Journal of Agricultural Education, 65(4), 192-209
https://doi.org/10.5032/jae.v65i4.2791

Development of a Concept Inventory of Inheritance for School-Based Agricultural Education

Dr. Ying Jin [1]
Dr. Chaney W. Mosley [2]

Abstract

This study developed three versions of a concept inventory of inheritance tailored for school-based agricultural education (SBAE). These inventories can be used as formative assessments to inform teachers about students' understanding of inheritance in the context of agriculture or as summative assessments to evaluate learning outcomes. Existing concept inventories in genetics related concepts lack real-world agricultural applications. Our inventories incorporate areas in agriculture, such as animal systems and plant systems, providing students with a deeper understanding of the concept of inheritance. Developed primarily by SBAE teachers, these inventories reflect real-world experiences and observations, offer valuable insights into students' misconceptions, gauge preparation, measure instructional effectiveness, and provide a comprehensive view of learning progressions. We utilized confirmatory factor analysis, Rasch model, and distractor analysis to examine the psychometric properties of the concept inventories.

Introduction

In classroom settings, a concept inventory is frequently used as a diagnostic test to help teachers identify students' misunderstandings regarding a specific instructional concept. Typically, a concept inventory consists of multiple-choice questions with distractors, which are incorrect response options representing common misunderstandings held by students about a particular concept (Cetnar, 2023). When utilized before instruction, student responses, particularly their choices among the distractors, can guide teachers in developing formative assessment activities for instructional purposes (Libarkin, 2008). On the other hand, when used after instruction, student responses on a concept inventory can inform learning outcomes or evaluate the effectiveness of instructional practices (Ward et al., 2016). Concept inventories have primarily been developed and widely utilized in the field of science education, including disciplines such as physics (e.g., the Force Concept Inventory by Halloun & Hestenes, 1985), chemistry (e.g., the Redox Concept Inventory by Brandriet & Bretz, 2014), biology (e.g., the Meiosis Concept Inventory by Kalas et al., 2013), and others.

Currently, the majority of concept inventories have been designed for introductory science courses at the undergraduate level. Concept inventories in various science disciplines can be found in Bowling et al. (2008), Presley et al. (2017), Newman et al. (2016), and Stefanski et al. (2016). However, only a limited number of concept inventories have recently been developed specifically for high school students, such as those focused on biology (Malone et al., 2021) and cosmology (Salimpour et al., 2023). Agricultural science serves as a subject that contextualizes various science disciplines (Myers & Washburn, 2008; National Research Council, 2012), such as biology (animal science), chemistry (plant and soil science), engineering (biosystems engineering), ecology (environmental natural resources), and economics (agriculture business), among others. These content areas are explicitly taught in the eight Agriculture, Food and Natural Resources

[1] Dr. Ying Jin is an Associate Professor in the Department of Learning, Technology, and Leadership Education at James Madison University, Harrisonburg, VA, 22807, jin4yx@jmu.edu.
[2] Dr. Chaney W. Mosley is an Associate Professor in the School of Agriculture at Middle Tennessee State University, Murfreesboro, TN, 37132, chaney.mosley@mtsu.edu

(AFNR) Career Cluster content areas (Swafford, 2018) and play significant roles in Career and Technical Education (CTE) in SBAE pathways.

Because agricultural science is not subject to the same high-stakes standardized tests as other science disciplines, there tends to be a minimal emphasis on formal assessment in agricultural education, yet school-based agricultural educators have stressed the importance of testing/assessment in agricultural programs to hold schools accountable for keeping these programs worthwhile in meeting the objectives of the schools (Phipps et al., 2008). State-level assessments are available to evaluate students' overall performance in CTE programs (including agriculture), such as Arizona's Technical Skills Assessment (Arizona Department of Education, 2023). However, these assessment tools were typically administered by the end of the program and were often used as summative assessments to evaluate program accountability (Pantleo et al., 2022).

The primary goal of this study was to develop a concept inventory on inheritance specifically tailored for the agricultural context. This inventory could serve as an assessment tool aiding in the evaluation of students' understanding of inheritance. Although several concept inventories have been developed concerning related aspects of inheritance (Abraham et al., 2014; Anderson et al., 2002; Bowling et al., 2008; Perez et al., 2013; Price et al., 2014; Ward et al., 2016), few are designed for high school students in the context of agriculture. Student performance on genetics/inheritance related questions on the 2019 National Assessment of Educational Progress (NAEP) ranged from 28% to 55% for the 12th graders. Given genetic/inheritance-related concepts are challenging to high school students, applying the concept of inheritance within the context of agriculture presents even greater challenges because successful application demands more advanced cognitive skills than mere memorization and comprehension of the concept itself (Duncan & Reiser, 2007). Therefore, addressing the challenge and promoting a profound understanding of inheritance and related concepts in agriculture through a validated formative assessment tool is vital for effective teaching and learning in school-based agricultural education.

Specifically, objectives of this research are:

1. Design a concept inventory to evaluate high school students' conceptual knowledge in inheritance in the context of SBAE.
2. Evaluate the psychometric properties of the concept inventory.

Conceptual Framework

The Evidence-Centered Design (ECD) framework (Mislevey et al., 2003) was employed to provide a comprehensive structure for designing the concept inventory. This framework consists of three interrelated models: the Student Model, the Evidence Model, and the Task Model. Each model plays a crucial role in the development of the concept inventory. Specifically, the Student Model outlines the core concepts to measure, such as genetic inheritance patterns and their applications in agriculture. It details the various levels of understanding and proficiency that students might demonstrate, capturing a range of expertise from basic knowledge to more advanced comprehension. The Evidence Model links observable data (e.g., student responses) to the constructs in the Student Model through establishing criteria for scoring and choosing the appropriate measurement model to make inferences about students' proficiency level based on their responses (e.g., Rasch model). The Task Model specifies the design and selection of assessment tasks that will elicit evidence of students' understanding. These tasks are carefully crafted to cover a range of contexts, ensuring that they are aligned with the constructs defined in the Student Model. By employing the ECD framework, we ensure our concept inventory is conceptually grounded and methodologically sound, capable of providing valid and reliable evidence of students' understanding of inheritance in SBAE.

Methodology

Item Development

Expert panel members for test item creation should be able to conceptualize the target group's competency levels (Impara & Plake, 1998; Dalum et al., 2022). The selection of panelists must consider factors such as their experience, qualifications, and ability to contribute meaningfully to the test development process (Peyrony et al., 2020; Flodén et al., 2020). Teachers who understand subject matter can make valuable contributions to an expert panel (Sparks & Lipka, 1991; Wan, 2015). The involvement of classroom teachers on expert panels can provide valuable insights into the practical aspects of test development, such as identifying specific ways to tailor assessments to meet the needs of diverse student populations, and can leverage the teachers' knowledge of classroom dynamics to create more effective assessments (Gerde & Foster, 2014). Further, the inclusion of teachers on expert panels can help ensure that the test content aligns with educational standards and reflects the realities of classroom instruction (Gerde et al., 2015). Therefore, an expert panel consisting of four agricultural teacher education faculty (i.e., researchers) and seven SBAE teachers with extensive teaching experience was formed to collectively develop the Concept Inventory of Inheritance (CII). Shaw et al. (2008) identified seven topics of misconceptions high school students have regarding genetics. For each identified topic, key ideas were used to categorize misconceptions. Table 1 shows the five key ideas associated with the topic *patterns of inheritance* (Shaw et al., 2008). The panel first reviewed the key ideas associated with the topic of patterns of inheritance and discussed if those ideas were consistent with observed misconceptions of secondary students enrolled in agricultural education courses, achieving 100% consensus. Next, a crosswalk was conducted to identify overlap between the key ideas and the AFNR Career Cluster Content Standards (National Council for Agricultural Education [NCAE], 2015) to verify the national appropriateness of the CII for SBAE. The AFNR Career Cluster Content Standards (National Council for Agricultural Education, 2015) are organized by eight career pathways, with standards in each pathway organized by Common Career Technical Core (CCTC) Standards (National Association of State Directors of Career and Technical Education, 2012), performance indicators, and sample measurements. The expert panel identified two AFNR career pathways as having standards that explicitly relate with the key ideas indicated by Shaw et al. (2008). Those pathways are Animal Systems (AS) and Biotechnology Systems (BS). The expert panel determined the Plant Systems (PS) career pathway did not have standards that explicitly relate with the key ideas but noted “topics represented by each strand are not all-encompassing” (NCAE, 2015, p. 108). Based on their years of classroom observations, the expert panel recommended a concept inventory related patterns of inheritance should include questions related to AS, BS, and PS career pathways. Finally, the expert panel developed items based on three of the cognitive demands for achieving performance expectations of the 2019 Science Framework of National Assessment of Educational Program (National Assessment Governing Board [NAGB], 2019). Table 1 also presents the item specification table to illustrate the item distribution across the identified themes and the cognitive domains.

Table 1

Item Specification Based on Key Ideas and Cognitive Domains

	Cognitive domains		
Key Ideas	*Knowing that*	*Knowing how*	*Knowing why*
Probability	3	3	3
Types of inheritance	3	3	3
Misunderstanding modes of inheritance	3	3	3
Interpretation of data from Punnett Squares	3	3	3
Understanding the origin of Chromosomal Anomalies	3	3	3

Note. The number 3 in each cell represents the number of items per cognitive domain and theme.

For items demanding the cognitive domain of *knowing that*, items were developed to understand students' performance expectation on factual statements regarding inheritance in the agricultural context (NAGB, 2019). An example item is, "In Shorthorn cattle, the roan color is an example of what type of inheritance?" This item was designed to examine students' comprehension of genetic concepts such as dominance, recessiveness, and how they manifest in real-world examples. In addition to the correct option, three distractors were created based on students' common misunderstandings of inheritance patterns.

The *knowing how* items were developed to understand students' ability to apply knowledge and concepts to solve problems, conduct investigations, and analyze scientific information (NAGB, 2019). An example item is, "In pea plants, purple flowers (A) are dominant over white flowers (a). Which answer correctly sets up a Punnett square that crosses a homozygous dominant plant with a heterozygous plant?" To answer this question correctly, students need to recall the principles of inheritance, apply their knowledge to set up a Punnett square, and analyze the possible genotypes and phenotypes of the offspring, demonstrating their ability to interpret genetic crosses and predict inheritance patterns.

Items demanding the *knowing why* cognitive domain go beyond evaluating the factual knowledge and application but emphasize deeper understanding and critical thinking about scientific concepts, enabling students to explain the reasons behind phenomena and draw connections between different pieces of scientific knowledge (NAGB, 2019). An example item is, "Scrapie is a fatal disease that affects the nervous systems of small ruminant species. In genetic testing for Scrapie (Codon 171), animals identified as RR are resistant to the disease, QR animals are carriers, and QQ represents susceptibility to the disease. A farmer is looking at purchasing a ewe lamb to add to the flock. She has been identified as QR for Codon 171. The farmer's primary focus is to eradicate Scrapie in the operation. The farmer is currently utilizing a ram that is RR. Only considering the probability of Scrapie susceptibility, should the farmer purchase the ewe lamb? Why/why not?" This item was designed to indicate if students to grasp the basic principles of inheritance, apply their knowledge of genetic inheritance patterns to assess the risk of Scrapie transmission in the farmer's flock, weigh the risk of disease transmission, and make an informed decision based on scientific reasoning.

Data Collection

To mitigate the impact of test fatigue on item responses (Ackerman & Kanfer, 2009), three versions (A, B, C) of CII were created by randomly choosing one item from each of the five key ideas associated with the topic patterns of inheritance. Each version contained 15 items, with each item representing unique combinations of key ideas and cognitive domains. Nine SBAE teachers were recruited to administer CIIs in their Agriscience classes. In the state where data were collected, Agriscience is the introductory level course for all AFNR programs of study and has course standards dealing with the concept of patterns of inheritance. It is important to note the CII is not a standards-based assessment, rather, it is a concept-based assessment.

Three teachers were randomly assigned to each of the three CII versions. Students took CIIs after receiving instruction on inheritance. Among the enrolled students, 72 students took version A, 105 students took version B, and 129 students took version C.

Validity and Reliability

In this study, the validity of CII scores was assessed by examining their construct validity. This aspect of validity ensures scores accurately reflect the underlying concept intended to be measured by the instrument. Additionally, exploring construct validity assists in assessing the instrument's dimensionality, which is crucial for conducting Rasch analysis under the assumption of unidimensionality. To evaluate construct validity, one-factor confirmatory factor analysis (CFA) was conducted using the lavaan package in R (Rosseel, 2012), focusing on the factor of *inheritance*. The establishment of construct validity of CII scores was achieved by examining the model fit statistics and indices associated with the one-factor CFA model, including the Chi-square test ($\chi 2$), $\chi 2/df$ (with values below 3 considered acceptable), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values above 0.90, and Root Mean Square Error of Approximation (RMSEA) values below 0.06 (Hu & Bentler, 1999). These metrics served as benchmarks for evaluating the construct validity of CIIs.

Establishing reliability for concept inventories has posed a challenge, particularly when relying on indicators of internal reliability such as Cronbach's alpha (Taber, 2017). This difficulty stems from the nature of the questions themselves, which are not necessarily intended to be interrelated. In concept inventories, questions often blend various themes of the main concept, complicating the assessment of reliability. Therefore, it is not uncommon to observe reliability of responses from other concept inventories to be lower than a widely accepted threshold (e.g., 0.7, McNeish, 2018). For example, responses on the same Redox Concept Inventory from two different studies (Brandriet & Bretz, 2014; Jin et al., 2020) yielded Cronbach's alpha values ranged from 0.57 to 0.69. Similar Cronbach's alpha value is anticipated in the current study.

Rasch Analysis

To further understand the performance of CIIs, psychometric properties of CIIs were examined using the Rasch model (Rasch, 1960). The Rasch model utilizes a logistic probability model to estimate the probability of a correct response based on person ability parameters (θ) and item difficulty parameters (b), as illustrated in the equation below:

$$P(y_{ij} = 1 \mid \theta_j) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \tag{1}$$

where $P(y_{ij} = 1 \mid \theta_j)$ represents the probability of a correct response ($y_{ij} = 1$) of item i for person j given this person's ability parameter θ_j (i.e., to what extent a person understands the concept of inheritance), b_i is the item difficulty parameter for item i. Since both θ_j and b_i are on the same standardized scale [i.e., $\theta_j \sim N(0,1)$; $b_i \sim N(0,1)$], individuals are more likely to provide correct responses when their $\theta_j s$ exceed b_i. In addition, given that θ_j and b_i are on the same scale, the alignment of the distribution of $\theta_j s$ and $b_i s$ can be assessed to gauge the appropriateness of the instrument to its target population. For example, if CII is administered in a regular secondary agriculture class to evaluate students' overall understanding of inheritance, the distribution of $\theta_j s$ should ideally follow a normal distribution, with an approximate normal distribution of $b_i s$ aligned closely. This alignment ensures one or more items can effectively measure students' abilities (i.e., understanding of inheritance) in a distinct range of the θ distribution. The alignment of distributions was examined through Wright person−item maps, which are presented in the results section.

The mirt package in R (Chalmers, 2012) was used to conduct the Rasch analysis. Distributions of item and person parameter estimates were presented in the Wright person−item maps. Rasch model-data fit indices were also reported to evaluate to what extent the observed responses align with the expectations of the model. These fit indices inform the quality of individual items and the overall performance of the measurement instrument. Specifically, INFIT and OUTFIT statistics were reported as they can identify unusual response patterns that deviate from the model's expectations. INFIT and OUTFIT statistics are sensitive to less-extreme unexpected responses and extreme unexpected responses, respectively (Wind & Hua, 2021). Their expected values are 1 with an acceptable range between 0.7 to 1.3 (Bond, 2007). Items with acceptable INFIT and OUTFIT statistics suggest that these items are functioning as expected by the Rasch model and are not overly influenced by outliers or extreme responses.

Distractor Analysis

Distractor analyses were illustrated using trace plots, wherein the x-axis depicted the person's ability estimates derived from the Rasch analysis, and the y-axis represented the probability of each response option across different levels of person ability estimates. These plots were generated separately for each item. An ideal trace plot would exhibit the curve for the correct response option surpassing those of all other options as the person ability estimates increase. These trace plots were created using the ShinyItemAnalysis package in R (Martinková & Drabinová, 2019).

Results

Validity and Reliability

Model fit statistics and indices of the one-factor CFA models presented in Table 2 provided evidence of construct validity of all three versions of CIIs (i.e., all model fit statistics and indices exceeded the recommended thresholds). In factor analyses, factor loadings (λ) measure the degree of association between an item and the underlying construct/factor (i.e., the concept of inheritance). Items with bolded factor loadings presented in Table 3 are significantly correlated with the concept of inheritance, indicating the variance in item responses can be appropriately accounted for by the concept. Conversely, most items with low factor loadings (i.e., the unbolded ones) are items related to the probability key idea, which requires additional knowledge (i.e., math/statistics) beyond the concept of inheritance. The Cronbach's alpha values for three versions of CIIs were 0.62, 0.66, and 0.65, respectively.

Table 2

CFA and Rasch Analysis for Each Version of CII.

Items	CII-A				CII-B				CII-C			
	λ	*b*	INFIT	OUTFIT	λ	*b*	INFIT	OUTFIT	λ	*b*	INFIT	OUTFIT
1	-0.22	-0.68	1.02	1.03	0.15	0.79	1.04	1.02	**0.49**	-0.18	0.91	0.90
2	0.19	1.06	0.99	0.96	0.19	0.53	1.00	1.03	**0.41**	-0.38	0.92	0.87
3	-0.12	0.29	1.05	1.06	**0.57**	-0.54	0.88	0.85	**0.40**	0.49	0.90	0.90
4	**0.81**	0.14	0.87	0.86	0.22	-0.26	1.00	0.99	**0.46**	-0.81	0.94	0.88
5	**0.38**	-0.21	0.89	0.89	**0.56**	-0.49	0.88	0.87	**0.31**	-0.90	0.94	0.91
6	**0.55**	-0.61	0.86	0.85	**0.39**	0.89	0.97	0.88	-0.23	0.35	1.13	1.13
7	0.23	0.60	0.96	0.95	**0.42**	0.68	0.94	0.91	**0.64**	0.49	0.84	0.82
8	**0.42**	-0.75	0.93	0.93	**0.31**	0.11	0.92	0.90	**0.60**	0.76	0.87	0.87
9	**0.59**	-0.21	0.87	0.86	**0.39**	0.11	0.92	0.90	**0.41**	0.31	0.89	0.87
10	**0.79**	-1.41	0.86	0.77	**0.58**	-0.59	0.86	0.83	**0.36**	-0.34	0.94	0.93
11	-0.16	0.68	1.06	1.10	**0.64**	-1.20	0.89	0.77	**0.55**	-1.06	0.92	0.80
12	**0.61**	0.00	0.87	0.86	**0.81**	-0.03	0.81	0.79	**0.82**	0.09	0.80	0.78
13	**0.62**	-0.55	0.93	0.93	**0.81**	-1.68	0.84	0.66	**0.38**	0.91	0.95	0.97
14	-0.04	0.60	1.03	1.01	**0.41**	0.63	0.92	0.92	**0.31**	-0.42	0.95	0.90
15	0.35	1.06	0.97	0.89	0.23	1.06	1.02	0.99	0.15	0.68	1.00	1.00
Model fit statistics to evaluate construct validity	$\chi^2(90) = 91.17$, $p=0.45$ $\chi2/df = 1.01$ CFI = 0.99 TLI = 0.99 RMSEA = 0.01				$\chi^2(90) = 74.19$, $p=0.89$ $\chi2/df = 0.82$ CFI = 0.98 TLI = 0.98 RMSEA = 0.02				$\chi^2(90) = 85.54$, $p=0.89$ $\chi2/df = 0.95$ CFI = 0.93 TLI = 0.91 RMSEA = 0.03			
Reliability	0.62				0.66				0.65			

Note. λ: factor loading; b: item difficulty parameter; reliability: Cronbach's alpha.

Rasch Analysis

INFIT and OUTFIT statistics presented in Table 2 were within the acceptable range for all items across the three versions, indicating acceptable model data fit for the Rasch model. The estimated item difficulty parameters (b) presented in Table 2 ranged from -1.41 to 1.06, -1.68 to 1.06, and -1.06 to 0.91, respectively for CII-A, CII-B, and CII-C. For each person-item map, the histogram on the top of the map represents the distribution of person parameter estimates. An examination of the three person-item maps indicated the item difficulty parameter distributions were mostly well aligned with person parameter estimates between -1 to 1, with few items available to evaluate high-achievers (e.g., $\theta > 1$) and low-achievers (e.g., $\theta < -1$) in understanding the concept of inheritance.

Figure 1

Wright Person-Item Map for CII-A

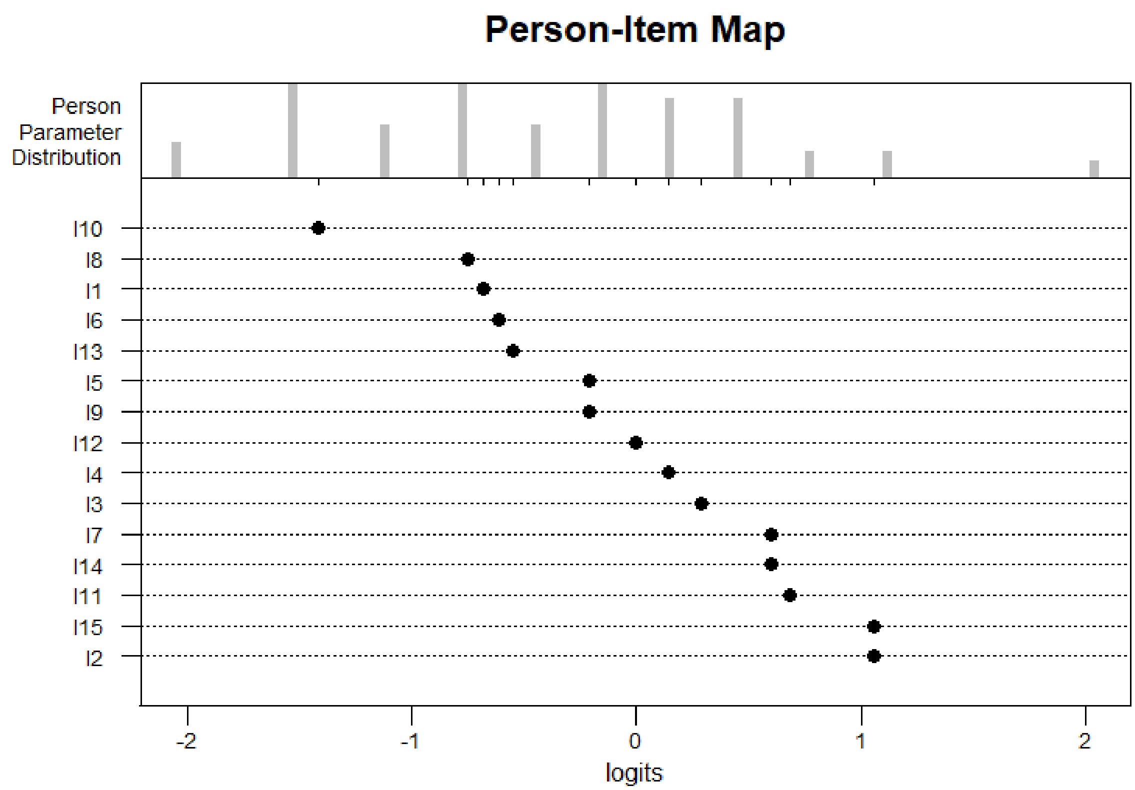


Figure 2

Wright Person-Item Map for CII-B

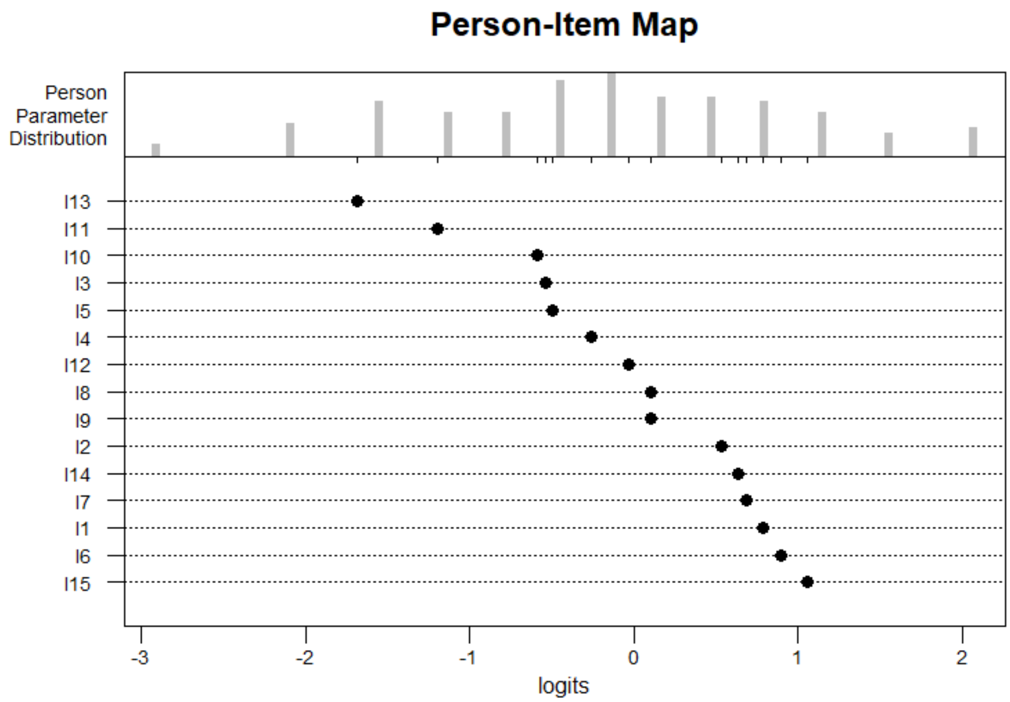
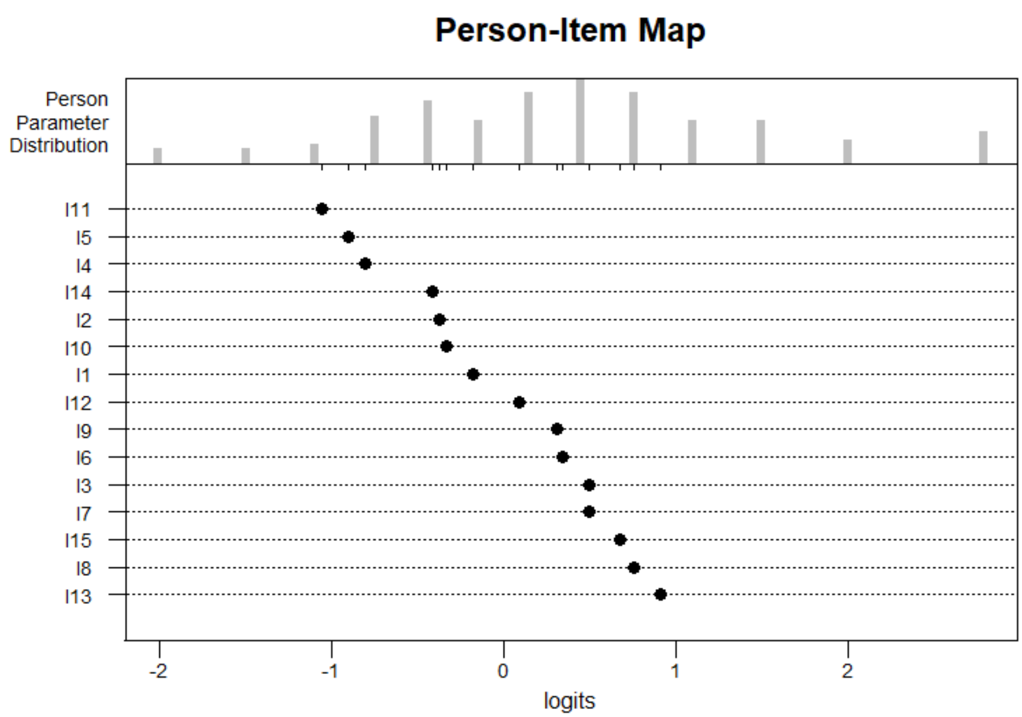


Figure 3

Wright Person-Item Map for CII-C



Distractor Analysis

Figures 4 and 5 show the distractor plots for item 4 in CII-A with a factor loading of 0.81 and for item 14 in CII-A with a factor loading of -0.04. These two items have the highest and lowest factor loadings in CII-A, respectively. The distractor plot of item 4 illustrates a trend in which the proportion of correct response (red line) started to surpass the proportions of all the other options as person ability estimates increase, suggesting that high achievers were more likely to answer this item correctly. Notably, items across all three versions of CII with significant factor loadings displayed a similar pattern in their distractor plots. Whereas for item 14, a distractor, rather than the correct response, had its proportion of responses surpassing the proportions of all the other options across the most range of continuum of the person parameter estimates. Figure 6 presents the item. Items with nonsignificant factor loadings tended to present a similar pattern as item 14 in CII-A. Distractor plots for all other items in CIIs can be requested from the authors. Interpretations of these distractor plots were detailed in the discussion section.

Figure 4

Distractor Plot for Item 4 in CII-A. 1=A, 2=B, 3=C, and 4=D.

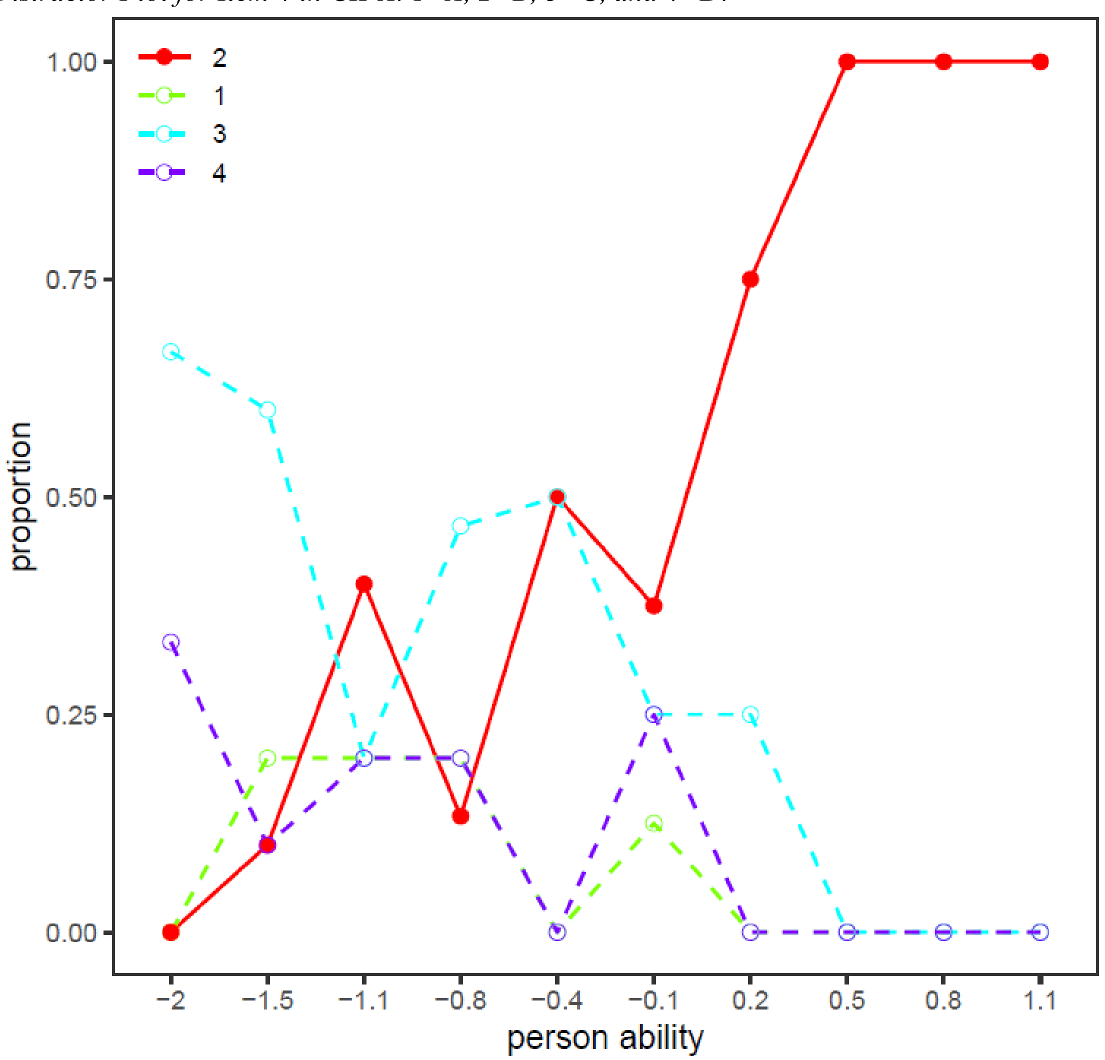


Figure 5
Distractor Plot for Item 14 in CII-A. 1=A, 2=B, 3=C, and 4=D.

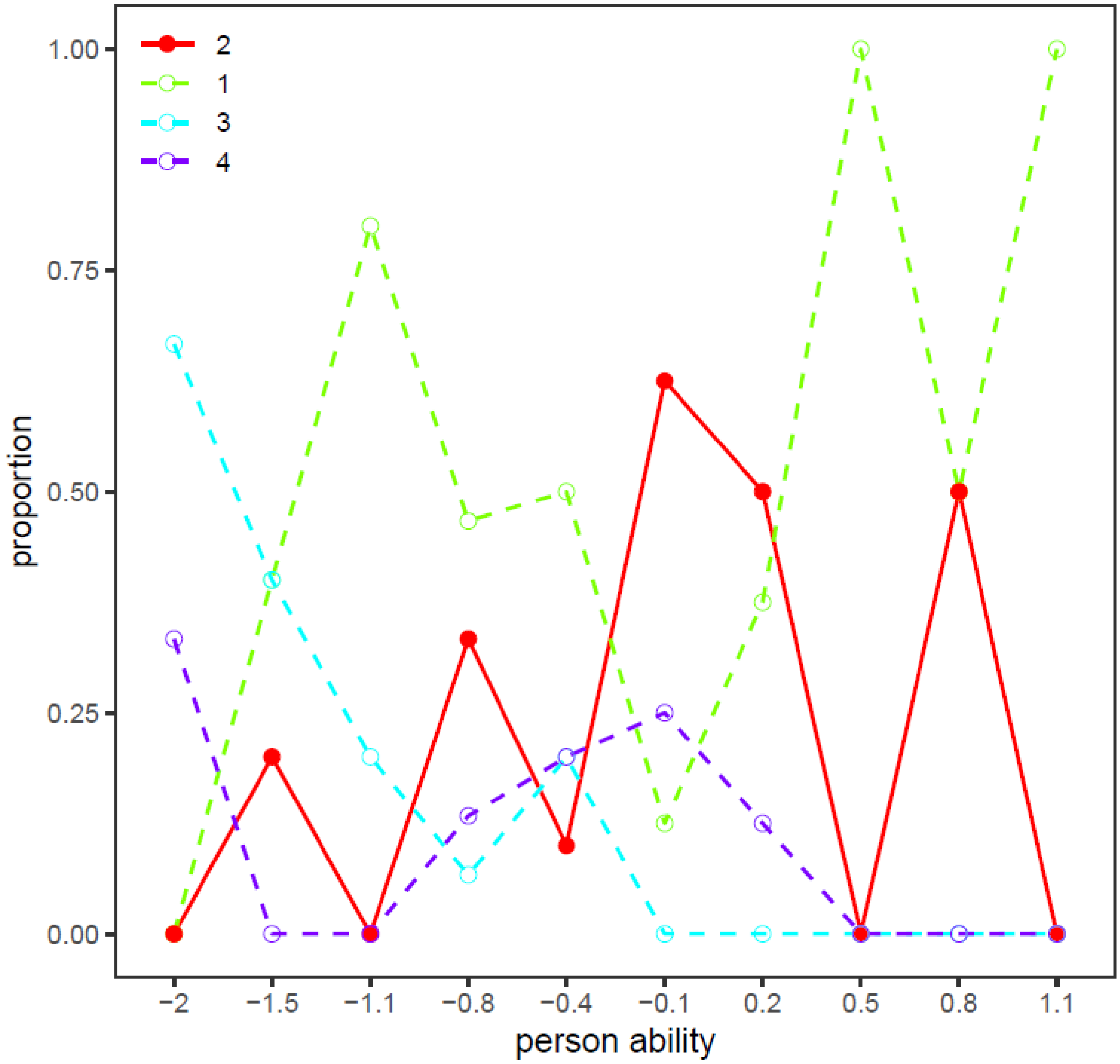


Figure 6

Item 14 in CII-A with a Factor Loading of -0.04.

Chromosomal abnormalities can cause serious problems in animal production. Which of the following best explains the chromosomal anomaly shown below?

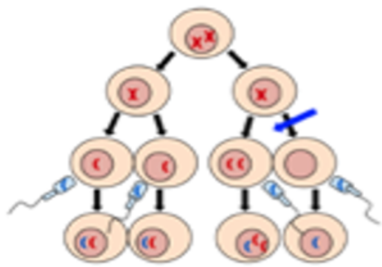


Image obtained from https://en.wikipedia.org/wiki/Nondisjunction

A. Failure of the sister chromatids to separate correctly
B. Failure of the homologues to separate correctly (correct)
C. Fertilization with non-ideal sperm
D. Production of a cell with a trisomy

Conclusions

This study developed three versions of a concept inventory on inheritance. Psychometric properties were examined by factor analysis, Cronbach's alpha, and Rasch analysis. Most items in CIIs demonstrated acceptable psychometric properties. They can be used as formative assessments to inform teachers about students' understanding of the concept of inheritance in the context of agriculture, or as summative assessments to evaluate students' learning outcomes. These concept inventories were developed specifically for SBAE and were primarily created by SBAE teachers who interact with their students daily, demonstrating their capability to develop high quality assessments.

Existing concept inventories in genetics evaluate students' understanding of genetics either without a context or primarily within the context of human bodies, such as the Genetics Concept Inventory (Elrod, 2007), the Genetics Concept Assessment (Smith et al., 2008), and the Meiosis Concept Inventory (Kalas et al., 2017). Although these assessments provide valuable insights into students' understanding of genetics, they often fail to connect the concepts with real-world applications, especially for specialized fields like agriculture. In contrast, the CIIs developed from this study offer a unique approach by incorporating various areas in the context of agriculture, including animal systems, biotechnology systems, and plant systems. By grounding the inheritance concept in agricultural contexts, these assessments provide students with a deeper understanding of how genetics directly impacts agricultural practices (e.g., livestock breeding and crop management). Students may feel more connected to the concept when being evaluated within a relevant context.

Another notable difference between the CIIs and other genetics concept inventories lies in their development process. The previously mentioned genetics concept inventories were all designed for undergraduate students in higher education programs, and their development involved conducting interviews with students to identify misconceptions and provide additional validity evidence (Bowling et al. 2008; Newman et al., 2016; Presley et al., 2017; Stefanski et al., 2016). In contrast, our development process could not incorporate student interviews due to IRB and school policy restrictions. We primarily relied on input from in-field teachers regarding students' misconceptions about inheritance, which were informed by Shaw et al. (2008). Given that teachers are direct observers of their students' learning processes, their insights were valuable in shaping the content and structure of the CIIs. Our approach ensured the CIIs were grounded in the real-world experiences and observations of educators who interact with students on a daily basis. This collaboration between researchers and teachers not only enriched the development process but also enhanced the relevance and applicability of the assessments in SBAE.

Discussion

Although most items in the CIIs exhibited acceptable psychometric properties, some items needed further attention due to their performance. For example, the extremely low factor loading for item 14 (Figure 6) could be attributed to the abnormal pattern observed in its distractor plot (Figure 5), which could be explained by several factors. First, instructors administering this item might not delve into the associated theme as comprehensively as they did with others. Second, despite instruction on this key idea, students may have retained misconceptions. Third, the wording of the question itself could have contributed to confusion among respondents. In item 14, the question presents an image that students must interpret to answer the question. The distractor response (A), in which the proportion of responses surpassed the proportions of all the other options, and the correct response (B), are closely worded, with the phrase "sister chromatids" in response A being replaced with "homologues" in response B. This subtle difference likely caused confusion among students, as they may not have fully grasped the distinction between chromatids and homologues. Consequently, they might have opted for the distractor that contained the more familiar term, "sister chromatids," over the correct term, "homologues," suggesting that the students' struggles could be rooted in their lack of understanding of these key concepts, leading them to choose the option that looked most familiar to them.

Items with low factor loadings also impacted reliability measured by Cronbach's alpha, which evaluates internal reliability. Highly correlated items, manifested by homogenously high factor loadings across all items, tend to yield high Cronbach's alpha (McNeish, 2018), and vice versa. Therefore, the Cronbach's alpha values observed for three versions of CIIs, ranging from 0.62 to 0.66, could be largely influenced by those probability items with low factor loadings. However, similar values were reported in previous concept inventory studies (Brandriet & Bretz, 2014; Jin et al., 2020).

For the Rasch analysis, a value of 1 for INFIT and OUTFIT statistics indicates a perfect model fit. Values less than 1 suggest that the model fits the data better than expected, which could imply that the items are too easy or that the model may capture both the underlying patterns and the random noise. This is an indication of overfitting, potentially leading to inconsistent performance with new datasets. Some items in CIIs showed signs of overfitting but remained above the lower bound of the acceptable range for these fit statistics (Bond, 2007).

It is crucial to highlight that items across all three versions of the CIIs predominantly fell within the middle range of difficulty levels. This suggested that teachers did not develop items that were either too easy or too challenging to assess low- and high-achieving students. This could be explained by the fact SBAE is an elective subject, unlike courses in mathematics, science, and English language arts, which are often assessed with high-stakes standardized tests. Consequently, teachers may tend to be lenient in their assessments, focusing on evaluating students' overall understanding of the concept across the general student population rather than differentiating between students at various achievement levels. Additionally, concept inventories, when used as formative assessments, are typically considered low-stakes assessments in regular classroom settings and are not employed as high-stakes tests to differentiate between high- and low-achievers (Haudek et al., 2011; Steif, 2005).

Despite some items requiring further investigation based on new datasets, the broader impact of this study can be viewed from two main perspectives. Firstly, the development of a concept inventory focused on a scientific concept within the agricultural context serves as a model for promoting formal assessment in SBAE. Future educators can use this research as a basis to develop concept inventories for other challenging scientific topics. By doing so, they can enhance the effectiveness of teaching these concepts to students. Secondly, previous attempts to demonstrate the impact of agricultural education on students' academic achievement in core subjects have faced limitations due to the absence of formal assessment methods. For example, Nolin and Parr (2013) used the number of agricultural classes taken by students as a proxy measure for agricultural education to assess its influence on graduation exam scores. Similarly, Balschweid (2002) examined the impact of incorporating an agricultural context into biology education by collecting students' self-reported survey data on learning outcomes. The development of a comprehensive set of concept inventories specific to the agricultural context provides a direct measure of students' core content knowledge in agriculture, and also introduces an additional dimension or indicator for evaluating the value of agricultural education. These inventories can be used in correlation studies with variables of interest, such as students' overall academic achievement. As a result, researchers and educators gain a more robust toolset to assess and understand the effects of agricultural education on students' learning outcomes in a broader academic context.

Implication

Concept inventories, whether used as formative assessments or summative assessments, can inform teaching practices, as evidenced by the distractor analyses. For SBAE, when employed as formative assessments, the CIIs provide instructors valuable insights into students' understanding of key concepts regarding inheritance. For example, the distractor plot of item 14 (Figure 5) revealed significant confusion between chromatids and homologues, which led to consistently low performance on this item across all proficiency levels of students. By utilizing concept inventories as formative assessments, SBAE teachers can adjust their teaching to address students' misconceptions or confusion effectively and measure the effectiveness of their instruction (Smith & Marbach-Ad, 2010). On the other hand, when the CIIs are utilized as summative assessments, they offer a means to evaluate the overall learning outcomes and achievements of students at the end of an instructional unit (DiVall et al., 2014), if the unit learning objectives align with items on the CII. Summative assessments using the CIIs could provide a comprehensive view of students' conceptual understanding and learning progressions over a specific period. For example, if SBAE teachers administer CIIs before and after their instruction on inheritance, the change in scores and distractor plots could reveal learning gains and areas needing further reinforcement. The CIIs could also help inform educators about the effectiveness of their teaching methods, with regard to inheritance. The CIIs in both formative and summative contexts can enable instructors to make data-informed decisions, improve teaching practices, and enhance student learning experiences. By integrating the CIIs into classroom assessment strategies, SBAE teachers can gain a holistic understanding of student learning, address misconceptions effectively, and continuously improve their teaching methodologies to optimize student outcomes.

Limitations/Future Study

Several factors could influence student responses on CIIs, affecting the analysis of the psychometric properties of the items. First, the information on what was taught about inheritance in SBAE was not collected from the participating teachers. This lack of context could result in nonrepresentative responses to certain items, which in turn may affect the interpretation of the distractor plots and the conclusions drawn to inform teaching practices. Second, due to the limited number of participating teachers and enrolled students per class, the sample size did not meet general guidelines (e.g., a 10:1 respondent-to-item ratio for factor analysis [Nunnally, 1978]). As a result, the insufficient sample size may lead to inconsistent estimates of factor loadings (MacCallum et al., 2001) due to sampling inadequacy, as indicated by KMO values of 0.52, 0.58, and 0.61 for each version of the CII, which could complicate efforts to compare the estimated factor loadings with established guidelines, such as the threshold of 0.7 or higher for strong factor loadings (Kline, 2016). This study performed the analyses to evaluate psychometric properties of all items in CIIs. No items were revised or removed based on current findings. For test security reasons, CII items are available upon request. We invite researchers and teachers to consider integrating CIIs into their research or teaching endeavors. Moving forward, we remain committed to refining the items based on feedback and data received from the SBAE community, thereby strengthening the evidence supporting their validity and reliability.

References

Abraham, J. K., Perez, K. E., & Price, R. M. (2014). The Dominance Concept Inventory: A tool for assessing undergraduate student alternative conceptions about dominance in Mendelian and population genetics. *CBE—Life Sciences Education*, *13*(2), 349–358. https://doi.org/10.1187/cbe.13-08-0160

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, *15*(2), 163. https://doi.org/10.1037/a0015719

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and Evaluation of the Conceptual Inventory of Natural Selection. *Journal of Research in Science Teaching*, *39*(10), 952–978. https://doi.org/10.1002/tea.10053

Arizona Department of Education. (2023). *CTE technical skills assessment overview guide*. https://www.azed.gov/sites/default/files/2023/10/CTE-Technical-Skills-Assessment-Overview-Guide.pdf

Balschweid, M. A. (2002). Teaching biology using agriculture as the context: Perceptions of high school students. *Journal of Agricultural Education*, *43*(2), 56–67. https://doi.org/10.5032/jae.2002.02056

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.) Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., Moskalik, C. L., & Huether, C. A. (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics*, *178*(1), 15–22. https://doi.org/10.1534/genetics.107.079533

Brandriet, A. R., & Bretz, S. L. (2014). The development of the Redox Concept Inventory as a measure of students' symbolic and particulate redox understandings and confidence. *Journal of Chemical Education*, *91*(8), 1132–1144. https://doi.org/10.1021/ed500051n

Cetnar, A. (2023). Introduction to concept inventories for medical physics education. *Journal of Applied Clinical Medical Physics*, *24*(10). https://doi.org/10.1002/acm2.14130

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Dalum, J., Christidis, N., Myrberg, I., Karlgren, K., Leanderson, C., & Sandborgh-Englund, G. (2022). Are we passing the acceptable? Standard setting of theoretical proficiency tests for foreign-trained dentists. *European Journal of Dental Education, 27*(3), 640-649. https://doi.org/10.1111/eje.12851.

DiVall, M. V., Alston, G. L., Bird, E., Buring, S. M., Kelley, K. A., Murphy, N. L., Schlesselman, L. S., Stowe, C. D., & Szilagyi, J. E. (2014). A Faculty Toolkit for Formative Assessment in Pharmacy Education. *American Journal of Pharmaceutical Education*, *78*(9), 160. https://doi.org/10.5688/ajpe789160

Duncan, R. G., & Reiser, B. J. (2007). Reasoning across ontologically distinct levels: Students' understandings of molecular genetics. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *44*(7), 938–959. https://doi.org/10.1002/tea.20186

Elrod, S. (2007). *Genetics concepts inventory*. Biological Sciences Department, California Polytechnic State University.

Flodén, A., Stadtler, M., Collazo, S., Mone, T., Ash, R., & Fridlund, B. (2020). Cross-cultural adaptation and psychometric validation of the Flodén atodai instrument in the North American context. *BMC Nursing, 19*(1). https://doi.org/10.1186/s12912-020-00444-8

Gerde, H. and Foster, T. (2014). Beyond the pencil: Expanding the occupational therapists' role in helping young children to develop writing skills. *The Open Journal of Occupational Therapy, 2*(1). https://doi.org/10.15453/2168-6408.1070

Gerde, H., Bingham, G., & Pendergast, M. (2015). Reliability and validity of the writing resources and interactions in teaching environments (write) for preschool classrooms. *Early Childhood Research Quarterly, 31*, 34-46. https://doi.org/10.1016/j.ecresq.2014.12.008

Haudek, K. C., Kaplan, J. J., Knight, J., Long, T., Merrill, J., Munn, A., Nehm, R., Smith, M., & Urban-Lurain, M. (2011). Harnessing technology to improve formative assessment of student conceptions in STEM: forging a national network. *CBE—Life Sciences Education*, *10*(2), 149–155. https://doi.org/10.1187/cbe.11-03-0019

Halloun, I. A. & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, *53*(11), 1043–1055. http://dx.doi.org/10.1119/1.14030

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. https://doi.org/10.1080/10705519909540118

Impara, J. and Plake, B. (1998). Teachers' ability to estimate item difficulty: A test of the assumptions in the angoff standard setting method. *Journal of Educational Measurement, 35*(1), 69-81. https://doi.org/10.1111/j.1745-3984.1998.tb00528.x

Jin, Y., Rodriguez, C. A., Shah, L., & Rushton, G. T. (2020). Examining the psychometric properties of the redox concept inventory: A Rasch approach. *Journal of Chemical Education*, *97*(12), 4235–4244. https://doi.org/10.1021/acs.jchemed.0c00479

Kalas, P., O'Neill, A., Pollock, C., & Birol, G. (2013). Development of a meiosis concept inventory. *CBE—Life Sciences Education*, *12*(4), 655–664. https://doi.org/10.1187/cbe.12-10-0174.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Libarkin, J. C. (2008). *Concept inventories in higher education science*. Manuscript prepared for the National Research Council Promising Practices in Undergraduate STEM Education Workshop 2, held October 13–14, 2008, in Washington, DC.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Sehee, H. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36(4), 611–637. https://doi.org/10.1207/s15327906mbr3604_06

Malone, K. L., Boone, W. J., Stammen, A., Schuchardt, A., Ding, L., & Sabree, Z. (2021). Construction and evaluation of an instrument to measure high school students biological content knowledge. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(12), em2048. https://doi.org/10.29333/ejmste/11376

Martinková, P., & Drabinová, A. (2019). ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational test. *R Journal*, *10*(2), 503–515. http://dx.doi.org/10.32614/RJ-2018-074

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, *2003*(1), i-29.

Myers, B. E., & Washburn, S. G. (2008). Integrating science in the agriculture curriculum: agriculture teacher perceptions of the opportunities, barriers, and impact on student enrollment. *Journal of Agricultural Education*, *49*(2), 27–37. http://dx.doi.org/10.5032/jae.2008.02027

National Assessment Governing Board (2019). *Science framework for the 2019 National Assessment of Educational Progress*. https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/science/2019-science-framework.pdf

National Association of State Directors of Career and Technical Education. (2012). *The common career technical core, programs of study & industry-based standards*. https://careertech.org/wp-content/uploads/2023/01/CCTC-IndustryStandards.pdf

National Research Council (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: National Academies Press.

Nolin, J. B., & Parr, B. (2013). Utilization of a High Stakes High School Graduation Exam to Assess the Impact of Agricultural Education: A Measure of Curriculum Integration. *Journal of Agricultural Education*, *54*(3), 41–53. https://doi.org/10.5032/jae.2013.03041

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Pantleo, M., Conrad, M., Parr, K., & Parr, B. (2022). Technical skill assessments as a predictor of agriculture students' success after high school. *Journal of Research in Technical Careers*, *6*(1), 30–46. https://doi.org/10.9741/2578-2118.1111

Perez, K. E., Hiatt, A., Davis, G. K., Trujillo, C., French, D. P., Terry, M., & Price, R. M. (2013). The EvoDevoCI: a concept inventory for gauging students' understanding of evolutionary developmental biology. *CBE—Life Sciences Education*, *12*(4), 665–675. https://doi.org/10.1187/cbe.13-04-0079

Peyrony, O., Hutin, A., Truchot, J., Borie, R., Calvet, D., Albaladejo, A., … & Roux, D. (2020). Impact of panelists' experience on script concordance test scores of medical students. *BMC Medical Education, 20*(1). https://doi.org/10.1186/s12909-020-02243-w

Phipps, L. J., Osborne, E. W., Dyer, J. E., & Ball, A. (2008). *Handbook on agricultural education in public schools* (6th ed.). Clifton Park, NY: Thomson Delmar.

Presley, M. L., Gehringer, R., & Hanuscin, D. L. (2017). Using Natural Selection Concept Inventories in College Biology Classrooms to Improve Teaching and Learning. *The American Biology Teacher*, *79*(2), 91–94. http://dx.doi.org/10.1525/abt.2017.79.2.91

Price, R. M., Andrews, T. C., McElhinny, T. L., Mead, L. S., Abraham, J. K., Thanukos, A., & Perez, K. E. (2014). The genetic drift inventory: a tool for measuring what advanced undergraduates have mastered about genetic drift. *CBE—Life Sciences Education*, *13*(1), 65–75. https://doi.org/10.1187/cbe.13-08-0159

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. The Danish Institute of Educational Research.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Salimpour, S., Tytler, R., Doig, B., Fitzgerald, M. T., & Eriksson, U. (2023). Conceptualizing the Cosmos: Development and Validation of the Cosmology Concept Inventory for high school. *International Journal of Science and Mathematics Education*, *21*(1), 251–275. https://doi.org/10.1007/s10763-022-10252-y

Shaw, K. R., Van Horne, K., Zhang, H., & Boughman, J. (2008). Essay contest reveals misconceptions of high school students in genetics content. *Genetics*, *178*(3), 1157–1168. https://doi.org/10.1534/genetics.107.084194

Smith, A., & Marbach-Ad, G. (2010). Learning outcomes with linked assessments – An essential part of our regular teaching practice. *Journal of Microbiology and Biology Education*, *11*(2), 123–129. https://doi.org/10.1128/jmbe.v11i2.217

Smith, M. K., Wood, W. B., & Knight, J. K. (2008). The genetics concept assessment: A new concept inventory for gauging student understanding of genetics. *CBE—Life Sciences Education*, *7*(4), 422–430. https://doi.org/10.1187/cbe.08-08-0045

Sparks, R. and Lipka, R. (1991). Selecting master teachers: A preliminary study. *Journal of Personnel Evaluation in Education, 4*(4), 367-373. https://doi.org/10.1007/bf00141443

Stefanski, K. M., Gardner, G. E., & Seipelt-Thiemann, R. L. (2016). Development of a lac operon concept inventory (LOCI). *CBE—Life Sciences Education*, *15*(2), ar24.

Steif, P. S. (2005). Psychometric analysis of a statics concept inventory and its use as a formative assessment tool. *Proceedings of the 2005 ASEE Annual Conference & Exposition*, Portland, OR.

Swafford, M. (2018). STEM education at the nexus of the 3-circle model. *Journal of Agricultural education*, *59*(1), 297–315. https://doi.org/10.5032/jae.2018.01297

Taber, K. S. (2017). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

The National Council for Agricultural Education. (2015). *Agriculture, food and natural resources (AFNR) career cluster content standards*. https://ffa.app.box.com/s/n6jfkamfof0spttqjvhddzolyevpo3qn/file/294160068843

Ward, L. D., French, B. F., Barbosa-Leiker, C., & Iverson, A. E. (2016). Application of exploratory factor analysis and item response theory to validate the Genomic Nursing Concept Inventory. *Journal of Nursing Education*, *55*(1), 9–17. https://doi.org/10.3928/01484834-20151214-05

Wind, S., & Hua, C. (2021). *Rasch measurement theory analysis in R: Illustrations and practical guidance for researchers and practitioners*. Bookdown.org, [Epub].