JOURNAL OF BALTIC SCIENCE EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

A BIOENERGY COMPETENCY ASSESSMENT TOOL: DEVELOPMENT AND VALIDATION

Tingxuan Li,
Kari L. Clase,
Weiling Li,
Anne Traynor

Abstract. *This research is motivated by the perspective that when empirical studies and assessment frameworks inform each other, assessments can enrich science education and strengthen its connections to modern science. The research proposes a bioenergy competency assessment for science education. It uses an argument-based approach to validation. Multiple types of validity evidence were collected to support the proposed scores use and scores interpretation. Along with reporting a series of psychometric properties of response data, the findings indicated that empirical response data corresponded to the hypothesized data structure. The latent logistic scale estabished through a generalized partial credit model (GPCM) seemed useful in measuring students' bioenergy competency.*

Keywords: *competency assessment, item response theory, K-12 education, validity evidence.*

Tingxuan Li, Kari L. Clase, Weiling Li, Anne Traynor
Purdue University, USA

Introduction

The world's future energy supply is an important topic in K-12 science education because it highlights the interaction between science, technology, and human activities (National Research Council [NRC], 2012). Topics related to future energy supplies have been intensely studied by modern science (Karlen et al., 2014; Matzenberger et al., 2015; US Department of Energy, 2017). To maintain relevance and reflect modern scientific discovery, K-12 science classrooms should help students understand the role that science and technology play in development of renewable resources (Bowman & Govett, 2015; Next Generation Science Standards [NGSS] Lead States, 2013). Bioenergy, "renewable energy derived from recently living biological material" (Dahiya, 2014, p. 2), has strong potential as a subject that can illustrate the integration of scientific knowledge and its application to practical problems (Metz, 2011). However, recent studies have found that students are not aware of the opportunities and challenges that exist in developing renewable energy. For example, Halder, Pietarinen, Havu-Nuutinen, and Pelkonen (2010) conducted a survey in Finland and found that two-thirds of the students in middle and high schools were not aware of how bioenergy production impacts society at a local, national, and global level. They have stated that bioenergy-related economic, political, and environmental impacts can serve as an important context to foster *scientific literacy*.

Bioenergy educational units are able to cover core disciplinary knowledge required in science education, such as the *carbon cycle* and *photosynthesis* (Krauskopf, 2010; Metz, 2011). Additionally, learning experiences in bioenergy can spur students' interests in renewable energy careers; students may become more engaged in science by seeing its connections to the real world (Williams, Papierno, Makel, & Ceci, 2004). For example, Alaska Airlines in 2016 used the first commercial flight where part of fuel energy was made of wood waste (US Department of Agriculture, 2018). Bioeconomy has created 280,000 jobs in the US, which mainly came from the ethanol or biodiesel

https://doi.org/10.33225/jbse/19.18.264

production (US Department of Energy, 2018). In a broader sense, science education does not merely prepare future researchers in a discipline, but also equips the public to make informed decisions in a democratic society (Albe, 2013). The current issues in bioenergy should be understood by future citizens and consumers in terms of biofuels' costs and benefits (Parker, de Los Santos, & Anderson, 2015).

Research Focus

Several bioenergy educational centers (e.g., Oregon State University Bioenergy K-12 Education, 2016; Great Lakes Bioenergy Research Center, 2010) have offered bioenergy curriculum material on their websites for use by K-12 science teachers. Each curriculum had an embedded set of assessment items. However, documentation about the design of these assessments, and validity evidence to support their use, was not available. Other existing assessment frameworks have emphasized energy-related concepts across the sciences or attempt to establish students' learning progressions in understanding energy concepts (Jin & Anderson, 2012; Neumann, Viering, Boone, & Fischer, 2013) by modifying existing items from large-scale national assessments (Lee & Liu, 2009). None of them has measured students' knowledge of *renewable energy*, specifically.

To address the issues mentioned above, this research aims to develop a coherent assessment tool in bioenergy for high school students. The anticipated end-users of this assessment tool will be classroom teachers who integrate bioenergy concepts into regular science teaching. Thus, the proposed *scores use* was that teachers can gain information about students' competency in bioenergy. The proposed *scores interpretation* was students' bioenergy competency. As an initial validation study, the following aspects were reported in this study: (1) the assessment design process and (2) psychometric properties of response data including establishing a latent logistic scale for students' bioenergy competency from generalized partial credit modeling (GPCM, Muraki, 1993). In validating the proposed scores interpretation, multiple validity evidence was collected by following the guideline provided by the *Standards for Educational and Psychological Testing* (hereafter referred to as *the Standards*, AERA, APA, & NCME, 2014).

Research Methodology

General Background

This research was embedded in a teacher professional development (PD) program—*Research Goes to School* for high school science teachers. The purpose of this program was to bring cutting-edge research into classroom activities based on nationally-recognized science standards as well as use problem-based learning (PBL) instructional approach. It may aid teachers in developing a comprehensive understanding of recent developments and discovery in science. Research lectures were presented to teachers by professional scientists from the C3Bio research center at Purdue University. The lecture material was then re-created into a set of high school instructional units by teachers and the scientists together. The PD sessions guided teachers to (a) develop instructional units, each consisting of a series of scenario-based tasks, and (b) experience firsthand how students might learn the content knowledge via the units. The units were implemented in teachers' classrooms in the following academic semester. Hence, high school students received bioenergy instructional units as part of the regular science curriculum. In addition to the bioenergy assessment tool examined in this research, a *self-reported bioenergy learning experience questionnaire* was designed for students to complete at the end of the instruction. The questionnaire contained seven items measuring the extent to which students increased their understanding of the topics covered. For instance, one of the items was: *I learned how the biofuels concepts are relevant in everyday life*. The items were rated on a 4-point Likert scale ranging from "strongly disagree" to "strongly agree."

Claims

The technical quality of classroom-based assessments in STEM, as in other areas, often is not well-established, and sometimes may be lacking (Liu, Lee, Hofstetter, & Linn, 2008). The assessment design process can contribute to the validity of any meaning(s) attributed to the resulting scores. Validating score meaning is not mechanically conducting a series of analyses after assessment administration, but rather as intentionally structuring design activities "in such a way that validity evidence emerges" (Mislevy, 2007, p. 467). In this research, the target latent variable being assessed was *students' bioenergy competency*; the target grade-band for assessment was ninth

https://doi.org/10.33225/jbse/19.18.264

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

through twelfth grade; the proposed scores interpretation was students' bioenergy competency. As described by Kane (2013), the validity of a proposed scores interpretation "depends on how well the evidence supports the claims being made" (p.1). Thus, the claims should be put forth in validating any newly developed assessment tool.

In an achievement testing design, a *claim* about score meaning reflects students' competency with respect to specific learning standards that have been put forth (Kane, 2013). For example, a possible claim in math could be: *Students are able to identify when two expressions are equivalent* (e.g., $x + x + x = 3x$). The score interpretation on the math assessment, therefore, involves claims about students having some particular standing on computational fluency. In *an argument-based approach to validation* (Kane, 2013), what is valued in instruction, what responses the items/tasks elicit, and what the assessment is designed to capture are all tied together by such claims. In this study, the main claims about score meaning referred to students' mastery of specific learning objectives in bioenergy. By using both national science education standards (NRC, 1996) and their accompanying classroom instructional planning guide (NRC, 2000), four claims were derived to define students' bioenergy competency. The score interpretation hence involved these four claims about students. In addition, a *specification table* (Table 1 below) is provided because it is a common approach to map out the content areas of an assessment (e.g., DiDonato-Barnes, Fives, & Krause, 2013). It showed the intended distribution of tasks across content in the bioenergy assessment.

Claim 1: The student is able to describe fermentation and photosynthesis.
Claim 2: The student is able to construct explanations regarding organisms.
Claim 3: The student is able to recognize the flow of energy within a system.
Claim 4: The student is able to construct explanations regarding the characteristics of biofuel production systems.

Testable Elements

Within each claim, a set of testable elements was adopted to further specify how the science standards were unfolded in the assessment. The national standards are written as individual statements, which provide limited guidance to establish a coherent assessment to measure particular science competencies. In this design activity, the existing literature was used to identify how bioenergy knowledge was unfolded in science classrooms and understood by professional scientists. The purpose of this design activity was to create the fine-grained categories which can be used in assessment design process. A short review is presented below to show the literature used to extract testable elements. Figure 1 shows the testable elements within each claim. The testable elements guided the next design activity - item development.

Table 1. Bioenergy competency assessment specification table.

Content	Proportion %
Matter Cycles	25
Energy Flows	60
Production and Distribution of Biofuels	15

Short review. Bioenergy such as ethanol or biodiesel is currently available in the market. The professional scientists aim to find the next generations' bioenergy by examining different types of biomass which include algae. Biomass such as wood chips have potential as an energy source because it does not compete with food sources (e.g., sugarcane). Some types of biomass (e.g., corn) may take away considerable nutrients from the soil. This poses a dilemma for both farmers and policymakers; communities need to consider the trade-off between agricultural resources such as land quality (or water availability) and financial gains (Wu et al., 2018). It is critical to establish a system for making projections about the complex interactions among food security, bioenergy sustainability, and resources management on a local, national, and global level (Hammar et al., 2019).

Matter, energy, and organization in living systems is core disciplinary knowledge in K-12 biology education (Bybee, 2014). Energy is required for organisms to live and grow, and energy transfer occurs at different organizational levels from cells to ecosystems. Lin and Hu (2003) have found that most students failed to describe the

https://doi.org/10.33225/jbse/19.18.264

inter-relationship among energy flow and matter cycling sub-topics because their understanding about the living world was lacking an integrated perspective. Chabalengula, Sanders, and Mumba (2012) have found that students had misconceptions about energy in biological phenomena; students thought living matter had a unique kind of energy distinct from that found in non-living matter. Even most freshmen in college biology class held such misconceptions of *vitalism* (Barak, Gorodetsky, & Chipman, 1997), although biological phenomena can be understood through the same energy concepts applied to physics and chemistry.

An energy efficiency system needs to be evaluated with an infrastructural approach. The US Environmental Protection Agency mandates a tool, biofuel Life Cycle Assessment (LCA), to identify the environmental impacts of potential biofuels from biomass acquisition to their manufacturing, use, and final disposition (Dunn, 2012). In K-12 science education community, *tracing energy and carbon cycle* can be naturally embedded in LCA; the instructional units with LCA can help students understand tracing energy and carbon through a fuel-production system. Krauskopf (2010) has found that when determining which fuel was better, students tended to solely reason based on miles per gallon; they failed to possess *scientific habits of mind*, that is a way of conceptualizing a system. For example, a fuel production system included "all of the steps before fuel is even put into the car—beginning, for example, with petroleum extraction, ocean transport, refinery operation, and the gasoline's transportation to the pump" (p. 35).

Items and Scoring Schema

Items. In the previous design activity, the bioenergy competency assessment framework (i.e., Figure 1) was introduced. That structure guided the subsequent design activity—the design of items and scoring system. The advisory panel on the assessment development consisted of the professionals (e.g., scientists) in bioenergy and local science teachers outside. The discussions with the panel occurred throughout the design process in order to carry out the various design activities. Based on the feedback provided by the panel, the claims and national science standards seemed aligned to describe bioenergy competency. It seemed likely that the responses elicited by these testable elements would reflect students' bioenergy competency.

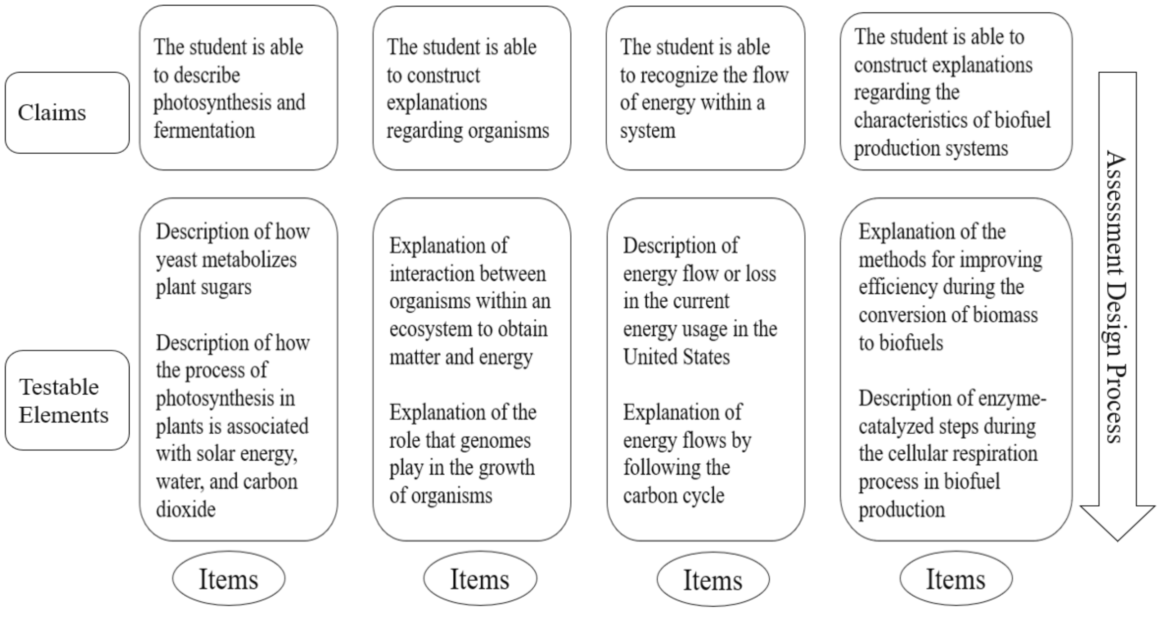

Figure 1. Bioenergy competency assessment.

The subsequent design decision made in this study was that the tasks should be both multiple-choice and open-ended items. The assessment delivery mode should be a paper-and-pencil mode. Some items were adapted or adopted from copyrighted sources, with permission from the owners. In total, 24 multiple-choice items and 5 open-ended items were developed. For multiple-choice items, students were asked to choose the correct answer from distractors. For open-ended items, students were asked to provide a short writing product,

https://doi.org/10.33225/jbse/19.18.264

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

which would be scored by item-specific rubrics developed in this study. Table 2 below shows the rubric for an item that was generated from Element 1 in Claim 4; it required students "to self-regulate their own cognitive process in order to monitor, and, if necessary, adjust their approach" (Tekkumru-Kisa, Stein, & Schunn, 2015, p. 666). The argumentative nature of this item required students to recognize that energy is lost during energy conversion; also to make sense of the fact that improving efficiency of the biofuel production system can be achieved through reducing energy waste and maintaining the maximal amount of energy input jointly; students were required to possess both declarative and procedural knowledge in order to fully communicate an argument that addresses the writing prompt.

Scoring. In the field testing procedure, one student was asked to complete all the items. The teacher also asked the student to provide feedback after trying the items. Based on the field notes taken by the observer, it appeared that the student had no difficulty understanding the items. At the end of bioenergy classroom instruction, response data were collected from 135 students. The students completed both *bioenergy competency items* and the *self-reported bioenergy learning experience questionnaire*. For bioenergy competency items, each multiple-choice item was scored as 1 for the correct answer or as 0 otherwise. For open-ended items, a numerical score of 0, 1, or 2 in the item-specific rubric was assigned. Two trained raters who were graduate students in biology scored the open-ended items for all students. The percentage of agreement between the two raters was 99.92%. Using the Krippendorff's alpha as an inter-rater reliability index (Hayes & Krippendorff, 2007), the magnitude of the index on this sample was 0.91.

Table 2. A scoring rubric for an open-ended item.

Scoring category	Level of competency	Description	Examples
2	Full	Fully elaborate relevant ideas by developing the linkage between the energy waste and energy input	• Energy is wasted during each step of the biofuel production system and these losses accumulate. For example, energy is wasted heavily at the combustion step. Biotechnology can be applied to avoid the combustion step completely, thus ensuring the maximal amount of energy input for the crop processing step. Additionally, the site at which the crops are harvested should be near to the crops' processing site. This minimizes the amount of energy wasted due to transportation. Overall, energy efficiency is improved through the reduction of energy waste, while maximizing the amount of energy input into each step.
1	Partial	Elaborate relevant ideas, but fail to develop the linkage	• Energy waste may be reduced during the energy distribution step. For example, designing a drop-in biofuel that is compatible with existing infrastructure can maximally eliminate energy waste. • Crop yields may be increased using genetic engineering (or in a broader sense biotechnology) to alter the plant genomes, resulting in plants that are hardier or more amenable for use in biofuel production.
0	Minimal	The response is incorrect or blank	• Growing a variety of crops • Using solar energy to process the crops

Research Results

Internal Structure

One additional source of the validity evidence that can be examined is the consistency between the internal structure of the observed response data and the hypothesized structure (the *Standards*, AERA, APA, & NCME, 2014). In the bioenergy competency assessment tool, all the items were designed to measure a single latent variable, so students' response data were hypothesized to have a one-dimensional structure. In order to determine the dimensionality of the empirical data, an *exploratory item response theory* (IRT*)* approach (Bock & Aitkin, 1981) was used in this study for the observed response data. When an exploratory IRT model is applied, as in exploratory factor analysis, the number of latent dimensions is assumed unknown.

Although the hypothesized dimensionality of the response data was one, because the instrument was newly developed, it was reasonable to empirically determine the number of latent dimensions. The basic steps included

https://doi.org/10.33225/jbse/19.18.264

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

choosing an IRT model and estimating the parameters in a one-factor exploratory model, a two-factor exploratory model. The maximum number of factors considered in this study was 3. Because the scoring system produced both binary and polytomous response data, the GPCM was used for item analysis and to establish the latent competency scale. The GPCM treats polytomous data by decomposing the responses into a series of ordered pairs of adjacent categories and then applies a dichotomous model to each pair. The *step difficulty parameter* for each category varies across items. For a multiple-choice item, there is only one pair of categories, which leads to one step difficulty parameter per item. For open-ended items, there are two pairs of categories, which lead to two step difficulty parameters per item. Unlike the partial credit model (PCM, Masters, 1982), the GPCM estimates one item-level discrimination parameter to represent the degree that each item differentiates on the latent competency across all the students. Because the instrument was newly developed, it was reasonable to examine the magnitudes of the discrimination parameter, rather than assuming they were equal across items.

Using a one-factor, and a two-factor exploratory framework for the GPCM with IRTPRO 2.1 (Cai, Thissen, & du Toit, 2011) software. The former model was nested in the latter model. Table 3 shows the Akaike information criterion (AIC) and Bayesian information criterion (BIC) model comparison criteria for the two models. The AIC and BIC perform well as model fit criteria, particularly when the response data are polytomous (Kang, Cohen, & Sung, 2009). For both indices, a smaller value is associated with better model performance (Lee & Ghosh, 2009). According to Table 3, one-dimensional GPCM was a more likely population model for the data than the two-dimensional GPCM. The internal structure of the observed response data therefore appearred to be consistent with the hypothesis that the assessment was measuring a single latent variable-bioenergy competency.

Table 3. Model comparison criteria.

Index	One-dimensional model	Two-dimensional model
AIC	4525.05	4530.15
BIC	4708.08	4794.53

Item Calibration

In order to provide item parameter estimation results and to generate a scoring report for each student in terms of his or her competency for bioenergy, item calibration was conducted using the one-dimensional GPCM. Equation 1 shows the parameters in the model, where each student *i* with latent competency θ provides a response in each item *j*'s *k*th category. The step difficulty parameter for each category *k*, which varies across items, is denoted as δ_{jk}. It is the relative difficulty in endorsing category *k* over the category (*k* -1). The item-level discrimination parameter is α_j which represents the degree each item differentiates on the latent variable.

$$P_{jk}(\theta_i) = \frac{\exp\left[\alpha_j(\theta_i - \delta_{jk})\right]}{1 + \exp\left[\alpha_j(\theta_i - \delta_{jk})\right]} \quad \text{(Equation 1).}$$

The item difficulty parameters are expected to spread along the latent scale. If most of the items are packed at one end as "difficult items" or at the other end as "easy items," a corresponding "floor effect" or "ceiling effect" may occur. Table 4 shows that difficulty parameters spread out reasonably well, suggesting that the items have optimal difficulty for students with a range of different bioenergy competency levels. For item discrimination power, Baker (1992) has suggested the following cut-offs: 0.65–1.34 for moderate power; 1.35–1.69 for high. Table 4 shows that among all the items, Item 11 is below the cut-off value, indicating low (insufficient) discrimination power to differentiate among students with different competency levels.

https://doi.org/10.33225/jbse/19.18.264

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Table 4. Item parameter estimation results.

Item ID	Discrimination Parameter	Difficulty Parameter Step 1	Difficulty Parameter Step 2
1	**1.81**	**-0.96**	**-**
2	0.78	1.16	-
3	3.28	-0.87	-
4	1.47	-1.05	-
5	0.94	-0.88	-
6	1.90	-1.06	-
7	1.74	-0.51	-
8	0.95	-0.91	-
9	0.81	0.95	-
10	1.11	-1.27	-
11	0.42	-0.68	-
12	0.81	-0.11	-
13	1.95	-0.47	-
14	1.25	0.71	-
15	2.42	-0.33	-
16	0.72	0	-
17	0.82	1.40	-
18	2.15	-0.50	-
19	0.72	-0.14	-
20	0.76	-0.57	-
21	2.09	-1.05	-
22	2.69	-0.58	-
23	1.44	-0.69	-
24	0.85	1.94	-
25	1.55	0.03	1.67
26	1.31	0.09	1.36
27	1.61	0.27	1.54
28	1.35	-0.07	1.55
29	1.07	0.06	1.50

Figure 2 is a visual representation to show the overview of parameters for the multiple-choice items. Each dot on the grey line represents the difficulty magnitude for each item. These magnitudes are sorted according to the items' discrimination parameters; the items at the left side have lower discrimination power than the items at the right side; the very left item on Figure 2 is item 11 which has the lowest discrimination magnitude (as shown in Table 4).

https://doi.org/10.33225/jbse/19.18.264

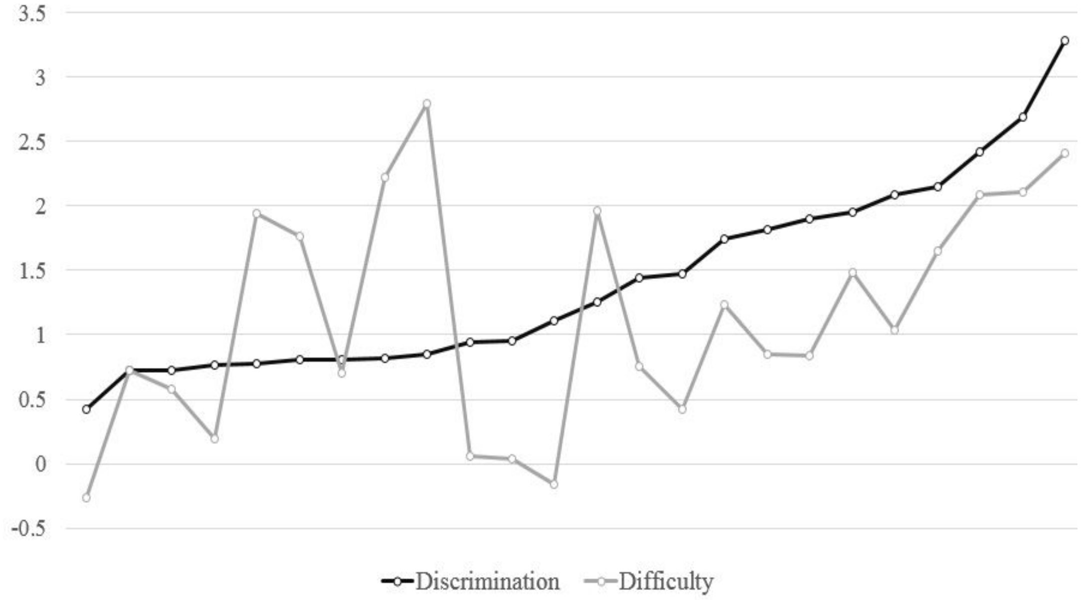

Figure 2. Item discrimination parameter and item difficulty parameter.

After item calibration, the students' competencies were represented by the latent scores $\theta_i \theta_i$ estimated from the GPCM. Therefore, the latent logistic scale in bioenergy competency was established. It was also reasonable to compute the test-level statistics to determine the range of the scores through which the assessment provided the most precise measurement. *Information* refers to the certainty about the estimated latent scores (de Ayala, 2009). The test information function is defined as the sum of the item information values obtained using the Fisher information matrix. Figure 3 below shows the test information curve for the bioenergy assessment. It suggests that the bioenergy assessment tool is able to provide the most information over the latent competency range from -1.5 to 0.8; the latent scores have greater measurement precision in this range.

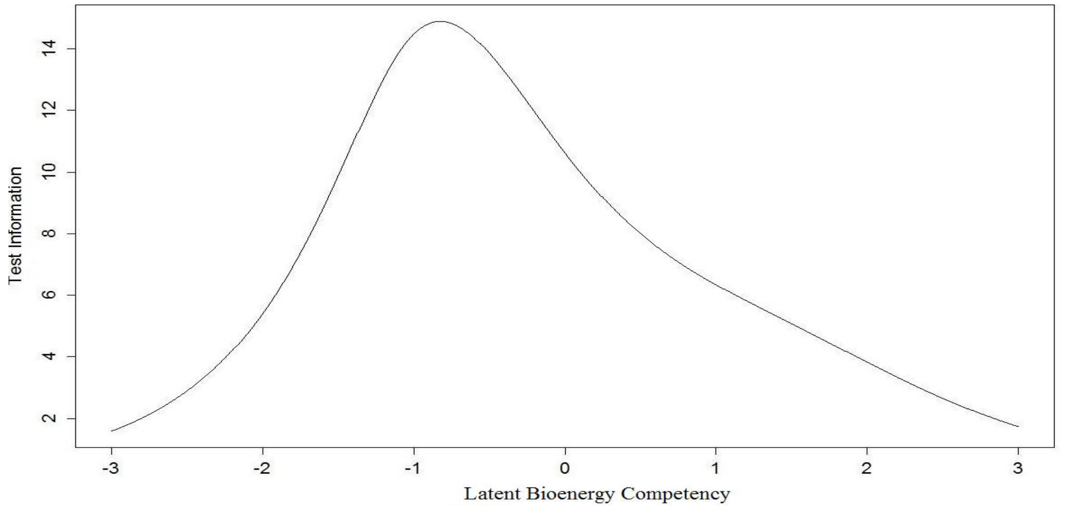

Figure 3. Test information curve.

https://doi.org/10.33225/jbse/19.18.264

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Association with an External Variable

According to *the Standards* (AERA, APA, & NCME, 2014), another type of validity evidence that can be examined is the association between the latent variable and an external variable. Typical external variables include (a) scores on other achievement tests, (b) self-reported scores, and (c) supervisors or teachers' ratings (Wilson, 2005). In this study, the *self-reported bioenergy learning experience questionnaire* was used as an external measure of students' bioenergy competency. Firstly, the Cronbach's alpha of the questionnaire items was 0.85. Then, the observed sum score for each student was calculated to represent the student's self-reported score. Finally, the correlation between the sum scores and the latent bioenergy competency scores was computed, resulting in a value of 0.60, which was positive and moderate. In this context, the correlation coefficient provided evidence supportive of the proposed interpretation; namely, that the latent scores produced from the GPCM measuring high school students' bioenergy competency.

Discussion

This research united the assessment design process and analysis of the psychometric properties on the response data to make sense of the latent variable of interest-students' competency in bioenergy. Multiple evidence was collected to validate the proposed score interpretation—as representing students' standing on bioenergy knowledge. Based on the item calibration results, some items possessed low or relative low discrimination power, which suggested these items should be revised before they were used again. Revision or development of any additional items can be grounded in the bioenergy competency assessment framework (see Figure 1) to ensure that content coverage is maintained. If teachers or curriculum developers are interested in generating more tasks to elicit students' bioenergy competency for use in classrooms, Figure 1 in this study is a possible framework to use.

In addition, this research possessed several design features. Firstly, the assessment possessed the properties that a NGSS-aligned assessment should possess. The NGSS calls for up-to-date assessment frameworks that can measure complex thinking as students carry out scientific processes. This may require, for example, mixed item formats including written responses and detailed scoring rubrics (NRC, 2014), as introduced in this study. Standardized large-scale assessments in science have been criticized as having overly simplified scoring rubrics (Liu, Lee, Hofstetter, & Linn, 2008). In contrast, well-designed, task-specific rubrics like those constructed in this study may be more sensitive to capturing different levels of competency. Moreover, a NGSS-aligned assessment has items that not only elicit students' *factual knowledge*, but also the evidence of *knowledge in use* (Debarger, Penuel, Harris, & Kennedy, 2015). In the present study, for example, the item on *improving efficiency of the biofuel production system,* required students to use knowledge that *energy is lost during conversion* (i.e., factual knowledge) in a specific system regarding *science practices*, which is what professional scientists do as they conduct bioenergy research.

Secondly, documenting design activities is a way to clearly articulate the proposed score interpretation in the assessment development. Merely writing down the definition of the domain or operational variables, is inadequate to establish a coherent assessment. The substantial effort should be made during the design process to tie together (a) the proposed score interpretation, (b) what is valued in classrooms and among professional scientists, and (c) the performance that the items aim to elicit. Moreover, a *specification table*, *claims*, and *item-specific rubrics* in various design stages may also help validate the proposed scores interpretation.

Finally, in analyzing the response data, the GPCM in IRT was used. Although classical test theory (CTT) has the drawback that "examinee characteristics and test characteristics cannot be separated" (Pellegrino, Chudowsky, & Glaser, 2001, p. 123), it is still common to encounter assessments in the science education literature that solely use CTT to report psychometric properties (e.g., see the assessment tool developed by Tiruneh, Elen, De Cock, Weldeslassie, & Janssen, 2016). The standard error of measurement in CTT is for all score values, on average, rather than for an individual score value. In contrast, the GPCM provides a test information index where the students' characteristics and test characteristics can be interpreted jointly, as shown in Figure 3.

https://doi.org/10.33225/jbse/19.18.264

Conclusions

The validation of the proposed score interpretation described here is arguably the first research concerning a bioenergy assessment tool in K-12 science education. Contemporary science has significant potential for contributing to K-12 science education. This research provided a small step forward in finding such potential from ongoing scientific discovery. It will be beneficial to science education if future studies collect more empirical evidence about students' *systems thinking habits* when solving problems about biofuel production systems. Biofuel production system also offer a context where learned facts, associated applications, and relevance are all tied together in students' learning experience. Students may learn, for example, that the biomass (e.g., soybean) used for biofuel competes with food in terms of water use and land use, which pertains to both ecosystem and sustainable development in society. Socioscientific perspectives can help students understand that biofuel production issues are societal issues at local, national, and global levels, thereby converting *learned science* to *applicable knowledge*. The reciprocal relationship is also true; that is, once students possess applicable knowledge, it may increase their motivation to engage with further learning in science.

Considering the end-users of the assessment, this assessment tool was provided for teachers to support their instructional needs. Teachers in general may find it is challenging to balance between teaching content, and teaching students how to convert learned science to applicable knowledge. Thus, gaining input from multiple stakeholders through teacher education programs or teacher PD programs should equip teachers to effectively implement socioscientific perspectives found in contemporary scientific discovery. Once teachers are equipped with a sufficient background in this area, they can more confidently integrate such topics in their regular science teaching. *Systems thinking*, along with *interdisciplinary science* and *greater engagement in the scientific process*, are the main goals in science teaching, described in the NGSS. To accomplish these goals, empirical studies and assessment frameworks can continue to inform each other, enriching science education and strengthening its connections to modern science discovery.

Acknowledgements

The authors would like to thank the reviewers for their helpful comments.

References

Albe, V. (2013). On the road to science education for sustainability? *Cultural Studies of Science Education, 8*(1), 185–192.

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

Baker, F. B. (1992). *Item response theory*. New York: Marcel Dekker.

Barak, J., Gorodetsky, M., & Chipman, D. (1997). Understanding of energy in biology and vitalistic conceptions. *International Journal of Science Education, 19*(1), 21-30.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-459.

Bowman, L., & Govett, A. L. (2015). Becoming the change: A critical evaluation of the changing face of life science, as reflected in the NGSS. *Science Educator, 24*, 51-61.

Bybee, R. W. (2014). The next generation science standards and the life sciences. *Science & Children*, *50*(6), 7-14.

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO for Windows [Computer software]. Lincolnwood, IL: Scientific Software International.

Chabalengula, V. M., Sanders, M., & Mumba, F. (2012). Diagnosing students' understanding of energy and its related concepts in biological context. *International Journal of Science and Mathematics Education, 10*(2), 241-266.

Dahiya, A. (2014). *Bioenergy: Biomass to biofuels*. Saint Louis, MO: Elsevier Science.

de Ayala, R. J. (2009). *The theory and practice of item response theory.* New York: The Guilford Press.

DeBarger, A. H., Penuel, W. R., Harris, C. J., & Kennedy, C. A. (2015). Building an assessment argument to design and use next generation science assessments in efficacy studies of curriculum interventions. *American Journal of Evaluation, 37*(2), 174-192.

DeBoer, G. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching, 37*(6), 582-601.

DiDonato-Barnes, N., Fives, H., & Krause, E. (2013). Using a table of specifications to improve teacher constructed traditional tests: An experimental design. *Assessment in Education: Principles, Policy & Practice, 21*(1), 90-108.

Dunn, J. (2012). *Biofuel life cycle analysis: Issues, results, and case simulations*. Retrieved from https://greet.es.anl.gov/files/ucws-biofuels

https://doi.org/10.33225/jbse/19.18.264

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Great Lakes Bioenergy Research Center (2010). *Biofuels vs. fossil fuels unit.* Retrieved from https://issuu.com/glbioenergy

Halder, P., Pietarinen, S., Havu-Nuutinen, J., & Pelkonen, P. (2010). Young citizens' knowledge and perceptions of bioenergy and future policy implications. *Energy Policy, 38*(6), 3058-3066.

Hammar, T., Stendahl, J., Sundberg, C., Holmström, H., & Hansson, P. (2019). Climate impact and energy efficiency of woody bioenergy systems from a landscape perspective. *Biomass and Bioenergy, 120*, 189-199.

Hayes, A., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77-89.

Jin, H., & Anderson, C. W. (2012). A learning progression for energy in socio-ecological systems. *Journal of Research in Science Teaching, 49*, 1149-1180.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.

Kang, T., Cohen, A. S., & Sung, H. (2009). Model selection indices for polytomous items. *Applied Psychological Measurement, 33*(7). 499-518.

Karlen, D., Birrell, L., Johnson, S., Osborne, J., Schumacher, J., Varvel, M, & Nafziger, J. (2014). Multilocation corn stover harvest effects on crop yields and nutrient removal. *BioEnergy Research, 7*(2), 528-539.

Kline, K. L., Msangi, S., Dale, V. H., Woods, J., Souza, G., Osseweijer, P., Clancy, J., Hilbert, J., Johnson, F., McDonnell, P., & Mugera, H. K. (2017). Reconciling food security and bioenergy: priorities for action. *GCB Bioenergy*, 9(3), 557-576.

Krauskopf, S. (2010). A life-cycle assessment of biofuels. *Science Teacher*, *77*(9), 35-40.

Lee, H., & Liu, O. L. (2009). Assessing learning progression of energy concepts across middle school grades: The knowledge integration perspective. *Science Education*, *94*, 665–688.

Lee, H., Liu, O., Pallant, A., Roohr, K., Pryputniewicz, S., & Buck, Z. (2014). Assessment of uncertainty - infused scientific argumentation. *Journal of Research in Science Teaching, 51*(5), 581-605.

Lee, H., & Ghosh, S. K. (2009). Performance of information criteria for spatial models. *Journal of Statistical Computation and Simulation, 79*(1), 93-106.

Lin, C. Y., & Hu, R. (2003). Students' understanding of energy flow and matter cycling in the context of the food chain, photosynthesis, and respiration. *International Journal of Science Education, 25*(12), 1529-1544.

Linn, R. L., & Gronlund, N. E. (1995). *Measurement and assessment in teaching*. Englewood Cliffs, NJ: Prentice-Hall.

Liu, O. L., Lee, H. S., Hofstetter, C., & Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures and evidence. *Educational Assessment, 13*(1), 33–55.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

Matzenberger, J., Kranzl, L., Tromborg, E., Junginger, M., Daioglou, V., Goh, C., Keramidas, K. (2015). Future perspectives of international bioenergy trade. *Renewable and Sustainable Energy Reviews*, *43*, 926-941.

Metz, S. (2011). *Fuel for thought: Building energy awareness in grades 9-12*. Arlington, VA: National Science Teachers Association.

Mislevy, R. J. (2007). Validity by design. *Educational Researcher, 36*(8), 463–469.

Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement, 4*(17), 351-363.

National Research Council. (1996). *National science education standards.* Washington, DC: The National Academies Press.

National Research Council. (2001). *Knowing What Students Know: the science and design of educational assessment.* Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: The National Academies Press.

National Research Council. (2014). *Developing Assessments for the Next Generation Science Standards.* Committee on Developing Assessments of Science Proficiency in K-12. Board on Testing and Assessment and Board on Science Education, J.W. Pellegrino, M.R. Wilson, J.A. Koenig, and A.S. Beatty, (Eds.), Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2014). *Developing Assessments for the Next Generation Science Standards.* Washington, DC: The National Academies Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: The National Academies Press.

Neumann, K., Viering, T., Boone, W. J., & Fischer, H. E. (2013). Towards a learning progression of energy. *Journal of Research in Science Teaching, 50*, 162–188.

NGSS Lead States (2013). *Next Generation Science Standards: For states, by states.* Washington, DC: The National Academies Press.

Oregon State University Bioenergy K-12 Education. (2016). *Bioenergy lessons*. Retrieved from http://agsci.oregonstate.edu/bioenergy-k-12/bioenergy-lessons.

Parker, J., de Los Santos, E., & Anderson, C. W. (2015). What learning progressions tell us about students' ability to participate in the global climate change and biofuels debates. *American Biology Teacher, 77*, 232.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science of design and educational assessment*. Washington, DC: The National Academies Press.

Reynolds, C. R., Livingston, R. B., & Willson, V. (2006). *Measurement and assessment in education*. Pearson/Allyn & Bacon, Boston: MA.

Tekkumru-Kisa, M., Stein, M. K., & Schunn, C. (2015). A framework of cognitive demand and content practices integration: Task analysis guide in science. *Journal of Research in Science Teaching, 52*(5), 659-685.

https://doi.org/10.33225/jbse/19.18.264

Tiruneh, D. T., Elen, J. G., De Cock, M., Weldeslassie, A., & Janssen, R. (2016). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education, 15*(4), 663-682.

US Department of Agriculture (2018, January 4). *USDA's National Institute of Food and Agriculture Invests in Integrated Biorefinery Optimization*. Retrieved from https://nifa.usda.gov/announcement/usda%E2%80%99s-national-institute-food-and-agriculture-invests-integrated-biorefinery.

US Department of Energy (2017, July 17). *Department of Energy Provides $40 Million for 4 DOE Bioenergy Research Centers*. Retrieved from https://www.energy.gov/articles/department-energy-provides-40-million-4-doe-bioenergy-research-centers.

US Department of Energy (2018). *Welcome to the bioenergy quiz!* Retrieved from https://www.energy.gov/eere/bioenergy/maps/beto-quiz-interactive-content#.

Williams, W. M., Papierno, P. B., Makel, M. C., & Ceci, S. J. (2004). Thinking like a scientist about real world problems: The Cornell Institute for Research on children science education program. *Journal of Applied Developmental Psychology, 25*(1), 107-126.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wu, Y., Zhao, F., Liu, S., Wang, L., Qiu, L., Alexandrov, G., & Jothiprakash, V. (2018). Bioenergy production and environmental impacts. *Geoscience Letter, 5*(14), 1-9.

Received: *December 29, 2018* Accepted: *March 18, 2019*

Tingxuan Li	Master of Science, Doctoral Student, Department of Educational Studies, Purdue University, 100 N University St, West Lafayette, IN, USA. E-mail: li2280@purdue.edu
Kari L. Clase	PhD, Professor, Department of Agricultural & Biological Engineering, Purdue University, 225 South University Street, West Lafayette, IN, USA. E-mail: klclase@purdue.edu
Weiling Li	PhD, Research and Evaluation Specialist, Evaluation and Learning Research Center, Purdue University, 100 N University St, West Lafayette, IN, USA. E-mail: li2288@purdue.edu
Anne Traynor	PhD, Assistant Professor, Department of Educational Studies, Purdue University, 100 N University St, West Lafayette, IN, USA. E-mail: atraynor@purdue.edu

https://doi.org/10.33225/jbse/19.18.264