Assessing the Effectiveness of Sustainability Learning

By Jill A. Marshall, Jay L. Banner, and Hye Sun You

This study investigated the interaction of disciplinary and interdisciplinary learning in a team-taught, first-year, interdisciplinary sustainability course. We surveyed (pre/post) both STEM (science, technology, engineering, and mathematics) and non-STEM majors (N = 241), assessing attitudes and content knowledge. Responses were analyzed using factor analysis, classical test theory, and Rasch analysis. Multivariate analyses of variance were performed to look for pre/post differences and differences between groups. Confirmatory factor analysis verified that the content knowledge items adequately assessed disciplinary and interdisciplinary understanding separately. Tests indicated an adequate item difficulty range, but also a need for more items at the higher end. All groups improved significantly over the semester. Slightly higher gains for STEM students and a correlation between disciplinary, and interdisciplinary learning might indicate a possible benefit of "disciplinary grounding." The instrument vetted through this study expands the number of validated sustainability content knowledge items. Further, it allows users to probe content knowledge along both interdisciplinary and disciplinary dimensions.

Given the charge to prepare citizens to embrace sustainable practices and policies, educating for sustainability should be the purview of college science teaching (Cortese, 2003). Sustainability education calls for "collaboration across disciplinary and institutional boundaries" (Remington-Doucette, Hiller Connell, Armstrong, & Musgrove, 2013, p. 405). Higher education, however, has been slow to move toward interdisciplinary curriculum and integrative pedagogical techniques. Institutions of higher education have such ingrained disciplinary structure that it is difficult to implement truly interdisciplinary courses. Team teaching and cross-listing are options, but questions remain as to how content will be developed to meet the needs of multiple disciplines and prevent the course from being segmented (Warburton, 2003).

There is also the issue of where such a course should be positioned. It might serve as a capstone experience, assuming students need a disciplinary foundation to address interdisciplinary problems (Boix Mansilla & Duraising, 2007; Remington-Doucette et al., 2013). Alternatively, it could serve as a cornerstone, assuming that early authentic, contextualized STEM learning (typically interdisciplinary) will enhance later studies. Sustainability objectives might also be infused into a variety of existing courses (Warburton, 2003), with the risk that interdisciplinary perspectives will be overshadowed by the focus, methods, and knowledge domain of the particular disciplinary departments where the courses are housed (Fisher & McAdams, 2015).

Assessing sustainability education

Assessing the outcomes of teaching for sustainability is complicated. Is it enough for students to develop knowledge and skills, or are enhanced attitudes and sustainable practices the actual goal? Even accepting an intermediate goal of strengthening understanding of sustainability issues, the interdisciplinary nature makes assessing learning outcomes difficult. Even experienced faculty recognized for teaching interdisciplinary courses report unease at assessing student outcomes (Boix Mansilla & Duraising, 2007).

Some progress has been made toward assessing sustainability education. Remington-Doucette et al. (2013) developed a rubric to assess students' holistic thinking and conflict resolution competencies. They assessed skill in analyzing problems from a systems perspective, but not sustainability content knowledge specifically. Remington-Doucette and Musgrove (2015) found that women improved their "sustainability competence" more than men as a result of a sustainability course.

Assessing knowledge

Environmental content knowledge surveys generally indicate poor understanding on the part of adult Americans (Leiserowitz, Smith, & Marlon, 2010). Zwickle, Koontz, Slagle, and Bruskotter (2014) developed an assessment to measure sustainability knowledge more directly, using item response theory to validate 16 multiple-choice items in the environmental, social, and economic domains. These authors assumed the independence of those domains, rather than performing a factor analysis. Their instrument is still being refined.

The critical role of fundamental disciplinary knowledge for sustainability (Clark & Dickson, 2003) and the need to investigate how disciplinary learning interacts with the development of interdisciplinary understanding (Warburton, 2003) require assessments with both disciplinary and interdisciplinary items. Shen, Liu, and Sung (2014) reported on the development of a tool for assessing understanding of a single topic (osmosis) in both dimensions. Experts from multiple disciplines identified key concepts, but they had considerable difficulty in merging their perspectives to define and delineate the osmosis construct (Sung et al., 2015).

Study design

Our study follows a similar design to Shen et al. (2014), but it explores the interaction of disciplinary and interdisciplinary learning by both STEM (science, technology, engineering, and mathematics) and non-STEM majors in the broader context of sustainability. We define *interdisciplinary learning* as addressing contexts that require application of overarching concepts or knowledge from multiple disciplines for problem solving. At the University of Texas (UT), Austin, sustainability is the focus of a team-taught interdisciplinary course designed as a first-year experience. To assess changes in knowledge and attitudes, a pre/post survey is administered. Knowledge items are directly related to environmental issues (Kopnina, 2012). The instrument includes single-discipline (D) and interdisciplinary (I) knowledge items. This study was designed to assess the validity of this instrument and investigate the interaction of disciplinary and interdisciplinary learning. Research questions were:

- RQ1: What are the psychometric properties of the instrument?
- RQ2: Are there differences in student responses to the survey before and after participation in the sustainability class?
- RQ3: Are there gender differences in responses to the survey?
- RQ4: Are there differences based on major (STEM vs. non-STEM)?
- RQ5: Is any relationship indicated between disciplinary and interdisciplinary learning?

Method

Setting and participants

Sustaining a Planet is a large-format signature course. Taking one signature course is a degree requirement for every first-year and transfer student at UT Austin. During this study, which occurred over the fall 2015 semester, 242 students were enrolled: 79% freshmen, 12% sophomores, and 9% upperclassmen (transfer students fulfilling the signature requirement). The course defines sustainability in the context of the environment-economy-equity triangle and focuses on the environment vertex recognizing that all three cannot be comprehensively covered in a single course. Although the course incorporates elements of social science, at its core it is a science course. Course learning goals include enhanced understanding of (a) sustainability concepts and challenges from the perspectives of the natural sciences (chemistry, physics, biology, geosciences) and engineering (energy, materials, industrial processes), and (b) the interdependence of different components of the Earth system (atmosphere, hydrosphere, lithosphere, biosphere). Goals also include enhanced student attitudes toward sustainability. Key concepts threaded throughout the course include the tragedy of the commons, the scientific method, feedback mechanisms, and the need for interdisciplinary approaches to address sustainability challenges.

Survey instrument

Student attitudes and interdisciplinary and disciplinary knowledge have been measured since the course was first offered. Attitudes are assessed with items from a survey developed at UCLA (Astin, Oseguera, Sax, & Korn, 2002). Knowledge items were developed by the course instructors. For this study, three additional items (K7, K12, K13 [see Appendix A]) were adapted from other tests (You, 2016) to increase coverage of interdisciplinary objectives.

To investigate attitudes, the final instrument (Appendix A) contained 9 Likert-scale items (scale: 1 = *not at all important*, 2 = *somewhat important*, 3 = *very important/essential*). Some of these items have multiple options, for a total of 29 scaled choices. The composite attitude measure was the total score on all Likert-scale items, ranging from 29 to 87.

The instrument also contained 16 multiple-choice knowledge items. Most had been vetted in previous years to determine item difficulty (Carter, 2013; Zwickle et al., 2014). The final instrument was administered online using Qualtrics (Provo, Utah) as beginning- and end-of-semester course assignments. Students received credit for completing the survey regardless of their answers.

A two-factor, full-information confirmatory factor analysis (CFA) was performed on the content items (Glynn, Brickman, Armstrong, & Taasoobshirazi, 2011) using Mplus v.7.4 (Muthén & Muthén, 1998–2015). A confirmatory factor analysis seeks to support the existence of hypothesized latent variables, in our case disciplinary (D) and interdisciplinary (I) understanding, that can explain the structure underlying responses to an assessment. Responses to items identified as D and I, respectively, should project (load) on two different dimensions in parameter space. Disciplinary items were identified as those that could be answered with knowledge or perspectives from a single discipline; interdisciplinary items required the integration of knowledge and skills from multiple disciplines or overarching knowledge beyond the limits of a single discipline. A team of experts in sustainability science independently identified each item as D or I, with uniform agreement on all but one item. The one item on which the experts were split was classified through negotiation as testing interdisciplinary knowledge.

Kline (2015) recommended a multistep evaluation approach, calculating first the model χ^2 ("badness of fit" or the extent that a proposed model varies from the data) and then additional fit indices. Possibilities include the weighted root mean square (WRMR), supplemented by the root mean square error of approximation (RMSEA) or comparative fit index (CFI; Yu, 2002). For our analysis, we report χ^2, WRMR, CFI, and RMSEA. Item difficulty for the knowledge items was estimated by traditional item difficulty calculation (% correct) and Rasch analysis (Shen et al., 2014).

TABLE 1

Demographic overview of sample by major and gender.

	STEM majors	Non-STEM majors	Total
Presurvey			
Females	58	71	130[a]
Males	61	50	111
Postsurvey			
Females	47	50	97
Males	32	32	64

Note: STEM = science, technology, engineering, and mathematics.
[a]One female student did not indicate a major.

TABLE 2

Model fit statistics for the proposed model. For χ^2, *df* is the number of pieces of information (15 items*16/2) minus the parameters being estimated (15 factor loadings, 15 error variances, and 1 covariance between the two factors in the model).

χ^2	*df*	*p*-value	Normed $\chi^2(\chi^2/df)$	CFI	RMSEA (90% CI)	WRMR
102.570	89	=0.1541	1.15	.912	.031 (.000; .055)	.866

Analysis of course outcomes

A multivariate analysis of variance (two-way MANOVA) was performed to test for gender differences and differences between students majoring in STEM and non-STEM fields on attitudes, interdisciplinary knowledge, and disciplinary knowledge. When a significant difference was found with the MANOVA, posthoc *t*-tests were conducted to identify the source of the difference.

As the pre- and postsurveys were conducted in an unmonitored online environment, students were assured of anonymity to counter possible collusion and use of web resources to identify answers, precluding a paired-samples analysis. Therefore, an independent-samples *t*-test was conducted to evaluate the impact of the sustainability course on students' attitudes and interdisciplinary and disciplinary knowledge by comparing pre- and postsemester scores for all participants.

Results

RQ1: Psychometric properties of the instrument

Factor analysis

A total of 241 respondents provided demographic information for the presurvey and 161 for the postsurvey. Table 1 gives the breakdown by gender and major.

CFA of the postsurvey supported the two-factor (interdisciplinary and disciplinary) structure of the knowledge items, with the exception of one item (K3) that did not load on either factor. Deletion of this item yielded an adequate measurement model for the remaining data. Table 2 reports model fit statistics.

The χ^2 of 102.570 (df = 89, p = .1541) is considered an appropriate fit (Hooper, Coughlan, & Mullen 2008). To reduce the effect of sample size, some researchers use the normed chi-square divided by the degrees of freedom (χ^2/df), with recommended values in the 1.0–3.0 range, placing our value well within the acceptable range (Glynn et al., 2011). The WRMR was 0.866; Yu (2002) indicated that a cutoff value for WRMR close to 1.0 is acceptable. The RMSEA was 0.031, in the "close fit" range well below the .06 cutoff value for a "relatively good" fit (Hu & Bentler, 1999). Although 0.90 is typically used as a lower cutoff for CFI (Glynn et al., 2011), Hu and Bentler (1999) recommended a value for CFI > 0.95, indicating possible problems with our CFI value of 0.912. However, raising the cutoff rate to 0.95 can increase the chances of rejecting a model that actually fits the data. This, in conjunction with the values for WRMR and RMSEA, supports the conclusion that the two-factor model is a good fit for our data. Table 3 gives item factor loadings. The bolded p values, for K4 and K8, are marginally nonsignificant, at .060 and .083, respectively.

TABLE 3

Standardized model results with item loadings on the disciplinary (top) and interdisciplinary (bottom) factors. K = knowledge items.

Disciplinary item	Estimate	*SE*	Est./*SE*	*P*-value
K1	0.596	0.112	5.325	0.000
K2	0.293	0.126	2.322	0.020
K5	0.506	0.212	2.391	0.017
K6	0.681	0.116	5.881	0.000
K8	0.216	0.124	1.736	**0.083**
K11	0.624	0.109	5.707	0.000
K14	0.382	0.117	3.252	0.001
K16	0.545	0.112	4.847	0.000

Interdisciplinary item	Estimate	*SE*	Est./*SE*	*P*-value
K4	0.153	0.124	1.884	**0.060**
K7	0.411	0.117	3.382	0.001
K9	0.623	0.106	3.632	0.000
K10	0.336	0.117	2.872	0.004
K12	0.539	0.103	5.236	0.000
K13	0.711	0.110	6.473	0.000
K15	0.696	0.097	7.161	0.000

Rasch analysis

Item fits from Rasch analysis for the 15 retained items ranged from 0.87–1.45 on the pretest and 0.81–1.31 on the posttest, indicating acceptable fit (Wright, Linacre, Gustafson, & Martin-Löf, 1994).

Item difficulty

The percentage of correct answers and the item difficulty parameter (b) from the Rasch analysis are given for each item in Table 4. Results for K3, the item deleted from the model during CFA, are not reported. Students gained an average of 19.6% on interdisciplinary (I) items and 32.4% on disciplinary (D).

Figure 1 shows the Wright (Item/Person) Map, providing a visual representation of the spread in item difficulty compared with student ability for both the pre- and postsurvey (Stone, Wright, & Stenner, 1999). The vertical scale represents person ability on the left (–3 to 3 logits) and item difficulty (Rasch b factor) on the right. Each "#" represents three students scoring at that ability on the pretest and two on the post. Each "." represents 1–2 students on the pretest and 1 on the posttest. Items on the pretest ranged from –1.77 to 2.29 in difficulty and from –2.71 to 1.32 on the posttest. The pretest matches the range of student abilities fairly well, but the posttest spread indicates that the items are slightly easy for the students, with no items testing at the highest ability levels.

RQ2: Pre/post differences in survey responses

There are significant, positive differences between the pre- and postsurveys, with moderate to large effect size (ES; Cohen's d) between the presurvey and postsurvey on both attitude, $t = 5.373$ ($df = 382$ [$N = 384$ completing all attitude items on either survey], $p < .001$, ES = 0.550), and knowledge, $t = 15.213$ ($df = 402$ [$N = 404$ completing all knowledge items on both pre and post], $p < .001$, ES = 1.518), and both disciplinary, $t = 16.890$ ($df = 402$, $p < .001$, ES = 1.685), and interdisciplinary scores, $t = 8.202$ ($df = 402$, $p < .001$, ES = 0.818), separately. Not surprisingly, student scores on all content knowledge items improved from the beginning to the end of the semester. This result must, of course, be interpreted with caution, as the pre/post groups were not matched.

RQ3: Gender differences in survey responses or pre/post gains

In the presurvey, there was a statistically significant gender difference in students' attitudes, favoring females (females rated the importance of the issues higher) when two outcome variables are considered simultaneously, $F(2, 222) = 8.562$, $p < .0001$; Wilk's $\Lambda = 0.928$. There is no significant interaction effect, meaning that the effect of major on the dependent variables is the same for males and females. For the postsurvey, there was a marginally statistically significant gender difference in students' attitudes when two outcome variables are considered simultaneously, $F(2, 151) = 3.108$, $p = .048$; Wilk's $\Lambda = 0.960$, again with no significant interaction effect. When examining the non-STEM students alone, however, the gender difference disappears on the postsurvey, $t = -0.495$ ($df = 76$, $p = .622$). There were no significant gender differences in content knowledge overall, $t = 0.883$ ($df = 239$, $p = .378$) on the pretest and $t = 1.048$ ($df = 159$, $p = .296$) on the posttest.

RQ4: Differences based on major

There were no significant differences by major on content knowledge overall, $t = 1.728$ ($df = 238$, $p = .085$) on the pretest and $t = 1.696$ ($df = 159$, $p = .092$) on the posttest. On the posttest only, there was a marginally significant, $t = 2.046$ ($df = 159$, $p = .042$), difference between STEM and non-STEM students (favoring STEM students) on the disciplinary score only. The effect size was small to medium (ES = 0.3245).

RQ5: Interaction between disciplinary and interdisciplinary learning?

CFA indicates that the structure of the content knowledge items is well described by two latent variables: disciplinary and interdisciplinary knowledge. Results also indicate a moderate (Cohen's d ES = 0.635) and highly statistically significant correlation between the interdisciplinary and disciplinary factors on the posttest ($p < .001$). In the pretest, there was no significant differ-

TABLE 4

Item difficulties from classical test theory (% correct) and item response theory analysis (Rasch item difficulty parameter, b) on the pre- (left) and posttest (right). K = knowledge items.

	Pre		Post	
	% Correct	Item difficulty (b)	% Correct	Item difficulty (b)
K1 (D)	16.5	1.30	69.1	0.03
K2 (D)	34.7	0.18	69.8	0.00
K4 (I)	21.5	0.94	51.2	0.95
K5 (D)	73.6	–1.77	94.4	–2.26
K6 (D)	34.7	0.18	43.8	1.32
K7 (I)	47.5	–0.44	56.8	0.68
K8 (D)	26.4	0.63	61.7	0.43
K9 (I)	42.1	–0.19	96.3	–2.71
K10 (I)	53.7	–0.74	60.5	0.49
K11 (D)	48.3	–0.48	80.2	–0.67
K12 (I)	47.5	–0.44	58.0	0.62
K13 (I)	53.7	–0.74	74.1	–0.26
K14 (D)	7.4	2.29	47.5	1.14
K15 (I)	43.4	–0.25	49.4	1.05
K16 (D)	47.9	–0.46	82.1	–0.81

ence between the single-discipline scores and interdisciplinary scores: $t = 1.704$ ($df = 241$, $p = .09$). In the posttest, however, there was a highly significant difference between the single-discipline and interdisciplinary scores, with higher disciplinary gains on average: $t = 6.802$ ($df = 159$, $p < .001$).

Discussion

Instrument evaluation

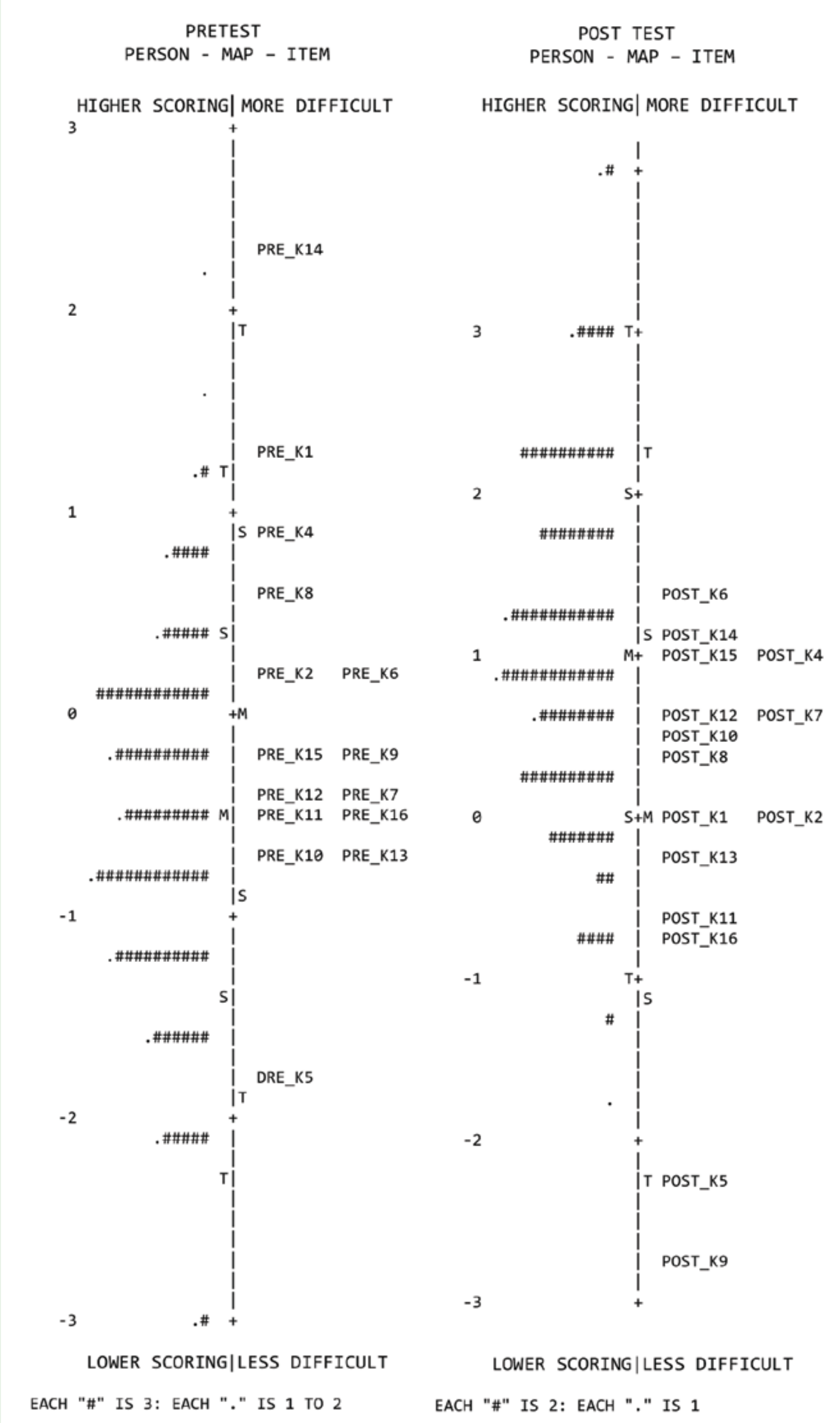
Our analysis supports the validity of the survey used in this study as a means of assessing sustainability learning and investigating the interaction of disciplinary and interdisciplinary learning. This instrument expands the number of vetted items in the environmental component of sustainability literacy over previous work (Zwickle et al., 2014) and permits an assessment of both interdisciplinary and disciplinary knowledge in additional areas (c.f., Shen et al., 2014). Parameters from the factor analysis were adequate to excellent, indicating an acceptable fit. Two items (K8 and K4) demonstrated a marginally significant difference from the model ($p = .083$ and $p = .06$, respectively) and merit further review. K4 is an interdisciplinary item, requiring synthesis of knowledge from biology, geology, and atmospheric science:

Most of the Earth's carbon resides in
A. I don't know
B. Soils and vegetation
C. The ocean
D. **Sedimentary rocks**
E. The atmosphere

It should be noted that experts involved in the study of the carbon cycle in various disciplines found this item difficult (You, 2016). It was among the most difficult items identified

FIGURE 1

Wright Item/Person Map showing distribution of item difficulties and student abilities for the pre (left) and post (right) tests.



by Rasch analysis on both pre- and posttests.

K8 is an item relating to engineering and was classified as measuring disciplinary knowledge: "The largest use of energy in the United States is currently____" It might be argued that this item should really be classified in the social domain (c.f., Zwickle et al., 2014). It tests knowledge of current energy use in the United States rather than STEM principles. Zwickle et al. (2014) argued against including time-sensitive items in sustainability assessments, as the goal should be to include only foundational knowledge questions requiring a deep understanding of STEM concepts.

The item that did not fit the two-factor model (i.e., "I know I can trust information in an article if the article is____") might also be classified in the social domain. It tests knowledge of the *practices* of science (Marshall, Erickson, & Sivam, 2015), as opposed to a scientific concept. Whether the assessment should be expanded to include this and additional items as part of a third domain merits further consideration.

The instrument also appears to be acceptable in terms of item difficulty, although the Wright Maps indicate a possible need for additional questions at extreme ends of the ability range. On the presurvey, a number of responses cluster at the low end of the ability range, but none of the items fall in this ability range, the easiest being K5 at –1.77. It is possible that the low-scoring ability responses do not represent true ability, but rather lack of effort in responding to the ungraded survey. In contrast, on the posttest there were a number of students with perfect scores. This indicates a "ceiling effect" and a possible need for additional items of greater difficulty.

Course outcomes

Postsurvey results show significantly higher scores on both attitudes and content knowledge than presurvey ($p < .001$ in both cases). The effect size was 0.550, indicating moderate practical significance, for the attitude component and 1.518, indicating high practical significance, for content knowledge. Pre/post comparison of survey results supports the efficacy of the course in enhancing sustainability literacy. However, further study using a matched-pairs analysis is certainly warranted.

Pre/post results show notable trends in regard to gender. First, the postsurvey did *not* show a gender difference in content knowledge in either the STEM or non-STEM populations, in contrast to typical outcomes in disciplinary science classes (Miyake et al., 2010). This argues that learning in an authentic, interdisciplinary context benefits students often marginalized in traditional STEM coursework (Mayberry, Welling, Phillips, Radeloff, & Rees, 1999).

In contrast, there was a gender difference in attitudes. Women overall scored higher on the attitude Likert-scale items compared with men in both pre- and postsurveys. This supports the contention that women associate greater importance with benefits to the environment, although the size of the gender difference is reported to depend on whether the metric is behaviors/concerns or activism (e.g., Agarwal, 2000; McCright & Xiao, 2014; Tindall, Davies, & Mauboules, 2003). Remington-Doucette and Musgrove (2015) also saw gender differences in favor of women, but only in interpersonal sustainability competency. However, it is also notable that in our study the difference was significant between STEM women and men on the postsurvey, but not between non-STEM women and men. STEM women may have demonstrated higher gains in attitudes because of the concentration of environmental science majors, a large majority of whom are women.

Relationship between disciplinary and interdisciplinary learning

Different relationships between disciplinary content learning and interdisciplinary content learning are possible: The two might be independent of each other; interdisciplinary learning might build on disciplinary learning or vice-versa. These models have implications for where students should experience sustainability coursework in their degree programs. Warburton (2003) argued that a disciplinary focus can inhibit "deep learning." Others have argued the need for a disciplinary grounding to enable interdisciplinary learning (Boix Mansilla & Duraising, 2007).

If interdisciplinary sustainability learning required disciplinary learning as a prerequisite, we might expect to see greater gains for STEM students than non-STEM on interdisciplinary assessment items, as the former will usually have experienced introductory STEM coursework between pre- and posttest, typically taking an introductory physics, chemistry, or biology class in their first semester and often having greater exposure in high school. Remington-Doucette et al. (2013) found non-STEM (business) majors did not improve in any competency measured, whereas sustainability majors improved in systems thinking and sustainability minors improved in all competencies. Our study, in contrast, found no significant interdisciplinary knowledge differences between STEM and non-STEM majors on the postsurvey. Therefore, if disciplinary grounding does enhance interdisciplinary learning, the additional exposure the STEM students had was not enough to make a difference or the effect is delayed. In contrast, the higher overall gains we saw on disciplinary knowledge compared with interdisciplinary might indicate that disciplinary understanding precedes interdisciplinary, as we have constructed it.

On the other hand, although we found no significant difference between STEM and non-STEM students' disciplinary knowledge on the pretest, on the posttest there was a mildly statistically significant difference with a moderate effect size, t = 2.046, df = 159, p = .042, ES = 0.3245). It is not surprising that students undergoing disciplinary coursework at the same time as the sustainability course might gain more in terms of disciplinary knowledge, as exposure to concepts in multiple contexts might enhance learning. Rogers, Pfaff, Hamilton, and Erkan (2015) found that disciplinary understanding was unchanged when sustainability was incorporated into STEM courses, but Fisher and McAdams (2015) found that disciplinary courses enhance understanding of sustainability *relevant to the course*, that is, the disciplinary component. To test whether interdisciplinary courses might also enhance disciplinary learning, comparison with a group not taking the sustainability class is needed.

Conclusions and limitations

This study validated an instrument to measure knowledge and attitudes. It expands the number of items to assess both disciplinary and interdisciplinary sustainability knowledge (c.f., Zwickle et al., 2014). Although the psychometric properties were acceptable, areas for improvement were identified, including the possible addition of items to measure a social/economic component and more difficult items to assess knowledge. Testing with larger populations and in different settings would substantiate these findings. A major limitation is that the instrument has only been tested in one course, with one team of instructors.

Results also support the value of a first-year, team-taught sustainability course in enhancing knowledge and attitudes, regardless of gender or major. However, a limitation was the lack of a matched-pairs sample. Finding no difference in interdisciplinary learning for STEM and non-STEM students speaks against a disciplinary knowledge prerequisite for sustainability education. Nevertheless, the higher disciplinary knowledge gains we saw overall may indicate that disciplinary learning precedes interdisciplinary learning. Studies of students experiencing coursework at different points in the undergraduate program are needed to test this possibility. ■

Acknowledgments

Support for this work was provided by a University of Texas (UT) Collaborative Teaching Grant, CTG2015-JSG-Banner (Measuring the Effectiveness of Interdisciplinary Courses), as well as the F.M. Bullard Professorship of the Geology Foundation and the Environmental Science Institute of UT Austin. We acknowledge the assistance of Ellison Carter and Angelina Jean Locker in acquiring the data.

References

Agarwal, B. (2000). Conceptualising environmental collective action: Why gender matters. *Cambridge Journal of Economics*, *24,* 283–310.

Astin, A. W., Oseguera, L., Sax, L. J., & Korn, W. S. (2002). *The American freshman: Thirty-five year trends, 1996–2001.* Los Angeles, CA: University of California Los Angeles, Higher Education Research Institute. Retrieved from https://www.heri.ucla.edu/PDFs/pubs/TFS/Trends/Monographs/TheAmericanFreshman35YearTrends.pdf

Boix Mansilla, V., & Duraising, E. D. (2007). Targeted assessment of students' interdisciplinary work: An empirically grounded framework proposed. *The Journal of Higher Education*, *78,* 215–237.

Carter, E. (2013, December 16). *Evaluating student learning in sustainability education*. Paper presented to the University of Minnesota Sustainability Faculty, Minneapolis.

Clark, W. C., & Dickson, N. M. (2003). Sustainability science: The emerging research program. *Proceedings of the National Academy of Sciences, USA, 100,* 8059–8061.

Cortese, A. D. (2003). The critical role of higher education in creating a sustainable future. *Planning for Higher Education*, *31*(3), 15–22.

Fisher, P. B., & McAdams, E. (2015). Gaps in sustainability education: The impact of higher education coursework on perceptions of sustainability. *International Journal of Sustainability in Higher Education*, *16,* 407–423.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48,* 1159–1176.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation

modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.
Kline, R. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.
Kopnina, H. (2012). Education for sustainable development (ESD): The turn away from 'environment' in environmental education? *Environmental Education Research*, *18,* 699–717.
Leiserowitz, A., Smith, N., & Marlon, J. R. (2010). *Americans' knowledge of climate change*. New Haven, CT: Yale University, Yale Project on Climate Change Communication. Retrieved from http://environment.yale.edu/climate/files/ClimateChangeKnowledge2010.pdf
Marshall, J. A., Erickson, T., & Sivam, K. (2015). Preservice teachers exploring the nature of science in simulated worlds. *International Journal of Gaming and Computer-Mediated Simulations, 7*(2), 24–45.
Mayberry, M., Welling, L., Phillips, J., Radeloff, C., & Rees, M. (1999). Feminism and science education: An interdisciplinary knowledge and practice project. *Journal of Women and Minorities in Science and Engineering, 5*, 1–16.
McCright, A. M., & Xiao, C. (2014). Gender and environmental concern: Insights from recent work and for future research. *Society & Natural Resources*, *27,* 1109–1113.
Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, *330*(6008), 1234–1237.
Muthén, L.K., & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén, Available at https://www.statmodel.com
Remington-Doucette, S., Hiller Connell, K. Y., Armstrong, C. M., & Musgrove, S. L. (2013). Assessing sustainability education in a transdisciplinary undergraduate course focused on real-world problem solving: A case for disciplinary grounding. *International Journal of Sustainability in Higher Education, 14,* 404–433.
Remington-Doucette, S., & Musgrove, S. L. (2015). Variation in sustainability competency development according to age, gender, and disciplinary affiliation. *International Journal of Sustainability in Higher Education*, *16,* 537–575.
Rogers, M., Pfaff, T., Hamilton, J., & Erkan, A. (2015). Using sustainability themes and multidisciplinary approaches to enhance STEM education. *International Journal of Sustainability in Higher Education, 16,* 523–536.
Shen, J., Liu, O. L., & Sung, S. (2014). Designing interdisciplinary assessments in sciences for college students: An example on osmosis. *International Journal of Science Education*, *36,* 1773–1793.
Stone, M. H., Wright, B. D, & Stenner, A. J. (1999). Mapping variable. *Journal of Outcome Measurement*, *3,* 308–322.
Sung, S., Shen, J., Stanger-Hall, K. F., Wiegert, C., Li, W., Brown, S., & Robertson, T. (2015). Toward interdisciplinary perspectives: Using osmotic pressure as an example for analyzing textbook explanations, *Journal of College Science Teaching*, *44*(4), 76–87.
Tindall, D. B., Davies, S., & Mauboules, C. (2003). Activism and conservation behavior in an environmental movement: The contradictory effects of gender. *Society & Natural Resources, 16,* 909–932.
Warburton, K. (2003). Deep learning and education for sustainability. *International Journal of Sustainability in Higher Education, 4*, 44–56.
Wright, B. D., Linacre, J. M., & Gustafson J.-E., & Martin-Löf, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.
You, H.S. (2016). *Toward interdisciplinary science learning: Development of an assessment for interdisciplinary understanding of 'carbon cycling.'* Unpublished doctoral dissertation, The University of Texas at Austin.
Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and* continuous outcomes. Unpublished doctoral dissertation, University of California at Los Angeles.
Zwickle, A., Koontz, T. M., Slagle, K. M., & Bruskotter, J. T. (2014). Assessing sustainability knowledge of a student population. *International Journal of Sustainability in Higher Education, 15,* 375–389.

Jill A. Marshall (marshall@austin.utexas.edu) *is an associate professor of STEM Education and* ***Jay L. Banner*** *is a professor in the Department of Geological Sciences and Environmental Science Institute, both at The University of Texas at Austin.* ***Hye Sun You*** *was a PhD candidate in STEM education at the time this article was written.*

Appendix: Survey ("A" connotes attitude items, and "K" connotes knowledge items).

Please indicate your gender.
- Male
- Female
- Other

Please indicate your intended area of study.
- Engineering
- Education
- Fine Arts
- Liberal Arts
- College of Natural Sciences (sciences, CS, mathematics)
- School of Geosciences
- Business
- Communication
- Social work
- UGS (School of Undergraduate Studies)
- Undecided

A1. How would you rate the importance to you of keeping up with political affairs?
- Not important at all
- Somewhat important
- Very important

A2. How would you rate your interest in raising a family someday?
- Not important at all
- Somewhat important
- Very important

A3. How would you rate the importance of being financially well off?
- Not important at all
- Somewhat important
- Very important

A4. How would you rate the importance of developing a meaningful philosophy of life?
- Not important at all
- Somewhat important
- Very important

A5. To what extent are the following environmental issues important to you: [scale: 1-not at all important, 2-somewhat important, 3-very important/essential]
A. Over-population
B. Finding alternative energy sources
C. Global climate change
D. Air pollution
E. Loss of biodiversity/species extinction
F. Water availability and quality

A6. To what extent are the following issues important to you: [scale: 1-not at all important, 2-somewhat important, 3-very important/essential]
A. Foreign relations
B. Economic issues
C. Environmental issues
D. Social justice issues
E. Moral issues
F. National security

A7. To what extent do you consider environmental issues faced by our society to be solvable?
- Not at all solvable
- Possibly solvable
- Very solvable

A8. To what extent do you consider participation by the following groups to be important with respect to solving environmental issues our society faces: [scale: 1-not at all important, 2-somewhat important, 3-very important/essential]
A. Government
B. Individuals
C. Businesses
D. Media
E. Education system
F. Celebrities championing a cause or causes
G. Charities

A9. To what degree are sustainability considerations important when making decisions about the following choices in your life: [scale: 1-not at all important, 2-somewhat important, 3-very important/essential]
A. Car purchase
B. Food/grocery products
C. Consumer goods
D. Career pathways
E. Voting

K1. Compared with other aquifers, limestone ('karst') aquifers usually are
A. more impacted by urbanization.
B. have flow paths that all have very similar characteristics such as permeability and porosity.
C. do not occur around major cities.
D. I don't know
E. None of the above.

K2. Which of the following affects global energy consumption the most?
A. I don't know.
B. Population
C. Wealth/income
D. Technological advances
E. B-D influence energy consumption to about the same extent.

K3. I know I can trust information in an article If the article:
A. was written by a professor.
B. contains quantitative information.
C. was published in a peer-reviewed journal.
D. All of the above
E. I don't know

K4. Most of the Earth's carbon resides in
A. I don't know.
B. Soils and vegetation
C. The ocean
D. Sedimentary rocks
E. The atmosphere

K5. It gets cold in the winter in [City] because
A. I don't know.
B. The Earth moves further from the Sun in its orbit.
C. The amount of carbon dioxide in the atmosphere increases in the southern hemisphere.
D. The Earth's axis of rotation is tilted.
E. None of the above

K6. Major greenhouse gases include all of the following except:
A. I don't know
B. Nitrogen
C. Water vapor
D. Carbon Dioxide
E. Methane (natural gas)

K7. Carbon cycling describes the movement of carbon (typically bound with other elements in compounds) through Earth's atmosphere, hydrosphere (oceans and other bodies of water), biosphere (plants and animals), and lithosphere (rocks and soils). Carbon exists in various chemical forms in each sphere. Which of the following carbon forms is NOT a primary form found in the sphere indicated?
A. Carbon dioxide in the atmosphere
B. Calcium carbonate the lithosphere
C. Organic matter in the biosphere
D. Bicarbonate and carbonate in the hydrosphere

E. None of the above; all are found in the sphere indicated.

K8. The largest use of energy in the United States is currently
A. Electricity generation
B. Transportation
C. Industrial energy use (other than electricity generation)
D. None of the above
E. I don't know.

K9.The dead zone in the Gulf of Mexico is caused by
A. I don't know.
B. fertilizer runoff into the Gulf causing bacterial blooms and oxygen depletion
C. acidification of the Gulf from acid rain
D. toxic pesticides entering the Gulf from the Mississippi River
E. warming ocean temperatures

K10. Which of the following things associated with climate change are known with high certainty?
A. CO_2 levels in the atmosphere have increased over the past 200 years.
B. If people continue to use energy at the current level, then CO_2 levels in the atmosphere will level off during this century.
C. The Greenhouse Effect came into effect with the onset of the Industrial Revolution.
D. If atmospheric CO_2 levels double this century, most types of vegetation will grow more slowly.
E. All of the above.

K11. Atmospheric circulation is caused by
A. I don't know.
B. Heat from Earth's interior
C. Pressure gradients caused by uneven solar heating
D. Lunar gravitational attraction
E. Plate tectonics

K12. The most popular timber product grown in the United States today is Pinus taeda known as Loblolly Pine. The pine trees' average height and circumference have been increasing since the 1960s. The level of CO_2 in the atmosphere has also been increasing rapidly since 1950. One scientist proposed that the increased tree growth is due to increased burning of fossil fuels.
A. This hypothesis does not make sense because burning fossil fuels releases CO_2, which causes acid rain and is harmful to plants.
B. This hypothesis does not make sense because burning fossil fuels releases CO_2 into the atmosphere, which is a pollutant that harms plant growth.
C. This hypothesis makes sense because increased CO_2 raises temperatures, and plants can convert more heat energy into chemical energy needed for plant growth.
D. This hypothesis makes sense because burning fossil fuels releases CO_2 into the atmosphere, which increases photosynthesis.
E. None of the above.

K13. If ice on the Earth's surface melts, it affects the Earth's albedo (i.e., the percentage of sunlight that the Earth reflects back into space). Which of the following statements would be true regarding the effect of melting ice and its effect on the Earth's albedo?
A. Melting ice would increase the Earth's albedo resulting in less heat absorbed, which in turn would result in more ice melting. This would be known as a positive feedback loop.
B. Melting ice would decrease the Earth's albedo resulting in more heat being absorbed, resulting in more ice melting. This would be known as a positive feedback loop.
C. Melting ice would increase the Earth's albedo resulting in more heat being absorbed, resulting in more ice melting. This would be known as a positive feedback loop.
D. Melting ice would increase the Earth's albedo resulting in more heat absorbed, which in turn would result in less ice melting. This would be known as a negative feedback loop.
E. Melting ice would decrease the Earth's albedo resulting in less heat being absorbed, resulting in less ice melting. This would be known as a negative feedback loop.

K14. Deserts
A. would form at the north and south poles if the Earth didn't rotate.
B. form today mostly at or near 30 degrees north and south latitude.
C. form where air from the upper troposphere loses its moisture, descends and warms.
D. All of the above
E. I don't know.

K15. Understanding environmental problems will be improved by
A. developing more efficient methods for extracting fossil fuels from geologic formations
B. building better and bigger water treatment facilities.
C. research studies that integrate multiple disciplines, such as biology, geology, chemistry and engineering
D. All of the above
E. None of the above

K16. Ocean water becomes denser
A. If it is cooled.
B. If it is warmed.
C. If it becomes saltier.
D. Both a and c
E. Both b and c

TABLE A1

Descriptions of interdisciplinarity (I for Interdisciplinary and D for items related to a single discipline) and answers for each item

Item	Interdisciplinarity	Answer
K1	D	A
K2	D	E
K3	I	C
K4	I	D
K5	D	D
K6	D	B
K7	I	E
K8	D	A
K9	I	B
K10	I	A
K11	D	C
K12	I	D
K13	I	B
K14	D	D
K15	I	C
K16	D	D