-order skills like memorization of facts and equations as important compared with higher-order skills like quantitative reasoning and scientific writing (intended vs. experienced curricula; Marbach-Ad *et al.*, 2019). Another survey of life sciences instructors and students found that students and instructors often did not align in their perceptions of whether six transferable skills were taught and assessed (enacted vs. experienced curricula; Matthews and Mercer-Mapstone, 2018). Relative to instructors, students reported more frequent instruction and assessment of all six skills, especially writing, teamwork, and ethical thinking. Finally, a study comparing biology undergraduate student performance and confidence on a quantitative skills assessment to instructor-defined goals found that students were underperforming relative to instructor expectations (intended vs. experienced curricula; Matthews *et al.*, 2016). In summary, instructor goals, classroom instruction and assessments, and student perceptions of curricula often do not align. In the absence of alignment, several things can happen: departments may be unintentionally misrepresenting curricula to accreditors, assessment data may be invalid for use to understand the effectiveness of instruction, and student motivation may be underleveraged if students do not perceive a direct connection between class activities and career-related learning goals (Dweck, 1986; Eccles and Wigfield, 2002).

Although a number of resources have been developed to help educators implement the *Vision and Change* curricular recommendations, little has been done to measure the alignment of the core concepts or competencies across the intended, enacted, and experienced curricula. Among existing resources that are aligned to *Vision and Change*, the majority focus on the core concepts (Smith *et al.*, 2019; Branchaw *et al.*, 2020). For the core competencies, the BioSkills Guide is a set of measurable program- and course-level learning outcomes developed to elaborate the *Vision and Change* core competencies (Clemmons *et al.*, 2020). Although not intentionally aligned with *Vision and Change*, a large number of valid assessments also exist to assess individual scientific competencies at the course level (e.g., Timmerman *et al.*, 2011; Gormally *et al.*, 2012; Brownell *et al.*, 2014; Dasgupta *et al.*, 2014; Stanhope *et al.*, 2017; Angra and Gardner, 2018; Hicks *et al.*, 2020; Reynders *et al.*, 2020; Tripp and Shortlidge, 2020). Finally, the Partnership for Undergraduate Life Sciences Education (PULSE) network developed a set of rubrics for departments to self-assess their progress toward fulfilling the pedagogical and curricular goals of *Vision and Change*, including teaching of the core competencies (Brancaccio-Taras *et al.*, 2016).

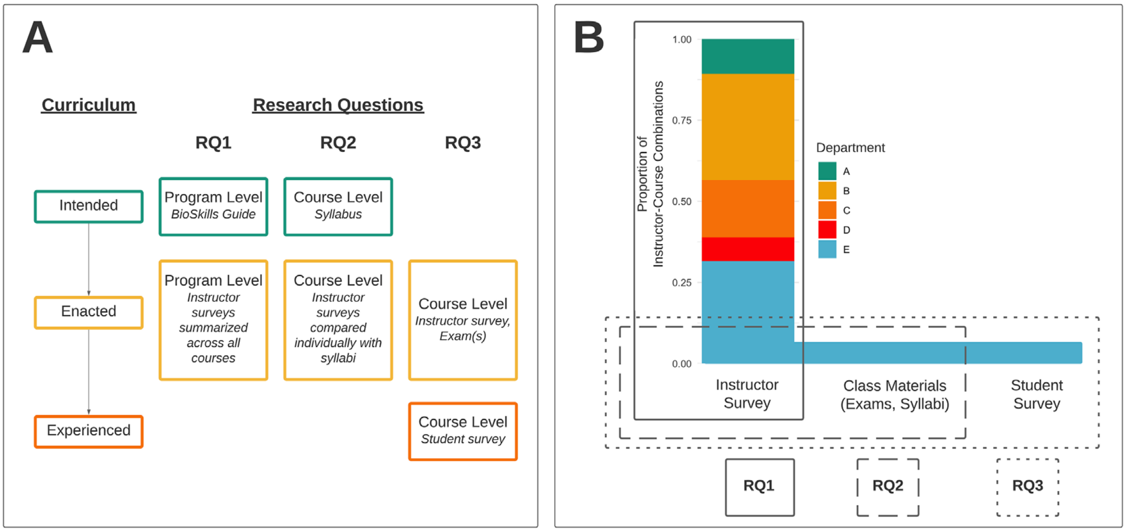


FIGURE 1. The relationship between (A) the research questions and the intended, enacted, and experienced curricula and (B) the research questions, the three types of data analyzed, and the five departments. (A) RQ1 examined the curricula at the program level, comparing results across all courses to the BioSkills Guide. RQ2 and RQ3 examined the curricula at the course level, analyzing alignment in individual courses. (B) The solid, dashed, or dotted lines around the research questions align with the same line types around the data sources. The multicolor stacked bar chart shows the distribution of instructor–course combinations across the five biology departments that participated in the pilot ($n = 187$ instructor–course combinations, defined as a particular instructor reporting on a particular course). Extended blue bar indicates that class materials and student surveys were collected from a subset of courses ($n = 10$) in department E.

In addition to using these resources, biology departments interested in adopting the *Vision and Change* curricular recommendations may want to map the core concepts and competencies onto their courses. "Curriculum mapping" refers to the process of comparing and aligning course offerings (enacted curriculum) to departmental learning outcomes (intended curriculum; Allen, 2004). Traditionally, curriculum mapping involves bringing faculty together in a room to individually report which program-level learning outcomes are addressed in each of their courses, often with an indication of depth of coverage (e.g., beginning, intermediate, mastery). The in-person data-collection portion of curriculum mapping can be logistically challenging for large departments, and there are few resources to guide departments through this process.

## Study Overview

In this study, we used the intended–enacted–experienced curriculum model as a framework to characterize how the core competency recommendations of *Vision and Change* are integrated into a sample of college biology courses. The *Vision and Change* core competencies are Process of Science, Quantitative Reasoning, Modeling, Interdisciplinary Nature of Science, Communication & Collaboration, and Science & Society. Specifically, we examined areas of alignment and misalignment among:

- the intended curriculum: programmatic- and course-level goals for core competencies;
- the enacted curriculum: core competencies being taught and assessed, from the instructor perspective; and
- the experienced curriculum: core competencies being taught, from the student perspective.

We framed our analyses using three research questions (Figure 1A).

*RQ1: Which Vision and Change Core Competencies Are Reported Taught and Assessed across Biology Departments?* We first compared the intended and enacted curricula for the core competencies at the program level. This is very similar to the comparison that is traditionally made during curriculum mapping. To simplify and standardize the process of curriculum mapping, we developed a Web survey, which we call the "BioSkills Curriculum Survey." The survey asked biology instructors about their teaching and assessment (enacted curriculum) of the 20 program-level core competency learning outcomes (two to six per competency) from the BioSkills Guide (intended curriculum). To address RQ1, we then analyzed results from a pilot survey of instructors in biology departments at five different institutions.

We used program-level learning outcomes from the BioSkills Guide as a proxy for the intended curriculum in RQ1. Although none of the five biology departments in our pilot had explicitly integrated the BioSkills learning outcomes into their departmental goals before this study, we felt the BioSkills Guide was appropriate evidence of the intended curricula for two reasons. First, the departments participated in our study because they were interested in tracking their progress toward the *Vision and Change* curricular recommendations, indicating their intent to begin the process of updating their programmatic learning outcomes and course offerings by retroactively examining their alignment. Second, the BioSkills Guide was purposefully designed to align with *Vision and Change* and was evaluated for evidence of content validity through a survey of more than 400 college biology educators (Clemmons *et al.*, 2020). In keeping with our overarching goal to support data-driven curricular transformation, the study design for RQ1 allowed us to compare alignment of the currently enacted curricula to the aspirational intended curriculum, not to existing departmental goals.

*RQ2: How Do the Core Competency Learning Outcomes That Instructors Report Teaching Compare with the Learning Outcomes They Include on Their Syllabi?*

*RQ3: What Predicts Agreement in Student and Instructor Perceptions of Teaching of the Core Competencies?* To complement our broader survey of instructors' perceptions of teaching and assessment at the program level, we next more deeply compared the intended and enacted curricula (RQ2) and the enacted and experienced curricula (RQ3) in 10 biology courses in a single department included in the pilot study from RQ1. No course can or should cover all learning outcomes in the BioSkills Guide (Clemmons *et al.*, 2020). Thus, for these analyses, we moved from examining coverage and gaps in core competency teaching and assessment to examining alignment across the three curricula within a given course, using the BioSkills Guide to standardize interpretations. In addition to using RQ1 survey responses from those 10 instructors (enacted curriculum), we analyzed syllabi (intended curriculum) and exams (enacted curriculum) and gathered students' perceptions of teaching (experienced curriculum) through a modified version of the BioSkills Curriculum Survey.

Considering all three research questions together, we aim to demonstrate a process that can support evidence-based reflection on the intended–enacted–experienced curriculum in order to produce actionable and relevant goals for curricular reform.

## METHODS

This study involved two phases (see Figure 1B). The first included the development and multi-institution administration of the BioSkills Curriculum Survey (RQ1). The second included a more targeted analysis of the intended–enacted–experienced curriculum at a single institution (RQ2 and RQ3). This study was approved by the University of Washington, Human Subjects Division as exempt (STUDY00001746).

### Development and Multi-Institution Administration of the BioSkills Curriculum Survey (RQ1)

*Development and Piloting of the BioSkills Curriculum Survey for Instructors.* The final BioSkills Curriculum Survey included three questions for each of the 20 program-level learning outcomes from the BioSkills Guide:

1. How often is [learning outcome] taught in this course? Response options were: "not taught," "one class session," "a few class sessions," "about half of class sessions," "most class sessions," or "almost every class session."
2. If the respondents indicated that the learning outcome was taught, by selecting any response other than "not taught," they were then asked: Is [learning outcome] assessed in this course? Response options were: "yes" or "no."
3. If the respondents indicated that the learning outcome was assessed, by selecting "yes," they were then asked: How is [learning outcome] assessed in this course (e.g., free-response exam questions, lab reports, homework assignments)? Respondents were provided a small textbox in which to respond in an open-ended way.

We split questions for the 20 learning outcomes into six blocks by *Vision and Change* core competency. Blocks were presented to instructors in a random order. Several additional questions about the characteristics of the course being reported on preceded questions about teaching and assessment: the course number, the biology subdisciplinary focus of the course, and the course components (e.g., lecture, lab, other) they would be reporting on. After completing all questions for one course, instructors were asked to repeat the survey for additional courses.

We used think-aloud interviews, an initial pilot in one biology department (data not shown), and a roundtable discussion at a national biology education research conference to improve the survey during its development (Supplemental Methods). We additionally evaluated the final survey using think-aloud interviews and Web probing questions that prompt survey respondents to explain their reasoning for selecting a particular response (Supplemental Methods and Supplemental Table 1). These methods provided evidence that the survey elicited valid response processes based on the intent of the survey (Tourangeau *et al.*, 2000; American Educational Research Association *et al.*, 2014).

The full questionnaire used for the instructor survey is included as Supplemental Material 5.

*Instructor Survey Participation and Data Cleaning.* We piloted the BioSkills Curriculum Survey with instructors in five biology departments at different institutions (two community colleges, two regional comprehensive universities, and one R1 research university). Departments were included on an opportunity sampling basis, via collaborations with department chairs or curricular committee chairs from each department included in the pilot (hereafter called "pilot collaborators") who volunteered to participate as part of broader curricular efforts specific to each department. One of the pilot collaborators is an author (D.A.D.). The survey was distributed via an email from the pilot collaborators that invited instructors who taught in the departments to voluntarily participate and encouraged participation by explaining how the survey data would be used to further departmental goals. Out of 176 instructors who were invited to complete the survey across the five departments, approximately 112 instructors began the survey (63.6% response rate overall). This response rate is much higher than other studies of faculty participation that use email recruitment (e.g., Moss-Racusin *et al.*, 2012; Barnes *et al.*, 2020). Each instructor reported on one or more courses, for a total of 215 instructor–course combinations (we define an "instructor–course combination" as a particular instructor reporting on a particular course) across 136 unique courses in the initial data set. After data cleaning, there were 187 instructor–course combinations in the final data set, representing 127 unique courses and approximately 108 unique instructors (Supplemental Methods and Supplemental Table 2). The number of instructor–course combinations varied from 14 to 61 per department. We carried out all survey data cleaning and analysis using R v. 4.0.2 (Wickham, 2016; R Core Team, 2018; Ram and Wickham, 2018; Wilke, 2020).

*Instructor Survey Data Recoding and Descriptive Statistics.* For the course characteristic of "level", we assigned each course as upper or lower level by cross-referencing course numbers with course catalogues and confirming the course numbering system with pilot collaborators. For example, in four of the five departments, course numbers of 100–299 indicated lower-level courses and >300 indicated upper-level courses.

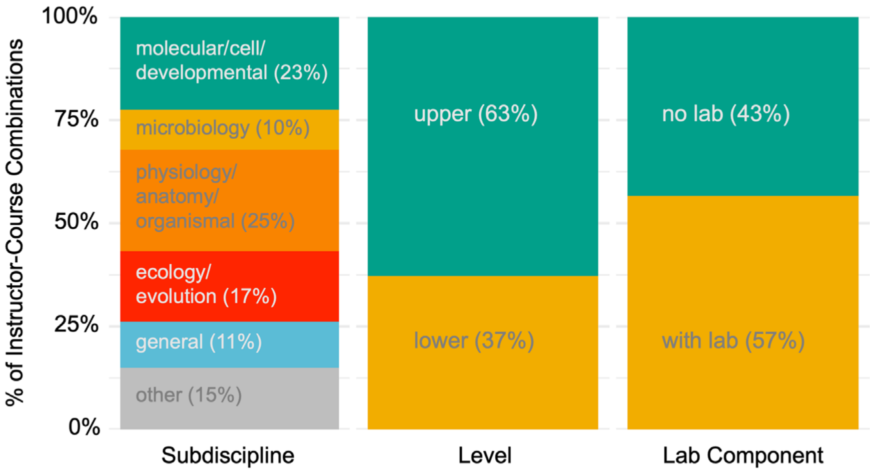


**FIGURE 2. The instructor survey data set included courses with a range of foci and characteristics (RQ1). Percentages are rounded (*n* = 187 instructor–course combinations). "Other" subdiscipline responses included integrative biology (4.8%), process of science (3.7%), biochemistry (3.2%), genetics (2.1%), and environmental/conservation/biodiversity (1.1%).**

For the course characteristic of "subdiscipline", instructors were asked to select the primary focus of the course (see Figure 2 for categories). Biochemistry was also included as a response option at one institution. Write-in responses to this question were independently evaluated by two authors (A.W.C., A.J.C.), and in some cases reassigned to an existing category. All disagreements were discussed until consensus was reached.

For the course characteristic of "lab component", we created a binary variable to indicate whether the instructor–course combination did (1) or did not (0) include reporting on a lab component. An additional option of "field" was provided for one department, which was coded as "lab" for our analyses. The survey instructed respondents to report on all components of a course they were familiar with and then asked them to indicate which course components they would be reporting on. Thus, not reporting on a lab component could be because: 1) there was no lab component associated with that course, or 2) the respondent was not sufficiently familiar with the lab component of the course and therefore chose not to report on it.

When analyzing teaching, we created a new binary variable called "reported taught" by binning responses to the question "How frequently is [learning outcome] taught in this course?" Responses of "a few," "about half," "most," or "almost every class session" were grouped together as "reported taught," and responses of "one class session" or "not taught" were grouped together as "reported not taught." When analyzing whether a learning outcome was assessed, note that the respondents were only asked "Is [learning outcome] assessed in this course?," if they said it was taught in the previous question. Thus, these analyses assume that a learning outcome was not assessed if it was not reported taught.

*Instructor Survey Assessment Mode Coding.* To analyze how different core competency learning outcomes are assessed, two authors (A.W.C., D.A.D.) qualitatively coded responses to the open-ended question "How is [learning outcome] assessed in this course?" Because respondents were only asked this question if they previously indicated that the learning outcome was assessed, summaries of assessment mode represent the percent of instructor–course combinations in which the indicated assessment mode was used out of instructor–course combinations *in which the learning outcome was assessed*. We used the constant comparative method to code responses (Glaser, 1965). Specifically, both coders reviewed a subset of responses to identify different categories of assessments and then discussed until consensus was reached on the meaning of each category. We identified five codes: exam/quiz, writing/presentation/project, practice exercise, self- and peer evaluation, and lab work (Supplemental Table 3). Responses could include multiple codes, because respondents often listed multiple assessment modes for a single learning outcome. After the coders came to consensus through iterative discussions using a training set of 171 responses, the remaining 1787 responses were divided between the two coders, including an overlapping 160 responses that were coded by both coders. We calculated Cohen's kappas for this overlapping set, which were 1, 1, 0.92, 0.91, and 0.85 for exam/quiz, writing/presentation/project, practice exercise, self/peer evaluation, and lab work codes, respectively. We deemed these interrater reliabilities to be sufficient to use independent coding for the remaining analysis (Cohen, 1960). We discussed points of discrepancy in the overlapping set until we reached consensus. The final data set included 1958 assessment mode responses.

*Instructor Survey Modeling.* We fit multilevel logistic regression models in a Bayesian framework using the rstanarm package (Lee *et al.*, 2018; Muth *et al.*, 2018). The structure and data set of each model are summarized in Table 1. Briefly, RQ1 models (models 1–5) used one of four outcome variables: whether or not a learning outcome was reported taught, reported assessed, reported assessed using writing/presentation/project, or reported assessed using exams. All models included core competency, either as the only fixed effect (models 1 and 3–5) or as an interaction with course characteristic of interest (with department in model 2a, with course level in model 2b, with whether course included a lab component in model 2c, and with course subdiscipline in model 2d). For models 2a–d, we made pairwise comparisons of each level of the interaction by systematically releveling the models to get estimates of the main effects of each level of core competency and course characteristic. All RQ1 models included three random effects: learning outcome, course, and respondent.

Models 1–2d, which all used "reported taught" as the outcome variable, included the full instructor survey data set of 187 instructor–course combinations (for a total of 3732 survey question responses about teaching frequency). Model 3, which used "reported assessed" as the outcome variable, only included instances when the learning outcome was reported taught (2662 total survey question responses about assessment). This

**TABLE 1. RQ1 and RQ3 models**

| RQ | Model no.[a] | Dependent variable | Predictors |
|---|---|---|---|
| RQ1 | 1 | Reported taught | Competency |
| | 2a | Reported taught | Competency × department |
| | 2b | Reported taught | Competency × course level |
| | 2c | Reported taught | Competency × lab component |
| | 2d | Reported taught | Competency × subdiscipline |
| | 3 | Reported assessed | Competency |
| | 4 | Reported assessed using writing/presentation/project | Competency |
| | 5 | Reported assessed using exam | Competency |
| RQ3 | 6 | Student–instructor agreement within one level | Competency |
| | 7 | Student–instructor agreement within one level | Competency, coded on syllabus, coded on exam(s), reported assessed |

[a]All models contained random effects for respondent, course, and learning outcome.

is because respondents were only asked if they assessed a learning outcome if they had said they taught it. Models 4 and 5, which used "reported assessed with a particular assessment mode" as the outcome variable, only included instances when the learning outcome was reported assessed (1860 total survey question responses about assessment mode). This is because respondents were only asked how they assessed a learning outcome if they reported assessing it.

See the Supplemental Methods, Supplemental Table 4, and Supplemental Figure 2 for additional details of model design, parameters, and evaluation of fit.

### Targeted Analysis of the Intended–Enacted–Experienced Curriculum at a Single Institution (RQ2 and RQ3)

RQ2 and RQ3 both involve the deeper analysis of an opportunity sample of 10 undergraduate biology courses in one of the departments from RQ1 (see Figure 1 and Supplemental Table 2). The pilot collaborator (also an author, D.A.D.) in that department recruited instructors to provide class materials (used for RQ2 and RQ3) and invite their students to take a modified BioSkills Curriculum Survey (used for RQ3). These data were then cross-compared with survey responses reported in RQ1 from those recruited instructors in order to address RQ2 and RQ3. Instructors participated voluntarily and were provided their anonymized and summarized course data in return.

### Class Material Coding

We collected the syllabus and all exams for each of the 10 courses. Two authors (A.W.C., A.J.C.), who had in-depth knowledge of the BioSkills Guide and no previous knowledge regarding these courses, independently coded syllabi and exams for the presence or absence of each of the 20 BioSkills program-level learning outcomes. We coded class materials in batches of two to three courses at a time and then met to discuss any ambiguities or discrepancies in coding until consensus was reached.

### Student Survey Development, Participation, and Data Cleaning

We investigated the experienced curriculum by adapting the BioSkills Curriculum Survey for students, asking about teaching of the same 20 BioSkills program-level learning outcomes, but removing questions about assessment. During one of two academic terms, instructors invited all students in their courses to participate during the final week of the academic term. Students were encouraged to participate through the inclusion of a small number of extra-credit points and the explanation that their responses could help improve the course and the program as a whole. A total of 306 students were enrolled across the 10 courses, and in total 270 student–course combinations were collected (we define a "student–course combination" as a particular student reporting on a particular course; Supplemental Table 5). The overall response rate was 88.2%, ranging from 66.7% to 100% per course. After data cleaning (as detailed in the Supplemental Methods), the final data set included 262 student–course combinations (9–59 per course), representing 215 unique students (determined using student names). Student names were excluded from all analytic data sets except when used to determine the number of unique students in the final data set.

Full text of the questionnaire used for the student survey is included as Supplemental Material 6.

*Student Survey Modeling.* See Table 1 for details of the structure and data sets of each RQ3 model. Modeling for RQ3 was carried out as described for RQ1, with the following exceptions: RQ3 models (models 6 and 7) used the binary variable "student–instructor agreement" as the outcome variable, where agreement was interpreted as a student's and the student's instructor's responses being within one response level (on the six-point teaching frequency scale) of each other for a given learning outcome. Both RQ3 models included core competency as a predictor variable, either as the only fixed effect (model 6) or as a fixed effect along with whether or not the learning outcome was included on the syllabus, included on the exam, or reported assessed (model 7). For model 7, we considered the presence of a learning outcome on one or more exams as presence on exam(s) for that course. Both RQ3 models included three random effects: respondent (i.e., student), course (which is equivalent to an instructor random effect), and learning outcome. Both RQ3 models included the full student survey data set of 262 student–course combinations, representing 5240 student survey question responses, each of which was compared with the corresponding instructor's teaching frequency response to determine student–instructor agreement.

## RESULTS

### RQ1: Which Vision and Change Core Competencies Are Reported Taught and Assessed across Biology Departments?

We used the BioSkills Curriculum Survey to compare the program-level intended curriculum with two facets of the enacted

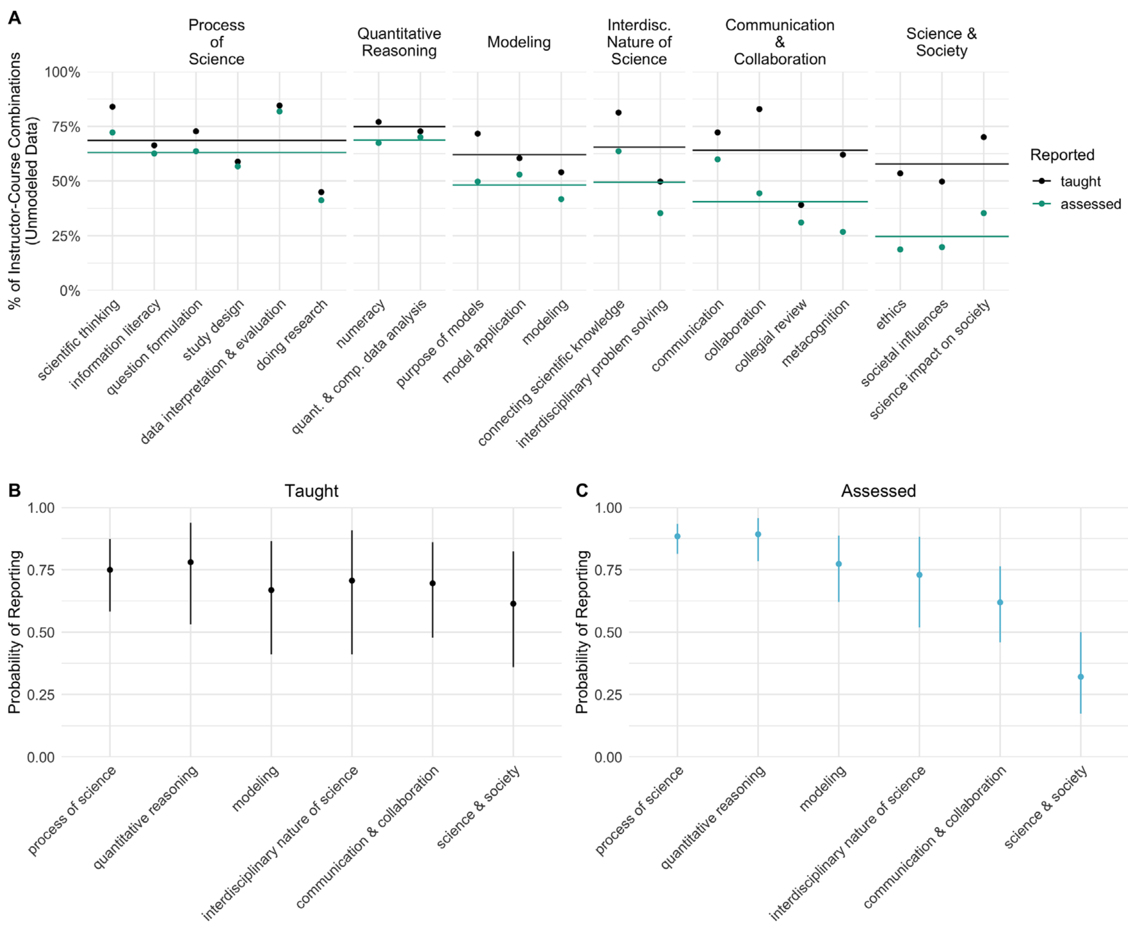


**FIGURE 3. Consistency in reported teaching but not assessment of core competencies (RQ1). (A) Unmodeled instructor survey data. Points show the percent of all instructor–course combinations ($n = 187$) for which the respondent reported teaching (black) or assessing (blue) that learning outcome. We considered a learning outcome taught if the instructor reported teaching it in "a few" or more class sessions. Lines show mean response across all learning outcomes in that core competency, treating the response scale as linear. See Supplemental Material 1 for full text of each learning outcome. (B, C) Model-based estimates of the probability of reporting teaching or reporting assessment of a learning outcome in that core competency. (C) The probability of assessment when taught, *not* the overall probability, because respondents were only asked if they assessed a learning outcome if they reported that they taught it. Points indicate means of the posterior distribution, and vertical lines indicate Bayesian 95% credible intervals.**

curriculum: instructor-reported teaching and instructor-reported assessment. Our RQ1 instructor survey data set included responses about teaching and assessment of the 20 BioSkills program-level learning outcomes from approximately 108 college biology instructors, each reporting on one or more courses, for a total of 127 unique undergraduate biology courses (Supplemental Table 2). This resulted in 187 instructor–course combinations, that is, a particular instructor reporting on a particular course. Courses represented a breadth of subdisciplinary focuses, levels, and structures (i.e., presence/absence of a lab component; Figure 2 and Supplemental Figure 1). In the following sections, we refer to the program-level outcomes by their shorthand names (e.g., Ethics, Interdisciplinary Problem Solving) for brevity; however, instructors (and students, for RQ3) were provided full text of the learning outcomes in the surveys (see Clemmons *et al.*, 2020; Supplementary Material 1 for full text).

*Which Core Competency Learning Outcomes Are Most and Least Frequently Reported Taught in Undergraduate Biology Courses across Five Departments?* To test whether instructors were more likely to say they taught learning outcomes in particular core competencies, we fit a model of the likelihood of reporting teaching with core competency as a predictor. To account for the nonlinear nature of the survey data and facilitate model interpretation, we binned teaching frequency responses into a binary variable termed "reported taught," for which responses of "one class session" or "not taught" were grouped together as "not reported taught" and responses of "a few class sessions" or greater were categorized as "reported taught" (Figure 3A). We found that there was no difference among core competencies in the probability of being reported taught (Figure 3B and Supplemental Table 6).

We were unable to model the likelihood of teaching with learning outcome as a predictor. However, by using the

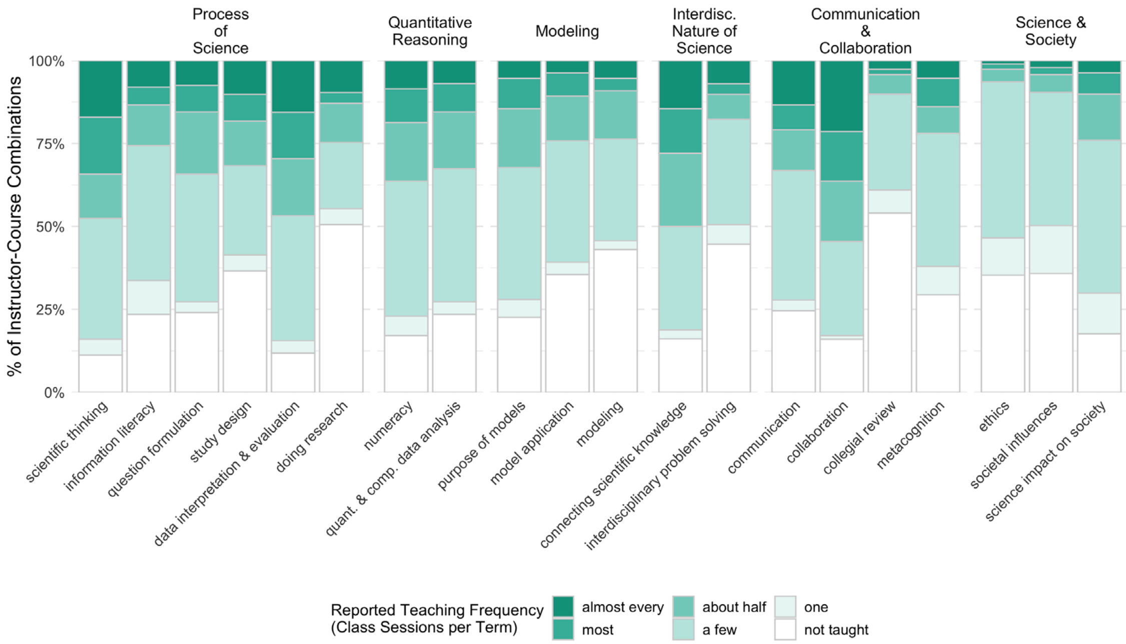


**FIGURE 4. Variation in reported levels of teaching frequency across the 20 program-level BioSkills Guide learning outcomes (RQ1). Unmodeled survey data showing the proportion of instructor–course combinations for each learning outcome, wherein each of the six frequency response options were used to answer the question "How frequently is [this learning outcome] taught in this course?" (*n* = 187 instructor–course combinations). See Supplemental Material 1 for full text of each learning outcome.**

unmodeled survey data to calculate the mean teaching frequency for each learning outcome, we observed some variation among learning outcomes within competencies that may warrant future examination. Notably, one of the Communication & Collaboration learning outcomes (Collegial Review: Provide and respond to constructive feedback in order to improve individual and team work) was among the three learning outcomes that instructors reported teaching at the lowest frequencies, whereas a different Communication & Collaboration learning outcome (Collaboration: Work productively in teams with people who have diverse backgrounds, skill sets and perspectives) was among the three learning outcomes that instructors reported teaching at the highest frequencies (Clemmons *et al*., 2020; Figure 4).

Our first model examined instructor-reported teaching across all courses in our sample. To determine whether different core competencies were more or less likely to be reported taught in particular types of courses, we fit four new, separate models of reported teaching, including the interaction of competency with: 1) department, 2) course level, 3) whether or not the instructor–course combination included reporting on a lab, and 4) subdiscipline (Supplemental Table 6 and Supplemental Figure 4; see Supplemental Figure 3 for unmodeled results). All four models revealed differences among different types of courses.

Some departments had a higher likelihood of reporting teaching Process of Science, Quantitative Reasoning, and Modeling (Supplemental Table 7). On the other hand, there were no differences between departments in the likelihood of reporting teaching Interdisciplinary Nature of Science, Communication & Collaboration, and Science & Society.

Process of Science and Science & Society were more likely to be reported taught in upper-level than lower-level courses (Table 2). Process of Science, Quantitative Reasoning, and Communication & Collaboration were more likely to be reported taught in courses with a lab component than courses without a lab component (Table 2). Conversely, Science & Society was more likely to be reported taught in courses without a lab component than courses with a lab component.

Communication & Collaboration was equally likely to be reported taught in courses across all subdisciplines (Supplemental Table 8). For all other core competencies, we observed differences between two or more subdisciplines. For those cases where differences were observed, anatomy, physiology, and/or organismal biology courses and general biology courses were less likely to report teaching core competencies relative to other subdisciplines.

*Does Alignment of Reported Teaching and Assessment Differ by Core Competency?* We next examined how likely instructors were to assess their students on the learning outcomes they reported teaching. We chose to analyze rates of reported assessment among cases in which the learning outcome was reported taught, rather than overall rates of assessment, in order to better understand whether instructors' focus in the classroom aligned with their assessment plans. We modeled reported assessment with core competency as a predictor (Supplemental Table 6). While core competency was not predictive of reported teaching when looking across all courses in our sample, we found that competency was, in fact, predictive of whether a learning outcome was reported assessed when taught (Figure 3C). Specifically, Process of Science,

TABLE 2. Pairwise comparisons of likelihood of reporting teaching in different types of courses

| Reference level | Comparison level | Process of science | Quantitative reasoning | Modeling | Interdisciplinary nature of science | Communication & collaboration | Science & society |
|---|---|---|---|---|---|---|---|
| Upper level | Lower level | Upper > lower OR: 2.1[a] | — | — | — | — | Upper > lower OR: 1.8[a] |
| With lab component | No lab component | Lab > no lab OR: 1.8[a] | Lab > no lab OR: 2.1[a] | — | — | Lab > no lab OR: 2[a] | No lab > lab OR: 1.7[a] |

[a]Text indicates direction of difference and "OR" indicates odds ratios (e.g., upper > lower OR: 2.1 indicates that upper-level courses were 2.1 times more likely to report teaching that competency than lower-level courses).

Quantitative Reasoning, Modeling, and Interdisciplinary Nature of Science learning outcomes were 17, 19, eight, and six times more likely, respectively, to be assessed when taught than Science & Society learning outcomes. Process of Science and Quantitative Reasoning learning outcomes were both five times more likely to be assessed when taught than Communication & Collaboration learning outcomes.

By calculating descriptive statistics for each learning outcome, we observed trends very similar to the modeled results (Figure 3A). All five learning outcomes with the poorest alignment between reported teaching and reported assessment—Ethics, Societal Influences, Metacognition, Science's Impact on Society, and Collaboration—were in the Communication & Collaboration and Science & Society competencies (refer to Supplemental Material 1 for full learning outcome text). These learning outcomes were reported assessed in 35–53% of instructor–course combinations for which they were reported taught. All five of the learning outcomes with the strongest alignment between reported teaching and reported assessment—Data Interpretation & Evaluation, Study Design, Quantitative & Computational Data Analysis, Information Literacy, and Doing Research—were in the Process of Science and Quantitative Reasoning competencies. These learning outcomes were reported assessed in 92–97% of cases in which they were reported taught.

*Are Instructors More Likely to Use Particular Assessment Modes for Core Competencies with Lower Alignment of Reported Teaching and Assessment?* To try to gain insight into observed differences in alignment of reported teaching and reported assessment by core competency, we qualitatively coded instructors' descriptions of assessment modes used (Supplemental Table 3). Summarizing the frequency of codes assigned across *all* learning outcomes, we found that exams and quizzes were the most commonly reported assessment type (54.3% of responses). There were also many reports of formative assessment (i.e., in- and out-of-class practice exercises, 40.3% of responses) as well as non-exam types of summative assessment (i.e., writing, presentations, and projects, 37.6%; Figure 5A).

Using modeling (Supplemental Table 6), we found no difference in which core competencies were likely to be reported assessed using writing, presentations, or projects (Figure 5B). Process of Science and Communication & Collaboration learning outcomes were less likely to be reported assessed using exams compared with Quantitative Reasoning (Figure 5C). Thus, the decreased use of the most common assessment mode (exams and quizzes) may in part explain why Communication & Collaboration is less likely to be reported assessed when reported taught relative to some other competencies. However, exams were also less likely to be used for Process of Science, a competency with strong alignment of reported teaching and reported assessment. Furthermore, a similar trend was not observed for Science & Society, one of the other competencies with low alignment between reported teaching and assessment.

### RQ2: How Do the Core Competency Learning Outcomes That Instructors Report Teaching Compare with the Learning Outcomes They Include on Their Syllabi?

We next examined syllabi from a smaller sample of courses in just one department from the pilot. Specifically, we qualitatively coded syllabi from 10 biology courses for the presence or absence of language indicating teaching of the 20 BioSkills Guide program-level learning outcomes in that course. We then compared the coded syllabi (intended curriculum) with that instructor's survey responses (enacted curriculum). Two of the 10 courses were two sections of a course-based undergraduate research experience course with shared lectures, syllabi, and exams (but separate lab sections led by separate instructors), and so the total number of syllabi examined was nine.

We found that Communication (eight of nine syllabi), Connecting Scientific Knowledge (six of nine), and Study Design (six of nine) were the learning outcomes most frequently mentioned on syllabi (Figure 6). Some learning outcomes were not represented on any of the nine examined syllabi (Scientific Thinking, Purpose of Models, Ethics, and Societal Influences) or represented on only one syllabus (Model Application and Interdisciplinary Problem Solving). Comparing instructor survey responses with syllabi, we found that learning outcomes represented on syllabi were almost always reported taught in a few or more class sessions. Among the 61 instances for which we coded core competency learning outcomes across the nine syllabi, the coded learning outcome was also reported taught in 54 instances (compare dark and light blue bars in Figure 6). A set of particular interest were learning outcomes that were reported taught but *not* represented on syllabi (light yellow bars in Figure 6). Looking at this set more closely, we found that these learning outcomes were likely to be taught at lower frequencies (i.e., "a few" or "about half of class sessions" rather than "most" or "almost all class sessions") compared with learning outcomes that were included on the syllabi (Supplemental Figure 5). In other words, learning outcomes that instructors reported teaching at the highest frequencies also tended to be represented on their syllabi.

*RQ3: What Predicts Agreement in Student and Instructor Perceptions of Teaching of the Core Competencies?* Finally, to complete the intended–enacted–experienced model and

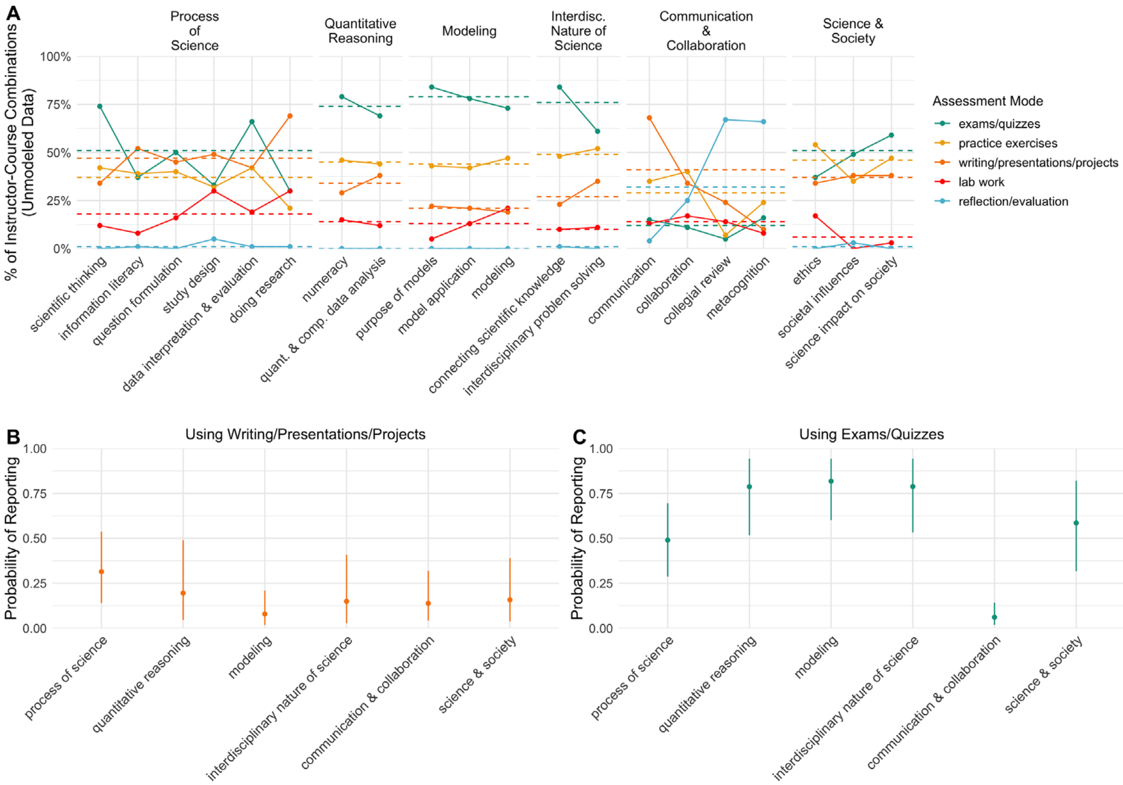


**FIGURE 5. Assessment mode varies by competency (RQ1). (A) Unmodeled instructor survey data. The percent of instructor–course combinations for which each assessment mode was reported used among instructor–course combinations for which the given learning outcome was reported assessed. Percent of instructor–course combinations is *not* out of total, because respondents were only asked how they assessed the learning outcome if they reported assessing it. Qualitative coding of responses could include multiple modes, so percentages do not sum to 100%. Connecting solid lines are included to ease tracking of individual assessment modes. Horizontal dashed lines show mean across all learning outcomes in that core competency. See Supplemental Material 1 for full text of each learning outcome. (B, C) Model-based estimates of the probability of reporting the use of writing, presentations, projects, or similar assessments (B) or exams or quizzes (C) when assessing a learning outcome in the given core competency. Points indicate means of the posterior distribution, and vertical lines indicate Bayesian 95% credible intervals.**

determine to what extent students' and instructors' views align, we administered a modified version of the BioSkills Curriculum Survey to students. This student survey data set included responses about students' perception of teaching of the 20 BioSkills program-level learning outcomes. We did not ask students about assessment. The survey was distributed during the final week of the academic term in the same subset of courses used for the RQ2 analysis. We collected responses from approximately 215 unique students across 10 college biology courses ($n$ = 9–60 per course; Supplemental Table 5). This resulted in 262 student–course combinations, that is, a particular student reporting on a particular course. We then compared student responses with their instructors' responses and asked what factors improved the likelihood that students and instructors had shared perceptions of core competency teaching.

*Does Agreement of Instructor and Student Perceptions of Teaching Vary by Core Competency?* By calculating descriptive statistics of student–instructor agreement for each learning outcome, we found that students tended to report more frequent teaching than instructors for all learning outcomes except Numeracy (Supplemental Figure 6). We next modeled the likelihood of student and instructor agreement (i.e., student and instructor responses being within one response level on the six-level teaching frequency scale) by competency. We found that students and instructors were 2.9 and 2.6 times less likely to agree on the frequency of teaching of Modeling and Interdisciplinary Nature of Science learning outcomes, respectively, than Quantitative Reasoning learning outcomes (Figure 7 and Supplemental Table 9).

*Does Representation of Core Competency Learning Outcomes in Syllabi, Exams, or Other Assessments Predict Student–Instructor Agreement That the Learning Outcome Is Taught?* Next, we asked whether the presence of a core competency on class materials, including syllabi and assessments, would increase the likelihood of student–instructor agreement. We coded all exams and the syllabus for each course for the presence or absence of each of the 20 learning outcomes (Figure 6 and Supplemental Figure 7). We also used instructor survey responses to the question "was [this learning outcome] assessed?" to account for the effect of other non-exam assessments.

We refit the previous model, including three new fixed effects in addition to core competency: presence or absence of the learning outcome on the syllabus, presence or absence of the

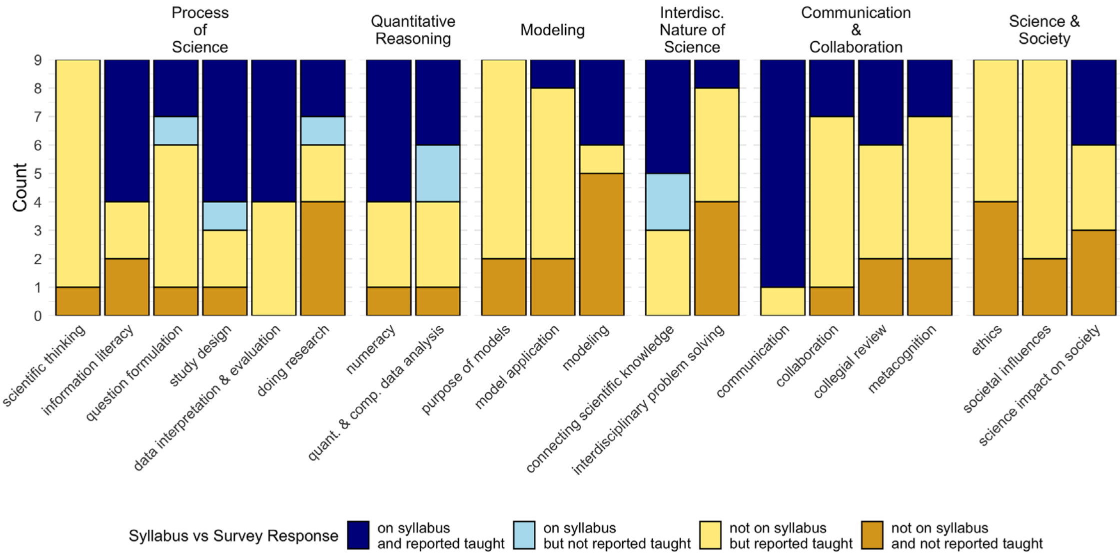


**FIGURE 6. Syllabi communicate only a subset of the learning outcomes taught in courses (RQ2). Comparison of whether an instructor reported teaching a learning outcome in a few or more class sessions and whether that learning outcome was coded on the syllabus (*n* = 10 courses; however, two courses shared a syllabus, so *n* syllabi = 9). Dark shading indicates alignment between syllabus coding and survey response. Light shading indicates lack of alignment. See Supplemental Material 1 for full text of each learning outcome.**

learning outcome on one or more of the class exams, and whether or not the instructor reported that the learning outcome was assessed. Presence of a learning outcome on an exam and instructor-reported assessment of learning outcomes increased the likelihood of student–instructor agreement by 1.8 and 1.7 times, respectively (Figure 8 and Supplemental Table 9, model 7). Presence on the syllabus, however, did not predict student–instructor agreement.

## DISCUSSION

This project was motivated by a desire to understand how and to what extent current undergraduate biology courses include the *Vision and Change* core competencies. Acknowledging the complex nature of "the curriculum," we explored core competency teaching and assessment practices using three lenses: 1) learning outcomes recommended at the program level or planned at the course level (intended curriculum), 2) learning outcomes reported by instructors as being taught and/or assessed or coded on course exams (enacted curriculum), and 3) learning outcomes reported by students as being taught (experienced curriculum). Using a newly developed curriculum mapping tool to measure the enacted and experienced curricula, we compared the three curricula to identify strengths and weaknesses in current core competency teaching and assessment practices across a sample of college biology courses. Here, we review areas of alignment and misalignment uncovered by these comparisons and call for future work exploring areas of misalignment.

### RQ1: Comparing the Intended and Enacted Curricula

*Reported Teaching Does Not Vary by Core Competency.* By looking across all core competency learning outcomes in the intended program-level curriculum, as recommended by *Vision and Change* and operationalized in the BioSkills Guide (AAAS, 2011; Clemmons *et al.*, 2020, p. S5), we found substantial evidence of alignment of the intended and enacted curricula in our sample set of biology departments. Aggregated across responses from all 108 instructors and 127 courses, core competency was not predictive of reported teaching. In other words, instructors reported that students gain exposure to one or more learning outcomes in all six core competencies, although unmodeled results suggest that variation exists in the frequency with which different learning outcomes within core competencies are taught (Figure 4). For example, while Collaboration was among the learning outcomes that were most often reported taught, other learning outcomes in the Communication & Collaboration competency (Collegial Review and Metacognition) were reported taught less frequently (refer to Supplemental Material 1 for full learning outcome text). Those interested in the instruction of particular learning outcomes may want to examine these trends more closely in the future.

Adding course characteristics to our models enabled us to observe course-dependent differences in reported core competency teaching; thus, the survey is able to detect differences in self-reported competency teaching where they exist. We found it noteworthy that lower-level courses and courses with a lab component were less likely to report teaching of Science & Society learning outcomes than upper-level courses or courses without a lab component, respectively. Further work is needed to determine whether the observed differences in instructor reports are the result of intentional choices to de-emphasize this core competency or whether less frequent teaching of Science & Society learning outcomes in these courses arose inadvertently due to a primary focus on teaching core concepts and/or technical skills, leaving less time to practice a wide range of core competencies. Science classrooms and science culture in general can be alienating to underrepresented students and lead to field-switching out of STEM (Seymour and Hunter, 2019). Therefore, it is important for departments and instructors to consider including Science & Society learning

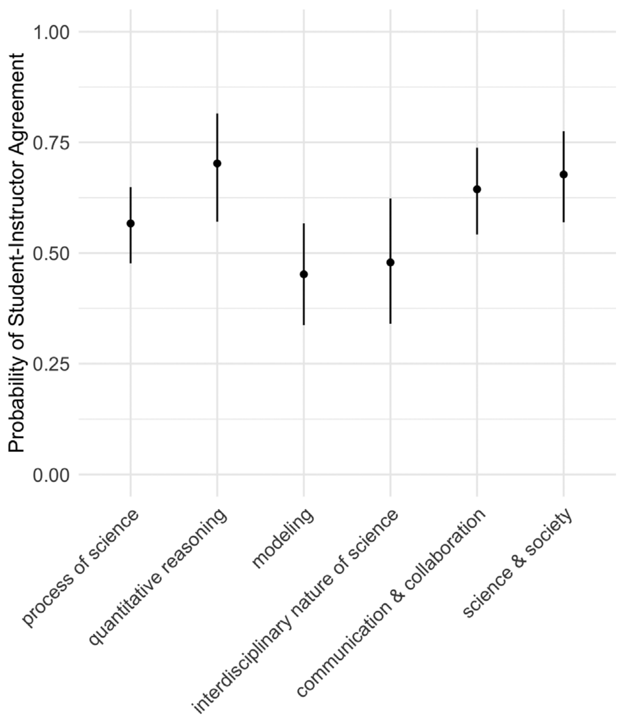


**FIGURE 7. Students and instructors are less likely to share perceptions of whether Modeling and Interdisciplinary Nature of Science were taught (RQ3). Model-based estimates of the probability of student–instructor agreement for whether learning outcomes in each core competency were taught (*n* = 262 student–course combinations, defined as a particular student reporting on a particular course). Points show means of the posterior distribution, and vertical lines show Bayesian 95% credible intervals.**

outcomes such as "Identify and describe the broader societal impacts of biological research on different stakeholders" and "Describe examples of how scientists' backgrounds and biases can influence science and how science is enhanced through diversity" (Clemmons *et al.*, 2020) in introductory-level courses as one way to invoke an asset-based approach to teaching science that creates a more inclusive environment (Estrada *et al.*, 2011, 2016; Johnson, 2019; MacSwan, 2020; Bradford *et al.*, 2021). We would argue that, in recognition of the systemic social and racial injustice that exists in STEM, including existing Science & Society learning outcomes in introductory courses is not sufficient, and we instead propose updating and expanding the *Vision and Change* recommendations to include a call for intentional integration of justice, diversity, equity, and inclusion learning outcomes throughout the undergraduate biology curriculum.

*The Alignment of Reported Teaching and Assessment Varies by Core Competency.* Similar to other reports (Matthews and Mercer-Mapstone, 2018), we found that Communication & Collaboration and Science & Society learning outcomes were less likely to be reported assessed, even though they were reported taught. There are several possible explanations for why instructors were less likely to report assessing Communication & Collaboration and Science & Society. First, instructors may be less likely to report using assessment modes that are less tangible. When asked "Is [this learning outcome] assessed in your course?," instructors may be more likely to think of assessments like exams and lab reports than informal feedback following class discussions or peer review. Nonetheless, by coding responses to the question "How is [this learning outcome] assessed in your course?," we know that at least some instructors did report using formative assessment and qualitative feedback to assess other learning outcomes.

Second, these core competencies may require more time- and resource-intensive modes of assessment. Especially for instructors who do not have support from graders or teaching assistants, it may be less feasible to grade and provide feedback on free-form assessments like writing assignments. We observed that instructors were less likely to report using exams, the most commonly reported assessment mode overall, to assess Communication & Collaboration learning outcomes; however, this was not true for Science & Society outcomes. Interestingly, we found no difference across core competencies in the likelihood of reporting assessment using writing, presentations, or projects. However, the difficulty of evaluating student performance may vary by core competency.

Finally, instructors may not assess these core competencies as much if they perceive them as lower priority. Indeed, four of the six BioSkills program-level learning outcomes (Ethics, Societal Influences, Metacognition, and Interdisciplinary Problem Solving) that had the lowest alignment between reported teaching and reported assessment in this study (Figure 3A) were also among those rated the lowest in terms of importance in a national survey of biology instructors (Clemmons *et al.*, 2020). Departments may want to use data such as these to revisit shared programmatic goals and refine assessment plans accordingly to reflect these goals.

Regardless of the reason for the observed differences, alignment between teaching and assessment is a foundation of instructional design (Wiggins and McTighe, 1998; NRC, 2000). If instructor self-reported data accurately reflect actual classroom practices, this misalignment may negatively impact student learning in these core competencies. Specifically, because Communication & Collaboration and Science & Society are competencies that prepare students to work effectively as part of a team and to be scientifically literate citizens (AAAS, 2011), misalignment in these competencies may have wide-ranging effects on students' careers and lives. Future work examining the experienced core competency curriculum using evidence of student learning could shed light on whether misalignment causes poorer learning gains.

For those who are interested in increasing the rate of assessment of Communication & Collaboration and Science & Society learning outcomes, validated assessments may be a good place to start (e.g., Rhodes, 2010; Timmerman *et al.*, 2011; Gormally *et al.*, 2012; Reynders *et al.*, 2020). However, simple dissemination of assessment tools alone is unlikely to be effective in changing assessment practices (Henderson *et al.*, 2010). Faculty may find it useful to talk with colleagues about assessment modes that have worked for them. We believe that one use of the most impactful possible uses of the BioSkills Curriculum Survey is to spur conversations among faculty about their core competency teaching and assessment practices.

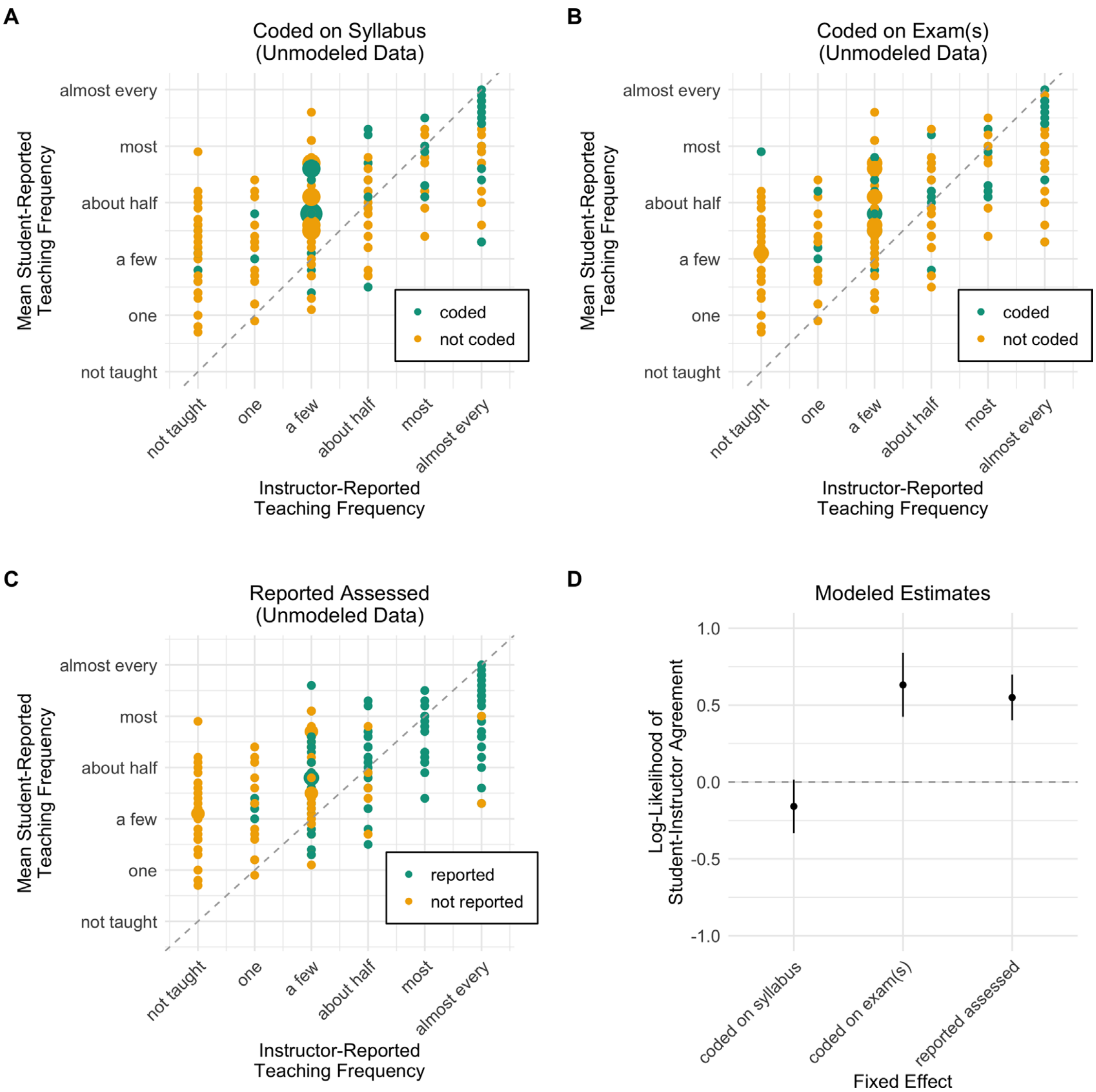


**FIGURE 8. Exams and other assessments, but not syllabi, help align student and instructor perspectives of what was taught (RQ3). (A–C) Unmodeled data in which each point is the comparison of one instructor's reported teaching frequency with the mean of the instructor's students' reported frequency for a particular learning outcome. Dashed line shows 1:1, representing perfect agreement between instructors and students. Results from 10 courses and 20 learning outcomes are shown (i.e., 200 total points on each chart in A–C). Point size is scaled by the number of student–course combinations. Points are color coded by whether or not the learning outcome was coded on the course syllabus (A), coded on one or more course exam(s) (B), or reported assessed by the instructor on the instructor survey (C). (D) Model-based estimates of the effect of inclusion of a learning outcome on class materials on the log-likelihood of student–instructor agreement, all else being equal. Points indicate means of the posterior distribution, and vertical lines indicate Bayesian 95% credible intervals. Credible intervals for "coded on exam(s)" and "reported assessed" do not cross 0, indicating these predict agreement. "Coded on syllabus" does cross 0 (credible intervals span −0.33 to 0.02), indicating no effect.**

## RQ2: Comparing the Intended and Enacted Curricula

No single biology course can or should include all six core competencies, and therefore we were also interested in how individual instructors reported (enacted) teaching compared with their intended teaching. To do this, we used syllabi from 10 biology courses in one department as a proxy for the course-level intended curriculum. By comparing learning outcomes included in syllabi with those reported taught by instructors, we found that the intended and enacted curricula aligned well in one direction: The enacted curriculum (learning outcomes reported taught) included almost all of the intended curriculum (learning outcomes coded on syllabi), but the enacted curriculum also included a number of learning outcomes that were not in the intended curriculum, as reflected by syllabi. There were only seven cases (4% of 180 comparisons of syllabi to survey reports) in which a learning outcome was represented on a syllabus and the instructor reported it was not taught. These cases likely represent learning outcomes for which the class material coders had a more expansive definition of a learning outcome than the instructor. Conversely, we observed a large number of cases (45% of 180 comparisons) in which an instructor reported a learning outcome as being taught but did not include it on the syllabus. These cases may represent learning outcomes that come up spontaneously or only briefly. While there is nothing inherently wrong with touching on skills and concepts that are not a primary focus of the course, the fact that these learning outcomes cluster in certain core competencies warrants further consideration. Learning outcomes were reported taught but not coded on the syllabi in 56% and 52% of comparisons for Science & Society and Modeling, respectively, compared with 39% and 33% of comparisons in Interdisciplinary Nature of Science and Quantitative Reasoning, respectively. Science & Society learning outcomes were also less likely to be reported assessed (when reported taught) relative to other competencies (Figure 3C). Taken together, these trends suggest that instructors are less likely to intentionally integrate Science & Society learning outcomes into enacted curricula. Thus, students may not be receiving explicit practice in these areas and/or may not recognize when they are receiving practice with these skills, which could lead to less effective learning.

### RQ3: Comparing the Enacted and Experienced Curricula

To compare the enacted and experienced curricula, we compared instructors' reported teaching frequency with their students' perceptions of teaching frequency at the end of the academic term in the same 10 courses examined for RQ2. In general, students tended to overreport core competency teaching relative to their instructors. We found that students and instructors were more likely to agree on whether Quantitative Reasoning learning outcomes were taught relative to Modeling or Interdisciplinary Nature of Science learning outcomes. Quantitative Reasoning includes learning outcomes related to the use of equations and graphs, which may be more discrete and recognizable to students than learning outcomes related to conceptual modeling or applying chemical concepts to the activity of molecules, for example.

We found that when a learning outcome was included on an exam or when the instructor reported that it was assessed (on an exam or elsewhere), students and instructors were more likely to agree on its teaching frequency. Students may be particularly attuned to the assessment portion of courses, because 1) they focus their practice more on content and skills they know will be included on assessments, and 2) they are externally motivated by grades. Inclusion of a learning outcome on the syllabus did not increase the probability of student–instructor agreement, suggesting that students may not refer to class syllabi as often as instructors would like. However, we would still have expected the core competencies stated on the syllabi to indirectly affect students' perceptions if instructors also tend to spend more time on those competencies. In cases in which instruction does not align with the syllabus, as discussed earlier, the syllabus would not be expected to increase student–instructor agreement. Future work should explore the role that syllabi play in courses and how they or other modes of communication might be used to more effectively message the intended curriculum to students.

We were not able to determine the extent to which different interpretations of learning outcomes explain areas of student–instructor agreement; however, the fact that agreement was poorest in Modeling suggests that this may be a competency for which a common vocabulary is needed. Other work examining jargon in biology courses has identified "model" as a term that students struggle with (Zukswert *et al.*, 2019). The BioSkills Guide learning outcomes were extensively reviewed and validated by college biology instructors, but their wording has not been as extensively tested for students. When looking at student survey probing question responses, we found a small number of cases in which students indicated that they did not know how to interpret a particular learning outcome. Therefore, we might have found stronger student–instructor agreement if we had used different language in some cases. However, because the BioSkills Guide learning outcomes are instructor-defined goals for students, it is worthwhile to understand the extent to which students are familiar with their instructors' language for describing core competencies.

Other curricular alignment efforts have uncovered interesting relationships among student perceptions, faculty perceptions, and class materials. For example, a study comparing STEM student and instructor ratings of the importance of five transferable skill sets found that students were more likely than faculty to value skills of lower cognitive complexity (e.g., remembering facts and formulas; Marbach-Ad *et al.*, 2019). Combined with the fact that the Bloom's levels of learning outcomes on instructors' syllabi may be higher than the Bloom's level of assessment questions (Momsen *et al.*, 2010), students may not hold the same priorities as instructors. During the validation of the Measurement Instrument for Scientific Teaching-Observable instrument, researchers compared student, instructor, and observer reports on whether various scientific teaching practices, including course design, pedagogical techniques, and some competencies, were present in a course (Durham *et al.*, 2018). Agreement was highest for scientific teaching practices that were present at lower frequencies (in that study, experimental design and data analysis). We similarly found that student–instructor agreement was greater for Science & Society, a competency for which instructors were more likely to report teaching at lower frequency levels (e.g., "a few class sessions" rather than "most class sessions"). It may be easier for multiple perspectives to agree on what is absent than on the frequency of what is taught. This finding reinforces that the BioSkills Curriculum Survey is best for documenting which core competency learning outcomes are taught and where, but not the exact frequencies with which they happen.

### Limitations

In this study we observed the intended, enacted, and experienced curricula using survey responses, class materials, and guiding documents describing national undergraduate biology curriculum recommendations. However, there are other ways to measure each of these curricular perspectives that would provide different or additional insights.

For the intended curriculum, we used the BioSkills Guide to represent national recommendations for program-level coverage of core competencies (RQ1) and instructor syllabi to represent the planned goals for a single course (RQ2 and RQ3). We address in the *Introduction* why we felt that the BioSkills Guide was an appropriate proxy for the intended curriculum for RQ1. In short, comparing the enacted and experienced curricula with national recommendations best addressed one of our overarching goals for this project: to determine the overall extent of implementation of the *Vision and Change* core competency recommendations in current undergraduate biology courses. For others who are interested in using the intended–enacted–experienced model of curriculum to look only at internal curricular alignment within a program, it would be better to use department-specific, programmatic learning goals.

For the RQ2 and 3 intended curriculum, we chose to use syllabi as a standardized way of collecting information about the intended curriculum for an individual course (rather than a whole program). Because syllabi are routinely used by departments and colleges to review new course offerings and to judge whether a course is appropriate for transfer credit or degree requirements, we reasoned that syllabi would reflect the core competencies (and core concepts) that the instructor intended to emphasize in the class. We found that instructors tended to report teaching more learning outcomes on the survey (enacted curriculum) than in their syllabi. This could be interpreted to mean that, in these cases, the competency teaching was unplanned and occurred organically in the classroom. Alternatively, these may represent cases where instructor understandings of the meaning of a learning outcome did not align with the coders' understandings. Additionally, in some cases, the syllabus may undersample the instructor's intended curriculum if there are other ways instructors communicate their course learning goals to students; for example, with daily learning objectives that are not included in the syllabus. This may be especially true if instructors only include in their syllabi those learning outcomes they plan to grade. Future studies, including instructor interviews combined with collection of all course materials, will be needed to differentiate between these possibilities.

For the enacted curriculum, we used instructor self-reports of teaching and assessment in their courses, collected via the BioSkills Curriculum Survey. We purposefully designed the survey to provide biology departments with actionable information about trends in core competency learning outcomes teaching and assessment across their programs (i.e., what is taught or assessed a lot, what is not taught or assessed very much). Surveys of this type have been shown to be an effective means of rapidly collecting curricular information in a standardized way (Porter, 2002; Desimone, 2009). However, if exact frequency measures or corroborating non–self-reported data are desired, we would recommend using a more precise, albeit labor-intensive approach, such as performing classroom observations or obtaining daily instructor logs (Porter, 2002).

For RQ1, our observations of the enacted curriculum included instructor self-reports of both teaching and assessment. Instructors were given some guidance on how to interpret "assessment" in the wording of the question: "How was [learning outcome] assessed in this course (e.g., free-response exam questions, lab reports, homework assignments)?"; however, they did not see this question unless they had answered "yes" to the preceding question "Is [learning outcome] assessed in this course?" Thus, we are unable to determine whether instructors answered the preceding question "no" because they truly did not assess a learning outcome or because they interpreted the term "assessment" more narrowly relative to other instructors. This means that assessment of some learning outcomes may be underreported relative to others. We partially addressed this limitation in the smaller data set used for RQ2 and RQ3 by corroborating instructor self-reports with exam coding. However, this may still have undersampled the learning outcomes assessed, as we did not have access to all of the assessments used in each course.

For the experienced curriculum, we used student *perceptions of what was taught* (Erickson and Shultz, 1992), as in other studies (Durham *et al.*, 2017; Mercer-Mapstone and Matthews, 2017; Marbach-Ad *et al.*, 2019). However, students may interpret what was "taught" differently than instructors, especially in cases of student-centered teaching wherein students may be less likely to classify active-learning approaches (e.g., group work, think–pair–share activities) as "teaching" compared with a traditional lecture (Wiggins *et al.*, 2017; Deslauriers *et al.*, 2019). We felt that identification of these differences in interpretation was an asset of our approach, revealing opportunities for instructors to better message the intended and enacted curricula to students. However, other researchers may want to modify the student version of the BioSkills Curriculum Survey to ask more explicitly about their experiences as learners. Furthermore, future work comparing student competency assessment data with what students and instructors report being taught would provide an additional important perspective to enrich our understanding of experienced core competency curricula.

Finally, the analyses presented in this study are from an opportunity sample of departments and courses, and thus are not generalizable. We encourage future work examining the intended–enacted–experienced curriculum in a wider range of undergraduate biology courses.

### Intended Uses for the BioSkills Curriculum Survey

Curriculum mapping is used to support a variety of practices, including standardizing instruction across multiple course sections, scaffolding instruction across a course series, building faculty consensus around programmatic goals, generating a programmatic assessment plan, updating degree requirements, deciding where to allocate funds for course updates or adding new courses, and providing evidence for accreditation (Allen, 2004; Hale, 2008; Partnership for Undergraduate Life Sciences Education, 2018). Similarly, we envision a variety of uses of the BioSkills Curriculum Survey. First and foremost, the survey data can catalyze biology departmental conversations about priorities and current practices in core competency teaching. Institutional change is only effective when faculty are engaged in the change planning and implementation processes (Henderson *et al.*, 2010). Simple top-down dissemination approaches to curricular reform are unlikely to work. Participating in the process of reporting and analyzing core competency teaching and assessment data may raise faculty awareness about—and interest in—core competencies in the biology curriculum. Departments can use data from the BioSkills Curriculum Survey to have concrete conversations about their existing enacted curricula and to generate buy-in for updating their program-level intended curricula. Indeed, during survey development, interviewees reported that simultaneously reflecting on both teaching and assessment of the learning outcomes was a valuable exercise for clarifying their course-level intended curricula. This sentiment was later echoed in feedback from pilot collaborators and in end-of-survey comments provided by respondents in the five-department pilot.

We hope that departments will complement their use of the BioSkills Curriculum Survey with other curricular analyses, including evidence of the experienced curriculum and other approaches to understanding the enacted curriculum. The most useful supplemental evidence for departments to collect will depend on their motivations and goals. More broadly, we hope future work will build on our findings by examining areas of alignment and misalignment across the intended, enacted, and experienced curricula in other settings and using other approaches. This will lay the groundwork to begin identifying ways to improve alignment across all three curricula, with the intention of making core competency instruction more coherent and effective for biology undergraduates.

## ACKNOWLEDGMENTS

This project was funded by the National Science Foundation (DUE 1710772). We thank Jerry Timbrook and the University of Washington (UW) Department of Biology Undergraduate Program Committee for constructive conversations about the initial drafting of the curriculum mapping survey. Thank you to Sara Brownell, Jenny McFarland, Erika Offerdahl, and the UW Biology Education Research Group for their continued feedback and assistance throughout this project. We thank Robin Altman, Alex Doetsch, Kelly McDonald, Heather McGray, and Mark Mort for coordinating the pilot survey administration. We thank instructors in the California State University Sacramento Department of Biological Sciences, the College of Southern Idaho Biology Program, the San Diego City College Life Sciences Department, the University of Kansas Undergraduate Biology Program, the UW Department of Biology, and the Western Washington University Biology Department for completing various versions of the curriculum survey. We thank Sara Brownell, Erika Offerdahl, Emily Scott, and anonymous reviewers for constructive feedback that greatly improved the article.

## REFERENCES

Allen, M. J. (2004). *Assessing Academic Programs in Higher Education*. Bolton, MA: Anker Publishing Company, Inc. Retrieved July 22, 2019, from www.ankerpub.com

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC. Retrieved December 15, 2021, from visionandchange.org

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Angra, A., & Gardner, S. M. (2018). The Graph Rubric: Development of a teaching, learning, and research tool. *CBE—Life Sciences Education*, *17*(4), ar65. https://doi.org/10.1187/cbe.18-01-0007

Barnes, M., Truong, J. M., Grunspan, D. Z., & Brownell, S. E. (2020). Are scientists biased against Christians? Exploring real and perceived bias against Christians in academic biology. *PLoS ONE*, *15*(1), e0226826. https://doi.org/10.1371/journal.pone.0226826

Bradford, B. C., Beier, M. E., & Oswald, F. L. (2021). A meta-analysis of university stem summer bridge program effectiveness. *CBE—Life Sciences Education*, *20*(2). https://doi.org/10.1187/cbe.20-03-0046

Brancaccio-Taras, L., Pape-Lindstrom, P., Peteroy-Kelly, M., Aguirre, K., Awong-Taylor, J., Balser, T., … & Zhao, J. (2016). The PULSE *Vision & Change* Rubrics, Version 1.0: A valid and equitable tool to measure transformation of life sciences departments at all institution types. *CBE—Life Sciences Education*, *15*(4), ar60. https://doi.org/10.1187/cbe.15-12-0260

Branchaw, J. L., Pape-Lindstrom, P. A., Tanner, K. D., Bissonnette, S. A., Cary, T. L., Couch, B. A., … & Brownell, S. E. (2020). Resources for teaching and assessing the *Vision and Change* biology core concepts. *CBE—Life Sciences Education*, *19*(2), es1. https://doi.org/10.1187/cbe.19-11-0243

Brownell, S. E., Wenderoth, M. P., Theobald, R., Okoroafor, N., Koval, M., Freeman, S., … & Crowe, A. J. (2014). How students think about experimental design: Novel conceptions revealed by in-class activities. *BioScience*, *64*(2), 125–137. https://doi.org/10.1093/biosci/bit016

Clemmons, A. W., Timbrook, J., Herron, J. C., & Crowe, A. J. (2020). BioSkills Guide: Development and national validation of a tool for interpreting the *Vision and Change* core competencies. *CBE—Life Sciences Education*, *19*(4), ar53. https://doi.org/10.1187/cbe.19-11-0259

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Dasgupta, A. P., Anderson, T. R., & Pelaez, N. (2014). Development and validation of a rubric for diagnosing students' experimental design knowledge and difficulties. *CBE—Life Sciences Education*, *13*(2), 265–284. https://doi.org/10.1187/cbe.13-09-0192

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences USA*, *116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

Durham, M. F., Knight, J. K., Bremers, E. K., DeFreece, J. D., Paine, A. R., & Couch, B. A. (2018). Student, instructor, and observer agreement regarding frequencies of scientific teaching practices using the Measurement Instrument for Scientific Teaching-Observable (MISTO). *International Journal of STEM Education*, *5*(1), 31. https://doi.org/10.1186/s40594-018-0128-1

Durham, M. F., Knight, J. K., & Couch, B. A. (2017). Measurement Instrument for Scientific Teaching (MIST): A tool to measure the frequencies of research-based teaching practices in undergraduate science courses. *CBE—Life Sciences Education*, *16*(4), ar67. https://doi.org/10.1187/cbe.17-02-0033

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040–1048. https://doi.org/10.1037/0003-066X.41.10.1040

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Erickson, F., & Shultz, J. (1992). Student's experience of the curriculum. In Jackson, P. W. (Ed.), *Handbook of research on curriculum* (pp. 465–487). New York: Macmillan. Retrieved January 1, 2021, from https://stars.library.ucf.edu/cirs/2237

Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutiérrez, C. G., … & Zavala, M. (2016). Improving underrepresented minority student persistence in STEM. *CBE—Life Sciences Education*, *15*(3), es5. https://doi.org/10.1187/cbe.16-01-0038

Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of Educational Psychology*, *103*(1), 206–222. https://doi.org/10.1037/a0020743

Ewell, P. T., & Jones, D. P. (1996). *Indicators of "good practice" in undergraduate education: A handbook for development and implementation*. Boulder, CO: National Center for Higher Education Management Systems.

Gehrke, N. J., Knapp, M. S., & Sirotnik, K. A. (1992). In search of the school curriculum. *Review of Research in Education*, *18*, 51. https://doi.org/10.2307/1167297

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, *12*(4), 436–445. https://doi.org/10.2307/798843

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE—Life Sciences Education*, *11*(4), 364–377. https://doi.org/10.1187/cbe.12-03-0026

Grunspan, D. Z., Nesse, R. M., Barnes, M. E., & Brownell, S. E. (2018). Core principles of evolutionary medicine: A Delphi study. *Evolution, Medicine* *and Public Health,, 2018*(1), 13–23. https://doi.org/10.1093/emph/eox025

Hale, J. A. (2008). *A guide to curriculum mapping: Planning, implementing, and sustaining the process*. Thousand Oaks, CA: Corwin Press. Retrieved January 1, 2021, from https://us.sagepub.com/en-us/nam/book/guide-curriculum-mapping?page=1

Henderson, C., Finkelstein, N., & Beach, A. (2010). Beyond dissemination in college science teaching: An introduction to four core change strategies. *Journal of College Science Teaching*, *39*(5), 18–25. https://doi.org/10.2505/3

Hicks, J., Dewey, J., Brandvain, Y., & Schuchardt, A. (2020). Development of the Biological Variation in Experimental Design and Analysis (BioVEDA) assessment. *PLoS ONE*, *15*(7), e0236098. https://doi.org/10.1371/journal.pone.0236098

Johnson, K. M. S. (2019). Implementing inclusive practices in an active learning STEM classroom. *Advances in Physiology Education*, *43*(2), 207–210. https://doi.org/10.1152/ADVAN.00045.2019

Lee, J., Sim, N., Ji, F., & Rabe-Hesketh, S. (2018, April 24). Introduction to multilevel modeling using rstanarm: A tutorial for education researchers. Retrieved September 25, 2020, from https://mc-stan.org/users/documentation/case-studies/tutorial_rstanarm.html#posteriormeans-posterior-standard-deviations-95-credible-intervals-and-monte-carlo-errors52

MacSwan, J. (2020). Academic English as standard language ideology: A renewed research agenda for asset-based language education. *Language Teaching Research*, *24*(1), 28–36. https://doi.org/10.1177/1362168818777540

Marbach-Ad, G., Hunt, C., & Thompson, K. V. (2019). Exploring the values undergraduate students attribute to cross-disciplinary skills needed for the workplace: An analysis of five STEM disciplines. *Journal of Science Education and Technology*, *28*(5), 452–469. https://doi.org/10.1007/s10956-019-09778-8

Matthews, K. E., Adams, P., & Goos, M. (2016). Quantitative skills as a graduate learning outcome of university science degree programmes: Student performance explored through the planned–enacted–experienced curriculum model. *International Journal of Science Education*, *38*(11), 1785–1799. https://doi.org/10.1080/09500693.2016.1215568

Matthews, K. E., Hodgson, Y., & Varsavsky, C. (2013). Factors influencing students' perceptions of their quantitative skills. *International Journal of Mathematical Education in Science and Technology*, *44*(6), 782–795. https://doi.org/10.1080/0020739X.2013.814814

Matthews, K. E., & Mercer-Mapstone, L. D. (2018). Toward curriculum convergence for graduate learning outcomes: Academic intentions and student experiences. *Studies in Higher Education*, *43*(4), 644–659. https://doi.org/10.1080/03075079.2016.1190704

Mercer-Mapstone, L. D., & Matthews, K. E. (2017). Student perceptions of communication skills in undergraduate science at an Australian research-intensive university. *Assessment and Evaluation in Higher Education*, *42*(1), 98–114. https://doi.org/10.1080/02602938.2015.1084492

Merkel, S., & ASM Task Force on Curriculum Guidelines for Undergraduate Microbiology. (2012) The development of curricular guidelines for introductory microbiology that focus on understanding. *Journal of Microbiology & Biology Education*, *13*(1), 32–38. https://doi.org/10.1128/jmbe.v13i1.363

Michael, J., Modell, H., McFarland, J., & Cliff, W. (2009). The "core principles" of physiology: What should students understand? *Advances in Physiology Education*, *33*(1), 10–16. https://doi.org/10.1152/advan.90139.2008

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE—Life Sciences Education*, *9*(4), 435–440. https://doi.org/10.1187/cbe.10-01-0001

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences USA*, *109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109

Muth, C., Oravecz, Z., & Gabry, J. (2018). User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *Quantitative Methods for Psychology*, *14*(2), 99–102. https://doi.org/10.20982/tqmp.14.2.p099

National Research Council. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press. https://doi.org/10.17226/9853

NRC. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press. https://doi.org/10.17226/13165

Partnership for Undergraduate Life Sciences Education. (2018). Northwest PULSE curriculum mapping webinar. Retrieved January 1, 2021, from www.pulse-community.org/nwwebinars

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, *31*(7), 3–14. Retrieved September 10, 2019, from www.project2061.org

Porter, A. C. G., & Smithson, J. L. (2001). Defining, developing, and using curriculum indicators, *CPRE Research Report Series, RR-048*, The Consortium for Policy Research in Education, University of Pennsylvania Graduate School of Education. Retrieved December 15, 2021, from https://www.cpre.org/sites/default/files/researchreport/788_rr48.pdf

Ram, K., & Wickham, H. (2018). *wesanderson: A Wes Anderson palette generator (R package version 0.3.6)*. Retrieved March 13, 2020, from https://cran.r-project.org/package=wesanderson

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved December 15, 2021, from www.r-project.org

Reynders, G., Lantz, J., Ruder, S. M., Stanford, C. L., & Cole, R. S. (2020). Rubrics to assess critical thinking and information processing in undergraduate STEM courses. *International Journal of STEM Education*, *7*(1), 9. https://doi.org/10.1186/s40594-020-00208-5

Rhodes, T. (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: Association of American Colleges and Universities.

Seymour, E., & Hunter, A. B. (2019). *Talking about leaving revisited: Persistence, relocation, and loss in undergraduate STEM education*. Switzerland: Springer International Publishing.

Smith, M. K., Brownell, S. E., Crowe, A. J., Holmes, N. G., Knight, J. K., Semsar, K., … & Couch, B. A. (2019). Tools for change: Measuring student conceptual understanding across undergraduate biology programs using Bio-MAPS assessments. *Journal of Microbiology & Biology Education*, *20*(2). https://doi.org/10.1128/jmbe.v20i2.1787

Stanhope, L., Ziegler, L., Haque, T., Le, L., Vinces, M., Davis, G. K., … & Overvoorde, P. J. (2017). Development of a Biological Science Quantitative Reasoning Exam (BioSQuaRE). *CBE—Life Sciences Education*, *16*(4), ar66. https://doi.org/10.1187/cbe.16-10-0301

Talanquer, V., Bucat, R., Tasker, R., & Mahaffy, P. G. (2020). Lessons from a pandemic: Educating for complexity, change, uncertainty, vulnerability, and resilience. *Journal of Chemical Education*, *97*(9), 2696–2700. https://doi.org/10.1021/acs.jchemed.0c00627

Tansey, J. T., Baird, T., Cox, M. M., Fox, K. M., Knight, J., Sears, D., & Bell, E. (2013). Foundational concepts and underlying theories for majors in "biochemistry and molecular biology." *Biochemistry and Molecular Biology Education*, *41*(5), 289–296. https://doi.org/10.1002/bmb.20727

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a "universal" rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, *36*(5), 509–547. https://doi.org/10.1080/02602930903540991

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511819322

Tripp, B., & Shortlidge, E. E. (2020). From theory to practice: Gathering evidence for the validity of data collected with the Interdisciplinary Science Rubric (IDSR). *CBE—Life Sciences Education*, *19*(3), ar33. https://doi.org/10.1187/cbe.20-02-0035

Wachtler, C., & Troein, M. (2003). A hidden curriculum: Mapping cultural competency in a medical programme. *Medical Education*, *37*(10), 861–868. https://doi.org/10.1046/j.1365-2923.2003.01624.x

Wickham, H. (2016). *tidyverse: Easily install and load the "Tidyverse." (R package version 1.2.1)*. Retrieved December 15, 2021, from https://cran.r-project.org/package=tidyverse

Wiggins, B. L., Eddy, S. L., Wener-Fligner, L., Freisem, K., Grunspan, D. Z., Theobald, E. J., … & Crowe, A. J. (2017). ASPECT: A Survey to Assess Student Perspective of Engagement in an Active-Learning Classroom. *CBE—Life Sciences Education*, *16*(2). https://doi.org/10.1187/cbe.16-08-0244

Wiggins, G., & McTighe, J. (1998). What is backward design? *In Understanding by Design* (pp. 7–19). https://doi.org/10.1016/j.cie.2006.03.005

Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2." (R package version 1.1.0)*. Retrieved December 15, 2021, from https://cran.r-project.org/package=cowplot

Zukswert, J. M., Barker, M. K., & McDonnell, L. (2019). Identifying troublesome jargon in biology: Discrepancies between student performance and perceived understanding. *CBE—Life Sciences Education*, *18*(1), ar6. https://doi.org/10.1187/cbe.17-07-0118