Assessing Sustainability Knowledge of a Student Population: Developing a Tool to Measure Knowledge in the Environmental, Economic, and Social Domains

(Forthcoming in the International Journal of Sustainability in Higher Education)

Adam Zwickle, Tomas M. Koontz, Kristina M. Slagle, Jeremy T. Bruskotter

Environmental and Social Sustainability Lab - http://ess.osu.edu
School of Environment and Natural Resources; The Ohio State University; Columbus, Ohio, United States

Corresponding author: Adam Zwickle
Zwickle.1@osu.edu

Acknowledgments:
We would like to thank the Office of Energy Services and Sustainability at The Ohio State University for their funding, encouragement and support.

Structured Abstract

Purpose:
In this article we present our tool for assessing the sustainability knowledge of an undergraduate population.

Design/methodology:
Multiple choice questions were developed through soliciting expert input, focus groups, pilot testing, distribution via a large scale online survey, and analysis using item response theory.

Findings:
The final assessment consists of 16 questions from the environmental, economic, and social domains covering foundational concepts within the topic of sustainability.

Research limitations:
This assessment represents an initial effort to quantify knowledge of the broad and abstract concept of sustainability. We plan to continue refining these questions to better differentiate between students with higher levels of knowledge and to replace those with answers that may change over time.

Practical implications:
With knowledge of sustainability concepts becoming increasingly included in institution-wide learning objectives there is a growing demand for a way to measure progress in this area. Our assessment tool can easily be used (via a campus-wide survey or distributed at the classroom level) by institutions to gauge current levels of knowledge, track changes over time, and assess the effectiveness of courses and curricula at meeting sustainability knowledge goals.

Originality:
This assessment of sustainability knowledge is the first of its kind to include all three separate domains of sustainability, and we expect it to be useful across a variety of college and university contexts.

Introduction

Colleges and universities are increasingly focusing on campus sustainability. On many campuses, a wide range of efforts from transportation and recycling to energy efficiency and water conservation are underway. In addition, the "sustainability across the curriculum" movement has emphasized the integration of sustainability concepts across majors, fields of study, and courses (Hopkinson and James 2010; Lidgren et al. 2006; Benn and Dunphy 2009). Such curricular efforts often aim to encourage students to gain sustainability literacy and adopt sustainable behaviors.

Along with the rise in campus sustainability efforts, the Association for the Advancement of Sustainability in Higher Education (AASHE) has developed and disseminated the Sustainability Tracking, Assessment, and Rating System (STARS). This rating system, adopted by over 200 institutions of higher education, awards points based on different sustainability-related practices (https://stars.aashe.org). One category, Education and Research (ER) Credit 13, "recognizes institutions that are assessing the sustainability literacy of their students." It goes on to state that "such an assessment helps institutions evaluate the success of their sustainability education initiatives and develop insight into how these initiatives could be improved" (STARS Technical Manual 1.2, p. 61). To earn points for this credit, an institution must conduct "an assessment of the sustainability literacy of its students. The sustainability literacy assessment focuses on knowledge of sustainability topics, not values or beliefs."

Given the increasing emphasis on campus sustainability, especially educational efforts to promote "sustainability across the curriculum," and the importance of assessing knowledge of sustainability topics, as indicated by the STARS rating system, colleges and universities need reliable and accurate means to assess students' sustainability knowledge. Unfortunately, tools to make such assessments are either nonexistent or their measures of knowledge rely on self-reported knowledge rather than more objective indicators. Thus we have developed an assessment tool to measure sustainability knowledge, and we have tested and refined this tool with a wide range of students at one of the nation's largest public universities. As described below, following prior literature, expert consultation, focus groups, and a pilot study, we developed and implemented a survey instrument with 30 questions that was subsequently analyzed using Item Response Theory (IRT) to select 16 questions for our final instrument. We subsequently show how our data can be used for some simple campus-wide comparisons, and we suggest avenues for further development.

Sustainability and Knowledge Assessment

The widely used term "sustainability" comes from the term "sustainable development," as used in the United Nations World Commission on Environment and Development report, *Our Common Future* (1987). A 2005 U.N. resolution (A/60/1) noted that sustainability was comprised of three key domains, environmental, economic, and social. Other organizations have identified these three domains as central to sustainability, for example the USEPA (n.d.) stated "Sustainability has many definitions but the basic principles and concepts remain constant: balancing a growing economy, protection for the environment, and social responsibility, so they together lead to an improved quality of life for ourselves and future generations." Some refer to these three domains as the "triple bottom line" of people, planet, and profit (Elkington, 1994). Readers interested in a thorough explication of each domain, (particularly the social domain), which is beyond the scope of this article, are encouraged to read the meta-analysis by Vallance, Perkins and Dixon (2011).

In recent years, several surveys have sought to assess knowledge about one sustainability domain – environmental. For example, a 2005 report on ten years of survey data from the Roper organization and the National Environmental Education and Training Foundation concluded that general environmental knowledge among adult Americans was low (Coyle 2005). The questions posed included those asking respondents to identify major sources of carbon monoxide, the definition of "biodiversity," and the function of Earth's ozone layer, among others. From a battery of 12 multiple-choice questions, two thirds of respondents were unable to achieve a "passing grade" (60% or more correct).

More recently, the Yale Project on Climate Change surveys have asked adult Americans to answer questions about global climate change. These true-false questions asked respondents 81 factual questions about causes and consequences of global climate change, for example, "The atmosphere carries heat from the north and south poles toward the equator " and "Global warming will cause some places to get wetter, while others will get drier." (Leiserowitz et al. 2010). Less than half of the respondents achieved a "passing grade" (60 % or more correct).

Assessing the other sustainability knowledge domains, however, is less common. A business report noted that "Most existing 'sustainability' management tools and systems seem to have been written by environmentalists and social scientists. Some refer to economic sustainability but are so patchy or vague that they would be inadequate for actually managing a real business" (Doane and MacGillivray 2001 section 5.0.1). The authors further state that "economic sustainability is the most elusive component of the 'triple bottom line' approach" (section 5.0.1). The social dimension is similarly under-assessed. In fact a recent issue (Winter 2012) of the journal *Sustainability: Science, Practice, and Policy* issue featured a special focus on social sustainability as a "missing pillar" in theorizing about sustainability. A telling example of this missing domain was when, in early discussions with our campus administrators who work in the area of sustainability, one asked, "I thought this was a survey about sustainability – why are you including the social questions?"

We did find one example of a sustainability knowledge survey that included more than just the environmental domain (Kim and Kozar 2012). This survey asked university apparel and textile students their level of agreement with factual statements about the use of child labor, workplace conditions, chemical pollutants, and recycling in textile production. Since this survey was specific to the apparel and textile industry, it would not be suitable for campus-wide use.

Developing and Implementing the Assessment of Sustainability Knowledge

Developing Knowledge Questions

Building onto previous knowledge assessments, we began with environmental knowledge questions developed by Coyle (2005) for the environmental domain. As Coyle's assessment was developed for an undergraduate student population it served as an appropriate starting point. Unfortunately, no similar knowledge assessments in the social or economic domains could be found.

For all three knowledge domains, we sought input from a range of experts across academic disciplines, including ecology, sociology, rural sociology, economics, business, forestry, political science, education, psychology, and anthropology. We developed questions to encompass fundamental concepts of sustainability in each domain. Additionally, we compared our questions to introductory textbooks in the relevant fields to ensure that the language and focus of the question were similar to what is being taught to undergraduate students.

With the exception of three questions (one in each domain) covering different aspects of the definition of sustainability, our goal was to use questions applying concepts in different ways. This approach requires both an understanding of the concept (which could be measured through a straight forward definition question) and some level of critical thinking to apply that concept to a specific context. For example, to measure whether a student understands the concept of economic and environmental externalities one could simply ask, "What is an externality?" Instead, we constructed a question in which the correct answer contained the concept of externality (see Table 1 question 13). To correctly answer this question a student must understand that there is pollution associated with producing energy as well as the fundamental concept of economic and environmental externalities.

[Insert Table 1 here]

A challenge in constructing multiple choice questions is the creation of incorrect answer options. Incorrect answers that are “too” incorrect make a question too easy, while incorrect options that are too close to the correct answer make the question too difficult or confusing. For some of our questions we did not expect students to already know the correct answer. By critically analyzing the answer options, however, it should be possible to make an educated guess. An example of this is the wealth disparity question (see Table 4 question 8). Logically, the wealthiest 20% of the U.S. population must hold greater than 20% of the privately held wealth, so the 20% option is clearly wrong. Additionally, a society where the wealthiest 20% of the population holds 35%, or even 50% of the wealth has a relatively even distribution of wealth. The higher level, 85 %, is considerably more uneven and has important implications for the social domain of sustainability. Finally, we provided a “don’t know” option for all of the questions to reduce random guessing and to reduce the uncertainty of determining whether an unanswered question was skipped or the respondent read the question but did not know enough to make an educated guess.

Refining the Assessment Tool

Following the construction of an initial pool of candidate questions, we held five rounds of focus groups. The focus groups consisted of 3-7 faculty members and 2-12 graduate students in the School of Environment and Natural Resources, most of whom were members of the Environmental and Social Sustainability Lab. The focus groups worked on eliminating and creating additional questions in an iterative fashion. Through focus group deliberation and consensus we reduced the pool of questions, eliminating those deemed to be too difficult, too easy, too specific, or not capturing a fundamental concept of sustainability.

Following the focus groups, we created a pilot questionnaire that we distributed to additional faculty members and graduate students, as well as select undergraduate students. This questionnaire was shared with content experts in the Ohio State University Office of Energy Services and Sustainability as well as faculty in various departments. Based on the results and feedback from these pilot tests we reduced our number of questions to 30. The remaining 30 questions were retained because they 1) covered core sustainability concepts within their domain, 2) were neither too easy nor too hard, and 3) resulted in an approximately even distribution across the three domains. Finally, we formatted the survey using the software package Survey Monkey (Finley 1999).

The 30 question assessment of sustainability knowledge (ASK) tool was placed at the beginning of a larger sustainability-focused survey emailed to 10,478 undergraduate students randomly selected from the over 40,000 currently enrolled undergraduate students at The Ohio State University. Following Dillman's tailored design method (Dillman et al. 2009), we had a notable university vice president send a letter containing the survey link and encouraging students to "Let their voices be heard." A reminder was sent every two days, at different times of day, to those who had not responded, for a total of five email contacts. Two weeks after the final reminder another email, this time from a member of our lab, was sent with a link to a much shorter "non-response" survey in order to compare respondents to non-respondents.

Survey response rates were relatively high for an online survey of the general student population (Dillman et al. 2009). In all, 2,019 students (19.3% of our sample) began the assessment, with 1,389 of those completing all 30 questions (13.3% of our sample). The dropout rate of survey participants was substantial, demonstrating a disadvantage of using a lengthy knowledge assessment tool. The 30 questions were spread over three pages, 10 questions per page. Each time a page was completed and respondents were presented with another page of questions, the dropout rate jumped. This was unique to the assessment portion, which was more cognitively taxing than the remaining eight pages of the survey (see Table 2).

[Insert Table 2 here]

With data collection complete, we were able to further refine the Assessment of Sustainability Knowledge using Item Response Theory to assess the information quality of our 30 questions. Our goal was to reduce the length of our assessment tool by retaining the most information-rich questions in each of the three domains. As described below, this process yielded a set of 16 questions in our final instrument.

Classical test theory provides a roadmap for assessing knowledge constructs. Under classical test theory, researchers start by specifying a construct or domain of knowledge to be measured (DeVellis 2003). In this early stage, the primary concern is reducing threats to validity by ensuring the domain being assessed is clearly specified. Next, researchers generate an “item pool”; that is, a group of response items that are believed to reflect the domain. Each of these items is a “parallel test” or “parallel indicator” of the domain of interest—meaning that each is considered at equally valid measure of that domain. The number of items needed to assess a domain is likely to vary, but the reliability of a measure should increase with more items (DeVellis 2003, Crocker and Algina 1986). The final stage of constructing a measure involves assessment of the items. This process can involve several steps, including expert review of items (to assess face validity), the use of factor analysis (to assess dimensionality), evaluation of scale length and reliability, etc. (see DeVellis 2003, Noar 2003). In recent years, Item Response Theory (IRT) has provided researchers additional tools for assessing both the scale (as a whole), as well as individual items.

Item response theory (IRT) starts with the assumption that all respondents have some level of the latent construct of interest (here, sustainability knowledge), and that this latent construct is normally distributed within the population of interest. The latent construct is denoted by ϑ, and a respondent's level of ϑ dictates how likely he or she is to respond correctly to an item. This likelihood of a correct response can be graphically depicted as an item characteristic curve (ICC), which approximates a logistic curve (*S*-shaped curve). This ICC (Figure 1, black line) is graphed with a y-axis representing the probability of a correct response and the x-axis representing ϑ, on a standard normal distribution typically ranging from -3 standard deviations below the mean (a very low level of ϑ), to zero (the average level of ϑ in the population), to 3 standard deviations above the mean (a very high level of ϑ). The point at which the ICC crosses the 50% probability of answering correctly indicates the level of ϑ a respondent would have if he/she answered the item correctly, and is called *b*. The slope of the curve, or discrimination parameter, called *a*, indicates how well the item discriminates between different levels of ϑ. For example, a flatter curve would not sharply discriminate between an average level of ϑ and one standard deviation above or below (Crocker and Algina, 1986). A final type of curve important to understanding IRT analyses is the information curve, which can be measured at both the item and the domain level (item level: Figure 1, dotted line; domain-level: Figure 2, solid black line). This curve is an indication of how much information an item or a set of items contains at different levels of ϑ (Embretson and Reise 2000). The blue line found in Figure 1 is simply the inverse of the ICC, and indicates the likelihood of incorrectly answering the item at a given level of ϑ.

[Insert Figure 1 here]

[Insert Figure 2 here]

While many IRT models exist, we chose the two parameter model appropriate for dichotomous responses (here, correct or incorrect on our sustainability knowledge assessment items (Thissen and Wainer 2001)). We assumed our three domains represent three separate latent constructs and treated each separately in our analyses. All IRT analyses were performed using the software package IRTPRO (Cai et al. 2011). Table 3 provides the difficulty and discrimination parameters for each item. To arrive at our final set of questions for each domain, we retained a combination to provide relatively higher discrimination values and as broad of a range of difficulty levels as we could while retaining questions that dealt with fundamental concepts of sustainability.

[Insert Table 3 here]

Results and Discussion

Comparing respondents to non-respondents

The results from the non-response data suggest that the sample drawn by our initial survey was more sustainably knowledgeable (responding, on average, with significantly higher scores on five out of the six questions, $p < 0.05$) and slightly more environmentally oriented (identifying more as an environmentalist, but $p > 0.05$) than the rest of the undergraduate student population, but had no differences in GPA. Therefore, these responses should be considered an upper bound of that found in the general student population.

Example Analyses

Our tool was developed as a means to obtain useful data about sustainability knowledge. In addition to fulfilling STARS ER Credit 13, the data generated can be useful for institutions interested in analyzing various aspects of sustainability knowledge among different student populations. For example, knowledge scores can be compared across the three domains, across students by class rank (freshman, sophomore, etc.), or across academic programs. Below we describe such analyses of our data for illustrative purposes.

Knowledge across domains

The average score for the final 16 question assessment by the 1,389 undergraduate students was 11.08 (69% correct, SD=3.21). When looking at the three separate domains, students scored similarly on the environmental and economic questions, and lower on the social questions (see Table 4). For more detailed results for each question see the Table 1.

[Insert Table 4 here]

Knowledge across class rank

Looking at the overall scores combining the three domains of sustainability knowledge, upperclassmen at Ohio State demonstrate greater levels of knowledge than underclassmen (ANOVA test, $F(3, 1330) = 8.09$, $p < 0.001$, see Figure 3. Post hoc tests showed significant differences between freshmen and juniors ($p < 0.01$) and freshmen and seniors ($p < 0.001$)). While knowledge increases for each class rank, the biggest difference is between freshmen and sophomores. Note that a one-time survey represents a snapshot in time and to make more substantial claims about how students are learning over time, a longitudinal comparison of the same student's knowledge at two or more times throughout their academic career should be made.

[Insert Figure 3 here]

Domain knowledge across academic programs

Knowledge scores can be grouped into the three separate domains and compared across students pursuing different majors. For example, it would be expected that economics majors would score higher on the economic domain questions than would students in other majors. Interestingly, this expectation was not fully met in our data, as the highest scoring major was aeronautical engineering (see figure 4).

[Insert Figure 4 here]

Limitations of the assessment tool

As an initial effort, our assessment tool is not without limitation. First and foremost are the limitations involved with using a multiple choice format to measure knowledge of a concept as broad and complex as sustainability. To truly be sustainably literate one must be able to integrate knowledge from the environmental, economic, and social domains and put this knowledge into practice. Although some of our questions incorporated integrative concepts, a multiple choice format is limited in its ability to assess knowledge at multiple levels of learning, such as analysis, synthesis, and evaluation. Thus this assessment tool cannot provide a full assessment of sustainability knowledge. Rather, it provides one way to engage in systematic comparisons across populations and times. While the assessment may not capture all of a student's knowledge, our IRT analysis shows that it does provide consistent data about students' sustainability knowledge across the three domains.

Conclusion

Sustainability literacy is a common goal across colleges and universities. As exemplified in the STARS rating system, assessing levels of sustainability literacy is a valuable action. Unfortunately, there is no widely agreed upon assessment tool for institutions to use to measure all three domains of sustainability knowledge: environment, economy, and society. Thus we have developed an assessment of sustainability knowledge (ASK) to assist in this effort.

The instrument presented here is the product of a rigorous development process comprised of existing question items, input from experts, iterative focus group testing, pilot testing, broad scale distribution, and finally the use of IRT to select the most discriminating questions. Throughout the process of developing this assessment we were required to make trade-offs between creating a more complex tool that captured more information, and a simpler tool succinct enough to foster higher response rates and be more readily analyzed by those without extensive social science expertise. In these cases, we nearly always ceded complexity for simplicity.

Although the assessment tool is relatively short, at only 16 questions, and it attempts to measure knowledge of a large and complex concept, we find that this is an acceptable trade-off in order to make it broadly applicable. This assessment can readily be conducted, whether via online survey as we did or in the classroom, on a regular basis. The survey results can easily be interpreted and used to help gauge the effectiveness of specific courses, programs, or educational goals.

The assessment tool has room for further refinement. While our IRT analysis shows that we are effectively capturing the lower ends of theta, allowing us to distinguish between those who possess varying degrees of below average sustainability knowledge, we have few questions capturing the upper ends of theta. We are working to develop a few more difficult questions that provide better information about students whose knowledge level are above the mean. Additionally, some of our questions have answers that may change over time. Any time a question is replaced in the assessment, it makes comparing the results over time more difficult. Therefore our goal is to include only foundational knowledge questions which are not bounded by time.

To further refine our assessment, establish a broad baseline of data, and compare our student population with other student populations, we are working with several colleges and universities in the United States and overseas. This raises and important question about the degree to which sustainability knowledge is contextually based. Many of our questions, such as the level of wealth inequality, ask about the U.S. context. Can this question be used effectively in other countries by substituting different countries into the question, or is wealth inequality a less critical piece of sustainability knowledge outside the U.S.? At the same time, other concepts such as the definition of economic sustainability seem to be less context-specific. We hope and expect that testing the applicability and utility of our assessment tool beyond our university will prompt deeper thinking about how we conceptualize and measure sustainability knowledge across all three domains.

REFERENCES

Association for the Advancement of Sustainability in Higher Education (2012), "Version 1.2 Technical Manual: February 2012", available at: http://www.aashe.org/files/documents/STARS/stars_1.2_technical_manual.pdf (accessed 24 January 2013).

Benn, S. and Dunphy, D. (2009), "Action research as an approach to integrating sustainability in MBA programs: an exploratory study", *Journal of Management Education*, Vol. 33 No. 3, pp. 276-95.

Brundtland Commission, (1987) *Our Common Future, The Report of the World Commission on Environment and Development*, Oxford University Press, Oxford.

Cai, L., Thissen, D. and du Toit, S (2011), IRTPRO 2.1 [computer software], Scientific Software International, Inc., Skokie, Illinois.

Crocker, L. M. and Algina, J. (1986), *Introduction to classical and modern test theory*. Holt Rinehart & Winston, New York, New York.

DeVellis, R. F. (2003), *Scale Development: Theory and Applications, 2nd edition*, *volume 26*, Sage, Thousand Oaks, California.

Dillman, D. A., Smyth, J. D., Christian, L. M, (2009), *Internet, mail, and mixed-mode surveys : the tailored design method*. Hoboken, N.J., Wiley & Sons.

Doane, D and MacGillivray, A, (2001), Economic Sustainability: The Business of Staying in Business. The Sigma Project R & D Report, March, available at: http://projectsigma.co.uk/RnDStreams/RD_economic_sustain.pdf (accessed 24 January 2013)

Elkington, J. (1994), “Towards the Sustainable Corporation: Win-Win-Win Business Strategies for Sustainable Development”, California Management Review 36, no. 2: 90–100.

Embretson, S. E. and Reise, S. P. (2000), *Item response theory for psychologists (Vol. 4)*. Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey.

Finley, R. (1999), SurveyMonkey [computer software], available at: http://www.surveymonkey.com (accessed 24 January 2013)

Hiller Connell, K.Y. and Kozar, J. M. (2012), “Sustainability knowledge and behaviors of apparel and textile undergraduates”, *International Journal of Sustainability in Higher Education*, Vol. 13 No. 4, pp. 394-407.

Hopkinson, P. and James, P. (2010), “Practical pedagogy for embedding ESD in science, technology, engineering and mathematics curricula”, *International Journal of Sustainability in Higher Education*, Vol. 11 No. 4, pp. 365-79.

Leiserowitz, A., Smith, N. and Marlon, J.R. (2010) *Americans' Knowledge of Climate Change.* Yale University. New Haven, CT: Yale Project on Climate Change Communication. Available at: http://environment.yale.edu/climate/files/ClimateChangeKnowledge2010.pdf (accessed 24 January 2013).

Lidgren, A., Hakan, R. and Huisingh, D. (2006), "A systematic approach to incorporate sustainability into university courses and curricula", *Journal of Cleaner Production*, Vol. 12, pp. 797-809.

Noar, S. M. (2003), "The Role of Structural Equation Modeling in Scale Development", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 10, Issue 4, pp. 622-647.

Thissen, D. and Wainer, H. (2001), *Test scoring*. Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey.

United Nations General Assembly, Resolution A/60/1 World Summit Outcome, Page 2, http://unpan1.un.org/intradoc/groups/public/documents/un/unpan021752.pdf (accessed 3/11/2013)

United States Environmental Protection Agency, No date, "What is Sustainability?" http://yosemite.epa.gov/r10/oi.nsf/b724ca698f6054798825705700693650/2da335d4a914342c88256fc400784f91!OpenDocument (accessed 24 January 2013).

Vallance, S., Perkins, H. C., & Dixon, J. E. (2011). What is social sustainability? A clarification of concepts. *Geoforum*, Vol. 42, Issue 3, pp. 342-348.