Research
DOI: https://doi.org/10.1128/jmbe.v19i3.1396
JMBE Journal of Microbiology & Biology Education

Designing a Curriculum-Aligned Assessment of Cumulative Learning about Marine Primary Production to Improve an Undergraduate Marine Sciences Program †

Ryan A. Weatherbee[1]* and Sara M. Lindsay[2]
[1]*Husson University, Bangor, ME 04401;* [2]*School of Marine Sciences and Maine Center for Research in STEM Education, University of Maine, Orono, ME 04469-5741*

We developed an assessment to track changes in understanding about marine primary production, a key concept taught across our undergraduate curriculum. Question content was informed by investigating student misunderstandings, conducting faculty interviews, and mapping primary production concepts to the curriculum. Content questions were paired with questions asking students how confident they were in their answers. Although students gained knowledge of marine primary production across educational levels, confidence data and item analysis indicated student misunderstandings on several concepts. Many students had difficulty on questions that required interpreting graphs or other higher-order thinking skills. The results set the stage for additional focused assessment and curriculum revision, and the questions may be useful in developing a large-scale, interdisciplinary marine sciences concept inventory.

INTRODUCTION

The oceans define our planet, supporting and shaping life on earth. Many challenges facing society involve the oceans, whether related to changing climate, fisheries, ocean health, or coastal development and human populations. Meeting these challenges requires citizens literate about the ocean and science, and improving science literacy has been a national educational priority (1–3). The framework for Ocean Science Literacy (http://oceanliteracy.wp2.coexploration.org) identifies essential principles and fundamental concepts to promote public understanding of the oceans and guide incorporation of ocean sciences content in K–12 curricula. Many of these concepts are included in marine science undergraduate curricula, including the Marine Sciences major at the University of Maine (UM). Although the framework identifies key concepts in ocean sciences, it is not a concept inventory that could be used to evaluate how student understanding of ocean science concepts changes over time. Indeed, comparatively little education research exists regarding teaching and learning of ocean science concepts (4, 5), in contrast to large bodies of education research in chemistry, physics, and biology (6–10). Because ocean sciences are interdisciplinary, encompassing biology, chemistry, geology, mathematics, oceanography, and physics, undergraduate programs seeking to develop program-level assessments of their curricula could deploy multiple concept inventories from the different disciplines (e.g., for biological sciences [11–14], chemistry [15–19], the recently published oceanography concept inventory [5], or the community-based GeoSciences Concept Inventory GCI 2.0 [20, 21]). However, many of these concept inventories intentionally target introductory levels and tend to be conceptually extensive, focusing on a wide range of topics (*sensu* 5), while others (e.g., Developmental Biology Content Assessment, [22]) may present concepts using examples/language less familiar to students learning about the topic in a marine context. Another approach is to develop a series of conceptually intensive surveys that assess whether students show evidence of stepwise learning about specific concepts taught across a curriculum.

We developed a survey focused on a single, highly important concept in ocean sciences: marine primary production. Primary production is the synthesis of organic carbon (i.e., chemical energy in the form of simple carbohydrates) from inorganic carbon (e.g., CO_2) by living organisms. In marine systems, primary producers include chemosynthetic and photosynthetic microbes, as well as macroalgae and marine plants. A complete understanding of this topic requires students to integrate knowledge from biological, chemical, physical and oceanographic perspectives. Students also must apply quantitative skills in graph interpretation that correspond to the ways that scientists investigate and communicate research about marine primary production.

*Corresponding author. Mailing address: Husson University, 1 College Circle, Bangor, ME 04401. Phone: 207-941-7031. E-mail: WeatherbeeR@husson.edu.
Received: 31 July 2017, Accepted: 22 August 2018, Published: 14 December 2018.
†Supplemental materials available at http://asmscience.org/jmbe

Thus, our construct included not just content knowledge about marine primary production but also the ability to interpret relevant graphical information. By focusing on a central topic in our curriculum, we aimed to inform undergraduate curriculum review and development and pilot a potential model for future program assessment.

In the UM School of Marine Sciences (SMS) assessment occurs primarily at the course level. Courses that meet departmental learning outcomes have been identified, but when we conducted this project, there was no effort in place to collect or analyze course-level or program-level student outcomes data. There was, therefore, little direct evidence that departmental outcomes were being met. We developed a curriculum-aligned assessment to determine whether students gained knowledge about marine primary production across the curriculum and to identify any persistent misunderstandings about this topic.

METHODS

We developed the marine primary production assessment (MPPA) using a mixed methodology approach that included mining existing student work to identify possible misunderstandings on the topic, interviews with faculty to better understand the curricular focus given to primary production, several rounds of question development, revision and faculty feedback, pilot testing, and student interviews (Fig. 1, details in Appendix 1). The MPPA included 14 questions that targeted concepts about which students had apparent misunderstandings and that were taught at different levels across the curriculum (Table 1 and see concept map, Appendix 2). The question topics are briefly described in Table 1; questions are grouped into subtopics identified from the implemented curriculum concept map (Appendix 2). The MPPA (Appendix 3) was administered electronically, and responses from 86 students who attempted all survey questions were analyzed. Details about the deployment and this population are provided in Appendix 1. All research was supervised and approved by the UM Institutional Review Board for Human Subjects Research (IRB #2008-12-19).

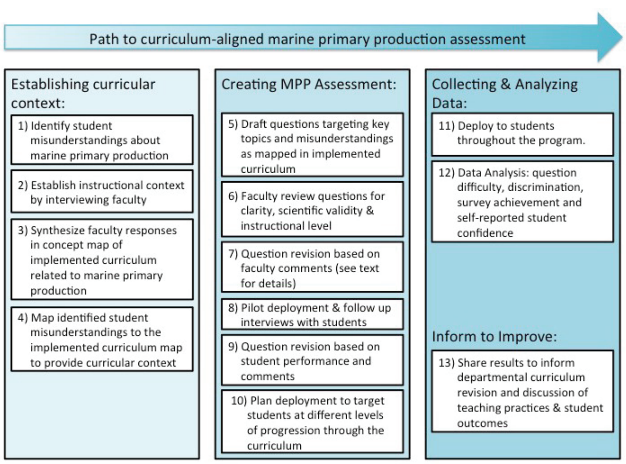

FIGURE 1. Flowchart describing general assessment development process. Potential misunderstandings about marine primary production were identified from existing first-year student homework and then mapped onto a concept map describing the implemented curriculum regarding marine primary production; assessment questions were drafted to target key topics for which we had evidence of some misunderstandings. Faculty experts guided question development, and some questions were revised based on student performance and comments following a pilot deployment. See Appendix 1 for details. MPP = marine primary production.

We used Rasch analysis (eRm package [23] within the R statistical computing software) to model how the responses to the MPPA questions represented an ordered acquisition of knowledge and skills associated with the construct, marine primary production. The Rasch model (24) assumes that any interaction between assessment questions and responses is determined by the difficulty of the questions and the ability of the person responding. Results provide information about question difficulty and how well each question fits the model of step-wise progression in knowledge (i.e., the item-parameter). In a Rasch analysis, question difficulty is the latent dimension output and is described on a log likelihood (logit) interval scale that allows us to estimate how much more difficult one question is than another (rather than just the rank order). The logit scale has negative and positive values, with easier questions having more negative values and harder questions having more positive values. A question with a logit score of 0 indicates a question of average difficulty to our population. Similar metrics describe student ability and how well each student fit the model of stepwise progression of knowledge gain (i.e., the person-parameter). The higher a student's ability relative to the difficulty of a question, the greater the probability of them getting the correct response. We also examined question discrimination (D) and performed item analyses to assess functioning of distractors in each question (25).

We assessed gross changes in knowledge about primary production by calculating the cumulative assessment score (as mean logit person-parameter value) of students in each educational cohort (i.e., educational levels 1 to 4). Educational levels were based on progression though required classes in the department and generally corresponded to first-year through senior levels. However, due to flexibility in the timing of the curriculum (i.e., upper division oceanography classes might be taken either junior or senior year), some departures from this pattern could occur (see Appendix 1, Table A1.1). Meaningful score differences between educational levels were determined by ANOVA and Tukey's least significant difference (LSD) post-hoc test. Due to low sample numbers and the pilot nature of this project, we chose an *a priori* alpha value of 0.1 for all statistical tests. Our goal was to identify potential patterns that will need further, more robust, testing with larger sample numbers to confirm any revealed patterns.

To gather evidence for changes in critical thinking skills across the undergraduate program, all survey questions were assigned a Bloom's Taxonomic level (Table 1) (26–28) that

TABLE 1.
MPPA question content.

Survey Question	Question Description	Instructional Category	Bloom's Taxonomy Level	Educational Level Introduced	Target Educational Level
1	Sources of biomass for marine primary producers	Photosynthesis/ respiration	2	FY	SO
2	Measures of primary production	Photosynthesis/ respiration	2	FY	FY
3	Influences of decreased light levels on production	Photosynthesis/ respiration	3	FY	FY
4	Competition for light	Ecology/biology	2	FY	J
5	Critical depth	Photosynthesis/ respiration	4	FY, SO, J[b]	S
6	Common photosynthetic pigments in eukaryotes	Photosynthesis/ respiration	1	FY, J[b]	J
7a&b	Stratification and phytoplankton bloom timing	Physical forcing	4,4	J/S, J/S	S
8[a]	Characteristics of trophic levels	Ecology/biology	2	FY	NA
9	El Niño – physical forcing of primary production	Ecology/biology	2.5	J/S	J
10	Energy transfer efficiencies	Ecology/biology	2	FY, SO	SO
11	Light attenuation and photosynthetic pigments	Photosynthesis/ respiration	NA	FY, J/S[b]	NA
12a&b[a]	Eutrophication and microbial decomposition	Nutrients	4,4.5	J/S, J/S	J, S

Question topics and primary alignment to curricular categories identified by faculty. Questions were assigned to a Bloom's taxonomy level as an indicator of cognitive demands of each question. Topics are introduced at different levels in the curriculum (i.e., Educational level introduced, see concept map in Appendix 2). Target educational level indicates the level faculty felt the question would be appropriate for.
[a] Open response question.
[b] Components of this question are introduced at different levels in the curriculum.
FY = first year; SO = sophomore; J = Junior; S = Senior; J/S = topic introduced in a class that may be taken Junior or Senior year; NA = not assessed. Q11 excluded from analysis.

designated the cognitive demands needed to obtain the correct response (knowledge through evaluation.). Two raters used the same protocol to assign a Bloom's level to each question; their scores for each question were either identical or within one Bloom level value on a Bloom's Taxonomy scale of one to six, and after discussion, the raters agreed on the levels presented in Table 1. We examined changes in the distribution of correct responses by Bloom's Taxonomic level for the full student group and by educational level using ANOVA.

Confidence data provide a context against which to judge achievement results. We asked how confident students were of their responses via a follow-up question immediately after each content question: "How sure are you of your answer to the question you just completed?" Students could select one response on a five-point Likert-like scale ranging from "I guessed" to "Extremely sure." We examined relationships between question confidence and survey score and between the mean student confidence for each question and the Bloom's level.

RESULTS

Question and item analysis

The Rasch analysis results indicate that the MPPA was moderately difficult for students, as indicated by an overall mean student ability value (not including question 11, see below) of -0.71 logits (±0.79 SD). Some questions were easier for students than others (e.g., Question 10 [Q10], Q2, Q3, Q4), with logit difficulty values less than -0.4 (Fig. 2a, see Appendix 3 for questions and student response distributions). The most difficult questions were Q5 and Q12b, with logit difficulty values of greater than 1.2. After deployment, we found wording flaws in Q11 and it was excluded from all analyses. Based on the log likelihood scaling of the difficulty parameter, we found that the hardest question for students (Q5) was twice as difficult as was the easiest question (Q10). The MPPA questions fit the Rasch model of step-wise increase in understanding. None of the questions fell outside the <-2 or >2 statistical limits (based on the standardized *t*-distribution) of the model that would indicate poorly fitting questions (Fig. 2a). However, the gap in the difficulty progression (y-axis) for moderately difficult questions (i.e., gap below questions 5 and 12b on Fig. 2a) indicates additional questions are needed within the difficulty range of ~0.5 to 1 logits. The person-parameter results (Fig. 2b) indicate student ability on the assessment; all but one student fit the model. That student was underfitting the model and their responses were noisy or erratic compared with what would be expected for a step-wise progression of understanding. Interestingly, this student had overall course credits equivalent to a senior but was just beginning the SMS major courses, suggesting they had recently transferred into the program when they took the assessment, which could help explain why their responses underfit the Rasch model. In addition, the underfitting student's GPA was 2.12 while the mean GPA for the study group was 3.05. However, GPA was not a good predictor of performance on the assessment (R^2 = 0.023, p = 0.197). Removing this student from the Rasch analysis had little effect on the results; the order of the questions in the item map remained the same and no additional students underfit the model in the person map.

Most survey questions had moderate (+0.1 to +0.3) or good (>+0.3) discriminating power (Table 2) based on recommendations by Nitko and Brookhart (25). All but one question had positive discrimination index values and thus distinguished high from low performers. Q7b had a negative question discrimination index value, indicating more low-performing students on the survey got it right than high-performing students. Possibly this occurred because Q7b was a follow-up to Q7a (Appendix 3), and students could answer correctly by selecting the graph consistent with their answer to Q7a even if their answer to Q7a was incorrect.

Item analysis (Table 2) identified four questions with distractors that either functioned poorly (not picked by low performers, Q6, Q9) or were ambiguous (i.e., picked more frequently than the correct answer by high performers, Q5, Q6, Q12a). For all but one question (Q4), students chose distractors that suggest they have possible misunderstandings of primary production (Table 3), ranging from misunderstanding the source of biomass for primary production and the process of photosynthesis, to appropriate measures of primary production and the influence of physical forcing on primary production. Students' answers to the open response question 12b also revealed confusion about the process of photosynthesis, with ~14% of students indicating that phytoplankton consumed or used oxygen during photosynthesis when in fact, photosynthesis produces oxygen.

Student performance

We examined gross changes in SMS undergraduates' knowledge about primary production (as mean person-parameter logit values) over all questions for students in each educational level (Fig. 3). The mean scores in each level ranged between -0.9798 logits and -0.3782 logits (the more negative the score, the lower the achievement on the assessment), and score distributions met ANOVA assumptions of normality and homogeneity of variance. Score increased with educational level (ANOVA, M.S. = 1.80, $F_{3,82}$ = 3.03, α = 0.1, p = 0.0329, posthoc power = 0.63, educational level 1 significantly different from levels 2, 3, and 4 [Tukey's LSD,

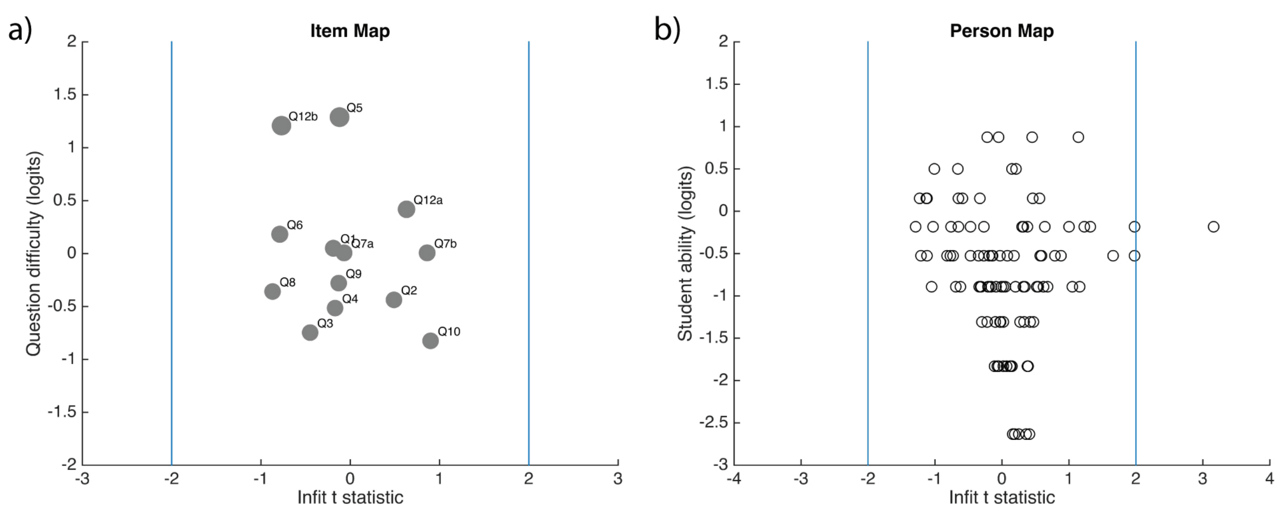

FIGURE 2. Rasch analysis results. (a) Item map showing how each survey question varied in difficulty (y-axis) and how well it fit the Rasch model (x-axis). Question difficulty increases with vertical logit value. The vertical size of each symbol shows the standard error of the estimated difficulty. (b) Person map showing each student's ability on the assessment (y-axis) and how well each student's responses fit the Rasch model (x-axis). Students with more ability had larger vertical ability values. To aid readability, error estimates are not shown by symbol size. The average standard error of estimated ability was 0.65 (±0.11 SD). For both maps, the fit of the questions/students to the model is indicated by their position on the x-axis. Items/students further from 0 fit the model less well. The vertical lines at -2 and 2 infit *t* values indicate the acceptable range for items fitting the Rasch model.

TABLE 2.
Summary of question and item analysis results.

Survey question	Question Metrics		Item Analysis		
	Question Difficulty (logits)	Question Discrimination (D)	Distractors not picked by low performers	Distractors picked more frequently by high performers	Items that may be ambiguous alternatives
1	0.05	0.11	—	B, D	—
2	-0.44	0.15	—	C, D	—
3	-0.75	0.15	—	C, D	—
4	-0.52	0.45	—	D	—
5	1.29	0.04	—	C, E	C
6	0.18	0.07	C	B, C, E	B
7a	0.01	0.20	—	D	—
7b	0.01	-0.10	NA	NA	NA
8[a]	-0.36	0.43	NA	NA	NA
9	-0.28	0.15	D	B, D, E	—
10	-0.82	0.28	—	C	—
11	NA	NA	NA	NA	NA
12a	0.42	0.07	—	B, D	B
12b[a]	1.21	0.23	NA	NA	NA

Left side of table shows question difficulty (logit value) and question discrimination (D) values for each question. Right side of table shows distractors that were flagged by item analysis. In these results, item A is the correct response and all other items are distractors. Questions are provided in Appendix 3. Q11 excluded from analysis.
[a] Open response question.
NA = not applicable.

TABLE 3.
Potential misunderstandings identified by MPPA.

Question	Potential Misunderstanding [Explanation of correct understanding]	Percent of Responses	Comments from Item Analysis
1	Majority of biomass gained by open-ocean photosynthetic marine algae comes from *dissolved forms of nitrogen* [Through photosynthesis, biomass is gained from processing carbon dioxide molecules]	25	Low performers chose this answer almost as frequently as the correct response.
1	Majority of biomass gained by open-ocean photosynthetic marine algae comes from *energy from the sun* [Through photosynthesis, biomass is gained from processing carbon dioxide molecules]	31	High performers chose this answer more than twice as often as other distractors
2	An appropriate measure of marine primary productivity would be in units of *mass of chlorophyll integrated over the water column.* [Marine primary productivity is typically measured in grams of carbon per square meter per year, hence mass of carbon per unit volume per time is the best answer]	38	Low performers chose this answer as frequently as the correct response; high performers chose this almost as frequently as the correct response
3	Decreasing light levels can immediately (within minutes) affect marine primary production by reducing the *amount of nutrients produced by phytoplankton* [Decreasing light levels would reduce the rate of photosynthesis, and consequently the rate at which organic carbon is produced; students seem to confuse nutrients and organic carbon as the product of photosynthesis]	28	Low performers chose this answer twice as frequently as the other two distractors.

Question	Potential Misunderstanding [Explanation of correct understanding]	Percent of Responses	Comments from Item Analysis
5	Graph Interpretation: For a phytoplankter that is continuously mixed to a specified depth shown in a graph and assuming no grazing, the depth at which the rate of photosynthesis and rate of respiration are equal (i.e., compensation depth) *indicates the maximum depth of the mixed layer at which net growth can occur*. [Because phytoplankton are being mixed in the water column, in the graph shown for this question the correct answer is the point that indicates the depth at which integrated photosynthesis and integrated respiration are equal, and below which net growth would not occur]	37	Answer chosen most often by high performers (2 times more frequently than the second most frequent item shown in the next entry) and by low performers
5	For a phytoplankter that is continuously mixed to a specified depth shown in a graph and assuming no grazing, the point at which rate of photosynthesis and rate of respiration are equal *indicates the depth of the photic zone*. [Depth of the photic zone is shown on the graph as the point where rate of photosynthesis is zero, but this is not what is labeled at point C].	26	Fewer high performers chose this answer compared with the previous entry.
6	*Chlorophyll-a and chlorophyll-b* are pigments that occur in the photosynthetic reaction center of all eukaryotic autotrophic algae. [Only chlorophyll-a is found in the reaction center of all eukaryotic autotrophic algae]	37	Both high and low performers chose this answer more frequently than the correct response.
7a	Graph Interpretation: Students chose the plot that shows spring stratification of surface waters in the Gulf of Maine *decreasing* due to increased fresh water inputs instead of increasing. [Students chose the opposite answer to the correct one, perhaps due to problems with graph mechanics; however, they might also misunderstand how salinity and density are related and that increased fresh water inputs would lead to less dense surface water]	27	This answer was chosen most often by low performers
9	El Niño conditions are best described by this sequence of events: *wind intensification>upwelling increases>more nutrients near the coast>phytoplankton bloom>oxygen used up in photic zone>anchovy suffocate>fishermen catch less* [This incorrect answer describes a sequence of events that might be associated with eutrophication, not El Niño, in which warm surface water caps upwelling, reducing nutrients at the surface so that phytoplankton die and the anchovy that feed on them starve and fishermen catch less.]	27	Low performers chose this answer 3 times more frequently than the other distractors.
10	Over a phytoplankter's lifetime, *10%* of the energy converted from solar energy to the energy stored in chemical bonds is lost as *heat*. [90% of energy is "lost" as metabolic heat; students are perhaps thinking about the oft-quoted 10% efficiency of transfer and confusing the percentage retained vs. transferred/"lost"]	27	Low performers chose this answer almost as frequently as the correct answer.
12a	Graph Interpretation: Students identified the processes illustrated by graphs of dissolved inorganic nutrient and chlorophyll-a in surface waters and dissolved oxygen at depth as being associated with the *spring bloom in the North Atlantic Ocean*, rather than eutrophication in Chesapeake Bay. [Increased phytoplankton abundance is associated with both the spring bloom and eutrophication, but only in eutrophication would you find seasonal bottom levels of dissolved oxygen near zero, as shown in the graphs]	52	Both low and high performers chose this answer most frequently, more than the correct response.

Summary of students' potential misunderstandings about marine primary production based on their responses to the survey questions; the misunderstanding is indicated in italics and an explanation of the correct understanding is provided in brackets. Only distractors chosen by 25% or more of those responding are shown. Survey questions and response distributions are provided in Appendix 3. MPPA = marine primary production assessment.

$\alpha = 0.1$, p values: level 1 vs. level 2 = 0.088, level 1 vs. level 3 = 0.027, level 1 vs. level 4 = 0.011, level 2 vs. level 3 = 0.934, level 2 vs. level 4 = 0.428, level 3 vs. level 4 = 0.409]). Over all pooled educational levels, student achievement decreased as the cognitive demand of the question increased ($R^2 = 0.366$, $p = 0.0284$, Fig. 4).

To test whether confident students were more likely to choose the correct responses, their self-reported

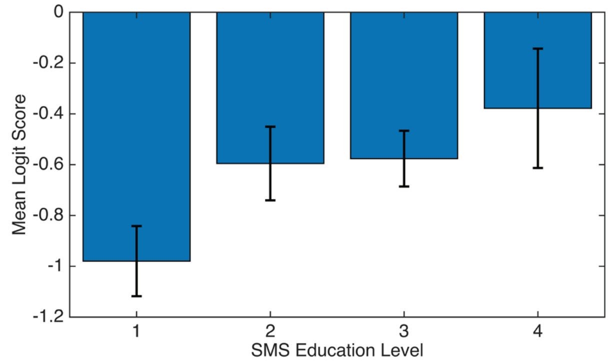

FIGURE 3. Mean assessment scores. Mean person parameter logit values in each SMS educational level. Larger values (less negative) indicate higher performance on the assessment. Error bars show standard error. Number of students in each educational level was level 1 = 20, level 2 = 17, level 3 = 34 and level 4 = 15. The mean score at level 1 was significantly different than all other levels (Tukey's LSD, α = 0.1). LSD = least significant difference; SMS = School of Marine Sciences.

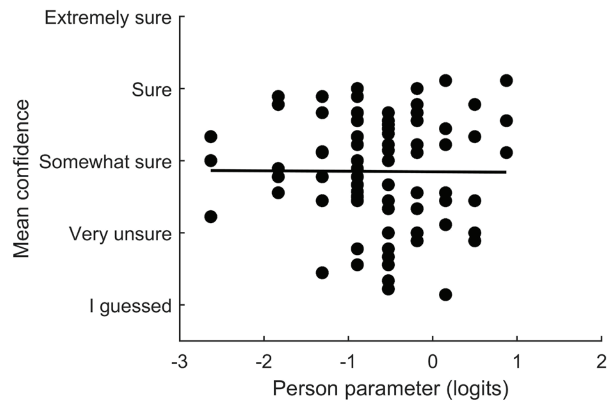

FIGURE 5. Relationship between self-reported response confidence and assessment score. The graph indicates a linear relationship between categorical mean self-reported question response confidence and overall assessment score over all questions and SMS educational levels. Each symbol represents a unique student (n = 86). The least-squares regression line is plotted (Pearson correlation, $R^2 = 0.00$, $\alpha = 0.1$, $p = 0.956$). SMS = School of Marine Sciences.

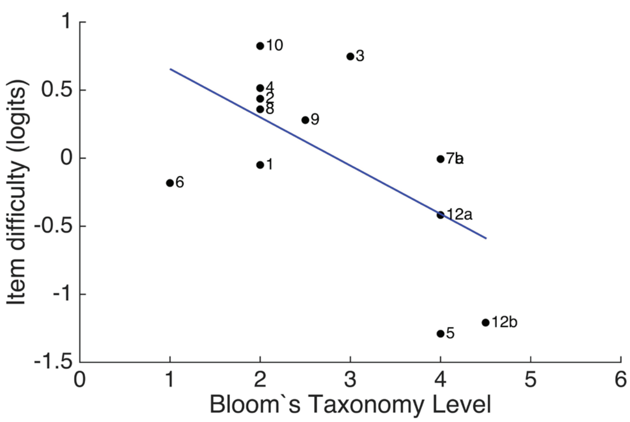

FIGURE 4. Relationship between assessment difficulty and cognitive demand of assessment questions. The linear correlation of mean logit assessment difficulty for each question is pooled over all SMS educational levels and the cognitive demands of each question as categorized by Bloom's Taxonomy levels ($R^2 = 0.37$, $p = 0.028$). Note that the signs of item difficulty values have been inverted to emphasize increased difficulty of questions with higher Bloom's levels. SMS = School of Marine Sciences.

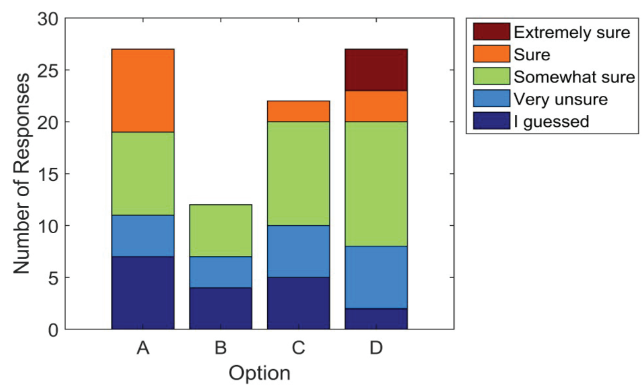

FIGURE 6. Self-reported confidence for survey Question 1. Responses are pooled over all educational levels, with confidence in each item indicated by color. The question was: The majority of actual weight (dry biomass) gained by open-ocean photosynthetic marine algae as they grow comes from which of the following substances? The possible responses were:

A. Molecules containing carbon. (correct answer)
B. Particulate matter.
C. Dissolved forms of nitrogen.
D. Energy from the sun.

confidence was correlated with survey score (person-parameter logit values). Over all pooled educational levels, categorical mean confidence was not significantly correlated with overall score (Pearson correlation, $R^2 = 0.00$, $\alpha = 0.1$, $p = 0.9565$, Fig. 5). We also examined confidence by individual question items. Results are presented for Q1 in Figure 6 and for all questions in Appendix 3. There was a weak negative correlation (Pearson correlation, $R^2 = 0.24$, $\alpha = 0.1$, $p = 0.089$) between student confidence and the cognitive demand of the questions pooled over all educational levels (Fig. 7), indicating that students tended to be less confident in their answers as the questions' cognitive demand increased.

DISCUSSION

The Marine Primary Production Assessment probed student understanding of a key concept in marine science that requires students to integrate knowledge from biological, chemical, and physical scientific perspectives, and that is taught from these different perspectives across the SMS undergraduate curriculum. We aimed to develop an assessment that we could use to understand the differences in student learning among educational cohorts based on

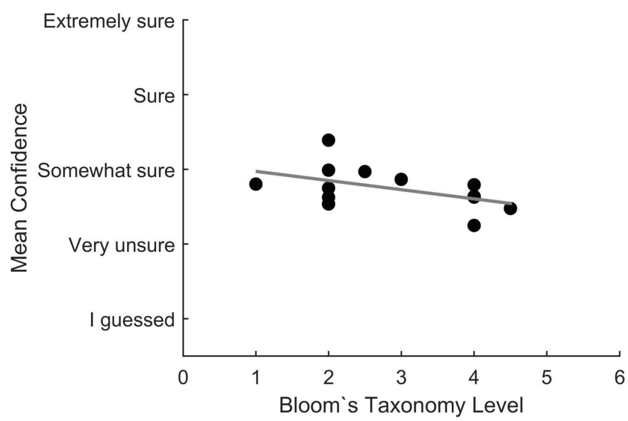

FIGURE 7. Relationship between question cognitive demand and mean student confidence. The graph shows a linear relationship between Bloom's Taxonomy level for each question and categorical mean self-reported question response confidence pooled over all educational levels. The least-squares regression line is plotted (Pearson correlation, $R^2 = 0.24$, $\alpha = 0.1$, $p = 0.089$).

their progression through the marine sciences curriculum. We also sought to identify student misunderstandings (and confidence therein) of key content knowledge to use this information to inform curriculum assessment and revision. The MPPA was not constructed to be and should not be used as a concept inventory for marine science.

Students gained content knowledge about marine primary production as they progressed though the SMS undergraduate program. Scores on the assessment increased by 61% from education level 1 to level 4 (Fig. 3), although the only statistically significant change was between level 1 and all the other levels. The overall increase in knowledge was consistent with increases noted in other life sciences programs (29, 30). Although students gained knowledge, the item analysis results suggest students also have potential misunderstandings. We define a misunderstanding as a seemingly incorrect idea that is inconsistent with expert knowledge of the topic. We make no claims about the root or cognitive construct of the misunderstanding here. Student responses showed evidence of potential misunderstandings in all but one (Q4) of the multiple-choice questions. Students' answers on Q1 and Q3 indicated confusion regarding the source of biomass for primary production and the influence of light and nutrients on photosynthesis. Students were also confused about appropriate measures of primary production (Q2), the types of photosynthetic pigments found in primary producers (Q6), and the influence of wind/physical forcing on primary production (Q5, Q9). Several distractors based on potential misunderstandings identified from the first-year homework assignment (see Appendix 2) were confirmed as misunderstandings by item analysis. For example, the preliminary data indicated that some students confuse nutrients (nitrogen and phosphorous specifically) as the primary source of biomass with carbon sources, which actually fill this role, and this was a reason we included item C as a distractor in Q1. The item analysis also identified this distractor as a possible misunderstanding, suggesting that it may be broadly held by the SMS undergraduate population. We found no factual errors in any of the distractors on the survey; thus, students' choosing distractors may indicate gaps in learning or the presence of misunderstandings.

Q1 explored what SMS students thought was the primary source of biomass for marine algae. This question was adapted from a study investigating undergraduate biology students' understanding of the roles of sunlight and carbon sources in terrestrial plant primary production (31). In that study, the following question was presented to biology students:

The majority of actual weight (dry biomass) gained by plants as they progress from seed to adult plant comes from which of the following substances?

a. Particle substances in soil that are taken up by plant roots.
b. Molecules in the air that enter through holes in the plant leaves. [Correct choice]
c. Substances dissolved in water taken up directly by plant roots.
d. Energy from the sun.

The strongest misconceptions among biology students answering this question were that biomass was primarily due to substances in the soil taken up by plant roots (choice a) or substances dissolved in the water (choice c). Very few students selected "Energy from the sun" (choice d) (31). In contrast, on the MPPA this last distractor in Q1 was chosen about twice as often as any other distractor by the highest performing subset of the SMS undergraduates who took the assessment (Table 2). Moreover, students were more confident in their selection of distractor D "Energy from the sun" than distractor C "Dissolved forms of nitrogen" (Fig. 6). Together, these data suggest that students in our study held a stronger misunderstanding that "Energy from the sun" is a source of biomass for phytoplankton compared with "Dissolved forms of nitrogen" as a source of biomass. We can only speculate about the reasons why marine science undergraduates would be more likely than biology students to think energy from the sun was a source of biomass. Perhaps the marine emphasis, especially on light attenuation with depth, and the heterogeneous, moving, three-dimensional ocean environment leads to some confusion among marine science students as to the true role of light in primary production.

Role of student confidence

We investigated relationships between student confidence and achievement to better understand whether mismatches in these two parameters might be negatively influencing learning by students who participated in the study. Our methodology differed from previous self-efficacy research (e.g., 32, 33). Rather than ask students to rank their beliefs in their general science abilities prior to measuring academic performance, we asked them to state how confident they were of their performance immediately after attempting an academic task. Our measure was designed to identify whether students' self-appraisal of their abilities matched their actual abilities.

Across pooled educational levels, spread in confidence tends to be similar regardless of proficiency on the assessment (Fig. 5), suggesting that high-performing students are as aware of their abilities as low-performing students. We don't see strong evidence of the Dunning-Kruger effect (34, 35), where low-performing students tend to be overconfident in their abilities compared with high-performing students. The observed evidence of both overconfidence and lack of confidence across performance levels suggests that the participating SMS students may have a general lack of science self-efficacy on content related to marine primary production. Overconfidence is especially worrisome as it may hamper students' efforts to learn new material (36).

Feedback to curriculum and ocean science education

The MPPA questions required reasoning skills that spanned the lower and higher levels of Bloom's Taxonomy to address issues raised by Momsen *et al.* (37) that undergraduate life-science assessments tend to target only lower-level cognitive skills, especially at the introductory level. In addition, including higher-order questions addressed a SMS programmatic goal that students should be able to synthesize prior learning to formulate and address scientific questions by the time they graduate. However, students tended to do better on questions that required lower cognitive demands (remembering and understanding) than on those requiring higher demands (applying, analyzing, and evaluating, Fig. 4). These results are consistent with previous work demonstrating only modest increases in reasoning ability between first year and senior levels in undergraduate programs (e.g., 38). Comparisons with independent measures of critical thinking ability (e.g., Cornell Critical Thinking Tests, capstone projects and evaluated portfolios [39]) would help elucidate whether this is a true representation of the students' abilities or simply a function of the questions we asked. The questions on our survey with the highest Bloom's levels were also questions that involved graph interpretation; thus, students' poor performance also might be due to graphing literacy deficiencies. Indeed, we know from McOsker's (40) work with SMS first-year students that a majority of students (73%) reported difficulty understanding the form of graphs from the primary literature presenting information about the effect of nutrients on primary productivity (e.g., the meaning of axis labels, units of measure, error bars, letters indicating the results of multiple comparisons, etc.). Our SMS students are not alone; Glazer (41) noted that "reading graphs and making sense of them is hard" for students, and interpreting graphs is influenced by prior content knowledge as well as proficiency with graphing skills. The MPPA questions were designed to be consistent with our curricular expectation that SMS students will become proficient at generating, interpreting, and applying marine science data. The results from the MPPA suggest that making connections between graphing literacy and content knowledge should be a curricular priority.

We designed the MPPA to quantify what students at each level of the undergraduate marine sciences curriculum knew about the topic, when/where they gained the knowledge, and how they retained it. Students make large initial gains in knowledge during their first-year introductory courses, and they appear to build on these in subsequent years. We identified specific student misunderstandings and are using that insight to inform curriculum development and revision. Specifically, we are now addressing misunderstandings about the source of carbon for photosynthesis in the introductory marine biology course and introducing in-class activities that reinforce creating and interpreting graphs related to several key concepts. Perhaps the most important lesson we learned is that program-level assessment is more efficient when the department has detailed content and skill-specific student learning outcomes. We also learned that even a short, targeted assessment can provide meaningful information. Having "started small" we can build a more comprehensive program-level assessment plan.

We acknowledge that testing the MPPA on larger numbers of students would be necessary to establish certainty in relationships we have seen. Results based on larger deployments will help determine whether the patterns we saw in things like distractor selection and question discrimination hold true or whether they reflected this unique set of students. More deployments would also provide data needed to establish test reliability over time, something we were unable to do with our single deployment. Questions could be added to the MPPA to fill gaps in the stepwise progression of knowledge on this topic identified by the Rasch analysis. Based on our single deployment, we recommend removing or editing several questions in future uses of the survey (described in Appendix 4). We consider our findings to be a starting point for further development of assessments of marine science content knowledge and skills.

We hope that our assessment will be used and improved upon by others and that the description of our approach (mapping known misunderstandings to the curriculum to develop content questions) may serve as a starting point for development of a comprehensive marine sciences undergraduate concept inventory. The undergraduate marine sciences education community would benefit from a good, well tested, set of questions for assessing content knowledge, skills, and abilities, so that each program would not need to reinvent the wheel for its own assessment needs.

SUPPLEMENTAL MATERIALS

Appendix 1: Methods of survey development
Appendix 2: Implemented curriculum concept map
Appendix 3: Marine primary production assessment (MPPA) reported on in this paper and graphs of student self-reported confidence by question item
Appendix 4: Recommendations for future deployments of the MPPA

ACKNOWLEDGMENTS

This project was conducted in fulfillment of requirements for the University of Maine Master of Science in Teaching program by Ryan Weatherbee. Additional thanks to William Ellis for help coordinating interactions with the undergraduate student body in SMS. We appreciate helpful comments on earlier versions of the manuscript from Natasha Spear and Molly Schauffler and several anonymous reviewers. Both RW and SL contributed to the study design, analysis and interpretation of the data, and writing of this manuscript. Assistance in the form of a graduate stipend for RW was provided by the University of Maine (UM) Research in STEM Education (RiSE) Center and the UM School of Marine Sciences (SMS). No potential conflicts to study participants exist. This study had no sponsors. It was conducted as part of a Master of Science in Teaching graduate degree at the University of Maine.

REFERENCES

1. American Association for the Advancement of Science. 2011. Vision and change in undergraduate biology education: a call to action: a summary of recommendations made at a national conference organized by the American Association for the Advancement of Science, July 15–17, 2009. Washington, DC.
2. National Research Council. 1999. Transforming undergraduate education in science, mathematics, engineering and technology. The National Academies Press, Washington, DC.
3. National Research Council. 2003. Evaluating and improving undergraduate teaching in science, technology, engineering and mathematics. The National Academies Press, Washington, DC.
4. Tran LU, Payne DL, Whitley L. 2010. Research on learning and teaching ocean and aquatic sciences. NMEA Spec Rep 3:22–26.
5. Arthurs L, Hsia JF, Schweinle W. 2015. The oceanography concept inventory: a semicustomizable assessment for measuring student understanding of oceanography. J Geosci Educ 63:310–322.
6. Meltzer DE, Thornton RK. 2012. Resource letter ALIP-1: active-learning instruction in physics. Am J Phys 80:478–496.
7. Hoskinson AM, Caballero MD, Knight JK. 2013. How can we improve problem solving in undergraduate biology? Applying lessons from 30 years of physics education research. CBE Life Sci Educ 12:153–161.
8. Dirks C. 2011. The current status and future direction of biology education research. Commissioned paper for the Committee on the Status, Contributions, and Future Directions of Discipline-Based Education Research, Board of Science Education, Division of Behavioral and Social Sciences and Education, National Research Council, Washington DC.
9. Doktor J, Mestre J. 2011. A synthesis of discipline-based education research in physics. Commissioned paper for the Committee on the Status, Contributions, and Future Directions of Discipline-Based Education Research, Board on Science Education, Division of Behavioral and Social Sciences and Education, National Research Council, Washington DC.
10. National Research Council. 2012. Discipline-based education research: understanding and improving learning in undergraduate science and engineering. The National Academies Press, Washington, DC.
11. D'Avanzo C. 2008. Biology concept inventories: overview, status, and next steps. Bioscience 58:1–7.
12. Smith MK, Knight JK. 2012. Using the genetics concept assessment to document persistent conceptual difficulties in undergraduate genetics courses. Genetics 191:21–32.
13. Anderson DL, Fisher KM, Norman JG. 2002. Development and validation of the conceptual inventory of natural selection. J Res Sci Teach 39:952–978.
14. Fisher KM, Williams KS. 2011. Concept inventories and conceptual assessments in biology (CABs): an annotated list. Available at: www.sci.sdsu.edu/CRMSE/files/Concept_Inventories_in_Biology_20110325.pdf
15. Mulford DR, Robinson WR. 2002. An inventory for alternate conceptions among first-semester general chemistry students. J Chem Educ 79:739–744.
16. Dick-Perez M, Luxford CJ, Windus TL, Holme T. 2016. A quantum chemistry concept inventory for physical chemistry classes. J Chem Educ 93:605–612.
17. Wren D, Barbera J. 2013. Gathering evidence for validity during the design, development and qualitative evaluation of thermochemistry concept inventory items. J Chem Educ 90:1590–1601.
18. Hestenes D, Wells M, Swackhamer G. 1992. Force concept inventory. Phys Teach 20:141–158.
19. Ding L. 2007. Designing an energy assessment to evaluate student understanding of energy topics. North Carolina State University, Raleigh, NC.
20. Libarkin JC, Ward EMG. 2011. The qualitative underpinnings of quantitative concept inventory questions, p 37–48. *In* Qualitative inquiry in geoscience education research: Geological Society of America Special Paper 474.
21. Libarkin JC, Anderson SW. 2005. Assessment of learning in entry-level geoscience courses: results from the geoscience concept inventory. J Geosci Educ 53:394–401.
22. Knight JK, Wood WB. 2005. Teaching more by lecturing less. Cell Biol Educ 4:298–310.
23. Mair P, Hatzinger R, Maier MJ, Rusch T. 2015. Package "eRm." CRAN. http://cran.r-project.org/web/packages/eRm/eRm.pdf.

24. Bond TG, Fox CM. 2015. Applying the Rasch model: fundamental measurement in the human sciences, 3rd ed. L. Erlbaum, Mahwah, NJ.
25. Nitko AJ, Brookhart SM. 2007. Educational assessment of students, 5th ed. Pearson Education, Inc., Upper Saddle River, NJ.
26. Bissell AN, Lemons PP. 2006. A new method for assessing critical thinking in the classroom. BioScience 56:66–72.
27. Byrnes JP. 2001. Cognitive development and learning, 2nd Edition. Allyn and Bacon, Boston, MA.
28. Crowe A, Dirks C, Wenderoth MP. 2008. Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. CBE Life Sci Educ 7:368–381.
29. Howard DR, Miskowski JA, Grunwald SK, Abler ML. 2007. Assessment of a bioinformatics across life science curricula initiative. Biochem Mol Biol Educ 35:16–23.
30. Vogt FD, Slish D. 2011. A model for programmatic evaluation by student assessment. J Sci Educ Technol 20:796–802.
31. Ebert-May D, Lim H, Batzli J. 2003. Disciplinary research strategies for assessment of learning. BioScience 53:1221.
32. Andrew S. 1998. Self-efficacy as a predictor of academic performance in science. J Adv Nurs 27:596–603.
33. Multon KD, Brown SD, Lent RW. 1991. Relation of self-efficacy to academic outcomes: a meta-analytic investigation. J Couns Psychol 38:30–38.
34. Bell P, Volckmann D. 2011. Knowledge surveys in general chemistry: confidence, overconfidence, and performance. J Chem Educ 88:1469–1476.
35. Kruger J, Dunning D. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 77:1121–1134.
36. Dunlosky J, Rawson KA. 2012. Overconfidence produces underachievement: inaccurate self evaluations undermine students' learning and retention. Learn Instr 22:271–280.
37. Momsen JL, Long TM, Wyse SA, Ebert-May D. 2010. Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. CBE Life Sci Educ 9:435–440.
38. Brabrand C, Dahl B. 2009. Using the SOLO taxonomy to analyze competence progression of university science curricula. High Educ Int J High Educ Educ Plan 58:531–549.
39. King PM, Wood PK, Mines RA. 1990. Critical thinking among college and graduate students. Rev High Educ 13:167–186.
40. McOsker 2009. Student understanding of error and variability in primary science communication. Masters Thesis. University of Maine, Orono, ME.
41. Glazer N. 2011. Challenges with graph interpretation: a review of the literature. Stud Sci Educ 47:183–210.