A peer reviewed, open-access electronic journal: ISSN 1531-7714

Leveraging Measures of Students' Prior Knowledge and Lived Experiences Toward More Personalized Culturally Responsive Assessment

Jesse R. Sparks, *ETS Research Institute*
Teresa Ober, *ETS Research Institute*
Blair Lehman, *Brighter Research*

Abstract: Approaches to personalized learning emphasize the role of students' knowledge, experiences, and interests in making classroom learning more relevant and engaging. Personalization that connects instruction to students' home and community funds of knowledge is consistent with culturally responsive pedagogy and culturally responsive assessment approaches emphasizing that classroom content should reflect students' cultural experiences. Measuring and accounting for variation in students' knowledge, interests, and lived experiences has implications for personalized socioculturally responsive assessments. Research on scenario-based assessments (SBAs) of reading comprehension, which situate test questions within an overarching scenario, context, and goal, indicates that measures of topical vocabulary knowledge embedded within the assessment explain variance in comprehension scores. We extend these lines of work to examine whether measures of students' funds of knowledge and lived experiences may further account for SBA performance and engagement, using empirical data from a geographically and racially diverse sample of approximately 1,300 middle school students. Specifically, we examined the degree to which students' topical vocabulary knowledge and self-reported funds of knowledge predicted SBA performance and test-taking engagement. Results suggest that funds of knowledge measures could be used in conjunction with topical vocabulary and engagement measures to provide adaptive personalization toward more socioculturally responsive SBAs.

Keywords: Personalized assessment, socioculturally responsive assessment, scenario-based assessment, reading comprehension

Introduction

Approaches to personalized learning emphasize the critical role of students' knowledge, experiences, and interests in making classroom learning more relevant and engaging (Bernacki et al., 2021; Walkington & Bernacki, 2020). A recent meta-analysis examining the impact of personalized and adaptive learning technologies in the K-12 literacy domain revealed an overall positive effect ($g = 0.29$) of personalized reading

interventions on learning, compared to interventions that did not employ personalization (Alrawashdeh et al., 2024). Personalization is a large umbrella that can cover many constructs; thus, it is important to consider what constructs are being leveraged to personalize the learning experience. Personalization that connects instruction to students' home and community funds of knowledge (González et al., 2005; Moll et al., 1992) is consistent with culturally responsive pedagogy and culturally responsive assessment approaches emphasizing that classroom content should reflect students' cultural experiences (Evans & Taylor, 2025; Gay, 2002; Hood, 1998; Ladson-Billings, 1995b; Lee, 1998; Qualls, 1998). Such approaches to culturally relevant education have demonstrated success at engaging students with educational activities and positively impacting academic achievement across the curriculum (see Aronson & Laughter, 2016, for a review). However, despite the many hypothesized benefits of personalized, culturally responsive assessments (see Bennett, 2023, 2025), there are still challenges affecting their implementation. For example, measuring and accounting for variation in students' knowledge, interests, and lived experiences has implications for the validity and fairness of such assessments (Bennett, 2023; Mislevy, 2018; Sinharay, Kane, et al., 2025).

We have previously argued that scenario-based assessments (SBAs)—performance-based assessments that situate test questions within an overarching scenario, context, and goal (Sabatini et al., 2018, 2020)—are an especially appropriate context in which to examine questions related to personalized socioculturally responsive assessments (O'Dwyer et al., 2023; Sparks et al., in press; Wang et al., 2025). Because SBAs are contextualized within goal-driven scenarios, involving thematically related materials focused on a specific topic, students may find these scenarios and topics more or less culturally relevant to them based on their prior experiences (Nakkula & Toshalis, 2025), which could limit their engagement with the SBA (Lehman, Sparks, Zapata-Rivera, et al., 2024). Given the importance of engagement to ensure valid interpretations of test performance (Wise & Smith, 2016), it is critical that such contextualized assessments employ a flexible approach to meet each student where they are. We propose personalized socioculturally responsive assessments as an approach to achieving this vision.

The degree of cultural relevance of scenarios in the context of reading comprehension SBAs has been shown to have implications for students' engagement and performance on those assessments (Wang et al., 2025). In other words, students with more background knowledge related to the topic of the SBA may perform better. For example, prior research indicates that measures of *topical vocabulary knowledge* embedded within SBAs explain variance in comprehension scores (McCarthy et al., 2018; O'Reilly, Wang, et al., 2019; Wang et al., 2021; Wang et al., 2025). In the current paper, we extend these lines of work to examine whether measures of students' *funds of knowledge and lived experiences* may further account for SBA performance and engagement beyond vocabulary-based knowledge measures, using empirical data from a geographically and racially diverse sample of approximately 1,300 middle school students. Specifically, we examined the degree to which students' topical vocabulary knowledge and relevant self-reported funds of knowledge predicted SBA performance and test-taking engagement. If these self-report instruments account for performance and engagement during comprehension tasks, such measures could be used in conjunction with topical vocabulary measures to provide personalized socioculturally responsive SBAs (Sparks et al., in press).

Socioculturally Responsive Assessment Approaches

Various approaches to creating more equitable assessments for all students, but especially those who have been historically underserved by traditional educational assessments, have been proposed (Bennett, 2025; Evans & Taylor, 2025; Hood, 1998; Lee, 1998; Qualls, 1998; Ramasubramanian et al., 2021; Randall, 2021; Randall et al., 2021; Walker et al., 2023). Common to these approaches is the aim "to design assessment for the social, cultural, and other relevant characteristics of diverse students and the contexts in which they live" (Bennett, 2023, p. 86). Several of these approaches draw from theories and models of culturally relevant pedagogy (Ladson-Billings, 1995a, 1995b, 2009, 2014), culturally responsive teaching (Gay, 2002, 2010, 2013), and culturally sustaining pedagogy (Paris, 2012), which generally aim to make instruction more

relevant and responsive to the identities and cultural backgrounds of students in diverse classrooms. Applying those instructional theories to the assessment context becomes increasingly critical as the U.S. student population becomes more economically, racially, ethnically, and culturally diverse (National Center for Education Statistics, 2024). Traditional standardized assessments tend to be reflective of a common (dominant) cultural paradigm and are generally not designed to capture the heterogeneous multicultural characteristics represented by today's students (Bennett, 2025; Randall, 2021).

Various assessment principles have been articulated for socioculturally responsive assessment and related conceptions (see Bennett, 2025, for a review). This includes *connecting* to individuals' cultural backgrounds, identities, and experiences, *adapting* to individual characteristics, including cultural identities, languages, interests, and prior knowledge, and *characterizing assessment results as an interaction* among student characteristics, features of the tasks, and the testing context (Bennett, 2023). Attention to the ways in which learners' prior knowledge and experiences and their interactions with other student characteristics and behaviors may affect assessment performance is a key tenet of frameworks for socioculturally responsive assessment. Assessments that draw on students' cultural and linguistic assets, funds of knowledge, and individual characteristics should engender greater student engagement and greater effort to complete the tasks successfully (Nakkula & Toshalis, 2025).

Work specifically focused on the context of performance assessment (of which SBAs are one example) underscores the need to ensure that assessment content is designed to be relevant to and reflective of students' lived experiences (Hood, 1998; Lee, 1998; O'Dwyer et al., 2023). Hood (1998) argued that poor outcomes on performance tasks may reflect a disconnect between students' cultural experiences and the experiences represented within assessment tasks, highlighting a major validity issue. Simply put, if the scenarios, goals, topics, and contexts used in performance tasks are unfamiliar to students, situating the work in those contexts may interfere with students' performance, rather than facilitate it. Consideration of cultural and linguistic backgrounds of students is essential in defining the constructs, skills, tasks, and topics the assessment will target (Hood, 1998) to ensure their appropriateness.

O'Dwyer et al. (2023) offered specific design principles that assessment developers can apply to produce more culturally relevant SBAs, including: (1) incorporating community-based learning into scenarios; (2) emphasizing empowerment through student-led action and problem-solving applied in local contexts; (3) incorporating multiple, diverse perspectives to support student perspective-taking; (4) negotiating trade-offs among familiarity, novelty, and engagement in the selection of contexts and scenarios; and (5) reflecting the multiple identities students bring to assessments by representing diverse characters (e.g., in terms of gender, race, ethnicity, age, and disability status). We draw on these principles in the current work and use them to guide the selection of learner variables we examine (i.e., prior knowledge, community-based experiences, engagement) to inform the design of personalized adaptations to SBAs.

Personalized Assessment

Personalized assessment can be viewed as a wider umbrella under which socioculturally responsive pedagogy and assessment are notable examples. Designers of personalized assessments must choose what student characteristics and/or behaviors will be selected to drive personalization. Most personalized assessments leverage current knowledge estimates, as is the case with computer-adaptive assessments (CATs; van der Linden & Glas, 2010), but there are also assessments that personalize based on students' test-taking engagement (Wise et al., 2006) and cultural backgrounds (Bennett, 2023; Hood, 1998; Lee, 1998; Qualls, 1998). Typically, personalized assessments provide: (a) assessment tasks that are sufficiently open-ended to allow students to personalize their experience, (b) different versions of the assessment to align with student characteristics, and/or (c) real-time adaptive support based on student behaviors during the assessment (Bennett, 2024). The determination of when and how to provide personalized support can be achieved through a variety of approaches that can be grouped into three categories: human-driven, machine-driven,

and hybrid (Arslan et al., 2024; Bennett, 2025). Human-driven personalization requires experts to author rules for when and how personalized support should be delivered, whereas machine-driven personalization often leverages machine learning (ML) algorithms and artificial intelligence (AI) to learn from the data when and how to personalize the learning experience. Hybrid approaches are often used to address transparency and interpretability issues by leveraging a combination of expert human insights and ML and AI methodologies (Paquette et al., 2014).

In the personalized assessment context, different versions of an SBA could be developed where one or more assessment components vary to provide a more engaging experience for students, while retaining the same measurement targets (Mislevy, 2018). These different versions could then be assigned to individual students through human-driven (student, teacher), machine-driven, or hybrid approaches. Similarly, real-time adaptive support can be delivered on-demand when requested by students (Sparks et al., 2024), when a teacher has been notified that a student requires support (Wise et al., 2019), or when the system detects that the student requires support (Wise et al., 2006). Particularly in the case of characteristics and behaviors that are less well defined for personalization, hybrid approaches are often more effective for developing models that identify non-cognitive states and can then drive personalization (Paquette & Baker, 2018). Regardless of the approach used for personalization, the goal for personalized assessments remains the same – to give students their best opportunity to achieve their educational goals in tasks that are engaging, relevant, and appropriate for them.

Personalization via Caring Assessments

The approaches to personalized assessment discussed so far have generally been deployed individually and with one characteristic or behavior as the focus of personalization. However, the culturally-enhanced caring assessments (CECA) framework (Lehman, Sparks, Zapata-Rivera, et al., 2024) advocates for personalizing assessments by considering a holistic picture of the student that goes beyond a single characteristic or behavior to incorporate students' cultural identities, language backgrounds, and lived experiences—and their intersections—to better understand, model, and respond to student needs (Bernacki et al., 2021; Sparks et al., 2024; Zapata-Rivera, 2017). Using this holistic view of students to guide personalization also requires a more holistic approach to personalization itself. Rather than adapting the assessment version or providing real-time adaptive support, CECA promotes the use of both approaches simultaneously to provide each student with optimal conditions for demonstrating what they know and can do.

CECA proposes to achieve this holistic approach to personalized assessment through an integrated learner model (i.e., a digital representation of the student that is updated as evidence is gathered; Shute & Zapata-Rivera, 2012; Zapata-Rivera et al., 2020; Zapata-Rivera & Arslan, 2021) that provides insights into students' intersectional characteristics and the ways in which those characteristics interact with their behaviors while completing assessments. This combined approach to personalization enables "caring" support to be delivered at multiple time points: before, during, and after the assessment. Caring before the assessment can be achieved by adjusting assessment versions based on learner profiles. Caring during the assessment involves just-in-time adaptations based on students' in-task states (e.g., engagement), with a general goal to maintain students' situational interest and keep them motivated to perform their best. Lastly, caring can occur after the assessment in the form of personalized feedback reports to teachers and students that leverage student behaviors during the assessment and student characteristics to contextualize performance feedback.

Despite the potential of CECA, there are still many challenges to realizing this approach to personalized assessments. One challenge is that the relevant student characteristics must be measured or captured in such a way that supports such an intersectional, holistic approach. For example, simply gathering student demographic information may not be sufficient to provide an engaging, culturally relevant assessment (Wang

et al., 2025). The present work addresses this challenge by examining relationships among several measures tapping different dimensions of students' knowledge (including funds of knowledge and lived experiences) and SBA performance. By developing measures designed to capture students' lived experiences relevant to SBA topics and investigating the interaction of lived experiences with measures of topical vocabulary knowledge, this study allows for an expanded understanding of how knowledge and lived experiences impact assessment performance with implications for designing personalized assessments. A second challenge for CECA is how to adapt assessments to meet students' needs once the relevant characteristics are measured. We return to this issue in the Discussion.

Measuring Prior Knowledge and Metacognition

Prior knowledge is a particularly important characteristic to consider when designing personalized assessments for reading comprehension (Hattan et al., 2024; Hwang et al., 2022; Kendeou et al., 2004; McCarthy & McNamara, 2021; O'Reilly, Sabatini, & Wang, 2019; O'Reilly, Wang, & Sabatini, 2019; Simonsmeier et al., 2022). However, reading assessments have traditionally been developed to *minimize* the role of prior knowledge (see O'Reilly & Sabatini, 2013; Pearson et al., 2021; RAND Reading Study Group, 2002). SBAs, on the other hand, allow for connections to different dimensions of students' prior knowledge (e.g., academic knowledge vs. topic knowledge; Wang et al., 2021; see also McCarthy et al., 2023). Emerging approaches to measuring knowledge within SBAs (Sabatini et al., 2020) using vocabulary knowledge- or factual content-based items indicate that such knowledge measures are valid and reliable predictors of comprehension (Higgs et al, 2023; McCarthy et al., 2018; O'Reilly, Sabatini et al., 2019), which provides an opportunity to support personalization on the basis of prior knowledge. The current study extends these lines of work to explore the measurement of additional dimensions of knowledge, elicited through self-reports of students' topic knowledge and lived experiences.

Another area ripe for personalization is the degree to which students are aware of limitations on their knowledge (i.e., metacognition; Flavell, 1979). If students do not have the relevant prior knowledge to answer a question about their prior knowledge, they may struggle to provide a response that reflects their true state of understanding. When students lack the necessary prior knowledge, they may guess or omit answers, leading to less valid assessments (Lau et al., 2011). One simple yet underexamined assessment design feature that enables students to respond even in cases of low knowledge or uncertainty involves the inclusion of an "I don't know" (IDK) response option. Prior experimental research suggests that incorporating such IDK or "unsure" options can improve response quality and interpretability of responses (Scoboria & Fisico, 2013; Wakabayashi & Guskin, 2010). The inclusion of IDK options may also contribute to the interpretability of responses by distinguishing between genuine knowledge gaps and potential misconceptions (Scoboria & Fisico, 2013; Wakabayashi & Guskin, 2010). More accurate assessment of students' prior knowledge can enable more effective personalization. Within an SBA context, O'Reilly, Sabatini et al. (2019) found that even after accounting for the effect of prior knowledge, greater use of an IDK option within the prior knowledge measure was associated with higher comprehension scores and learning from text among a sample of middle school students. The current study further extends these findings by examining how use of the IDK option within prior knowledge measures relates to students' prior knowledge, comprehension, and engagement. We also examined whether students perceive this novel SBA feature as supporting their motivation and performance.

The Current Study

The current study reflects a critical step toward development of personalized socioculturally responsive SBAs that provide all students with engaging, supportive experiences based on their individual and sociocultural characteristics (Sparks et al., in press; Wang et al., 2025). Our approach seeks to use evidence of students' topical vocabulary knowledge and self-reported funds of knowledge to drive adaptations to SBA content, if such measures indeed predict students' engagement and comprehension performance within

SBAs. To examine students' funds of knowledge (González et al., 2005; Moll et al., 2005), we developed and administered Likert-type self-report measures assessing students' funds of knowledge related to organic farming, as well as related (potentially more familiar) topics of organic food and organic gardening. For example, students reported whether there were organic farms in their community, or their families had access to organic foods (*organic food access*), and whether they were involved in buying food, preparing meals, or participating in organic gardening (*food preparation/gardening experience*). By developing measures designed to capture students' lived experiences relevant to the SBA topic, as well as administering measures of topical vocabulary knowledge like those used in prior SBA studies, we can examine how such knowledge and experience measures may interact to influence students' engagement with and performance on the assessment, with implications for personalized assessments.

In this paper, we address the following research questions.

RQ1. Do measures of self-reported prior knowledge and experiences predict reading comprehension as measured by SBA scores, over and above topical vocabulary measures?

RQ2. To what extent do measures of students' prior knowledge and experiences predict task engagement in the SBA?

RQ3. To what extent does task engagement predict reading comprehension, over and above measures of students' prior knowledge and experiences?

RQ4. Does students' use of the IDK option within the prior knowledge measures predict reading comprehension or task engagement?

RQ5. Do students' perceptions of the IDK option within the prior knowledge measures interact with the relationship among use of IDK, performance, and engagement?

RQ6. To what extent do students' levels of prior knowledge and experience interact with one another or with other factors (i.e., demographics, engagement) to predict SBA performance?

Methodology

Participants

We collected data from a sample of 1,412 6th to 8th grade students from four school districts in the U.S. This includes two small urban middle schools in a high-needs district in the Northeast, two schools from suburban districts in the South, and one school from a rural district in the South. For this study, relationships among knowledge, performance, and engagement were explored using data from 1,347 students who completed the SBA. School staff provided information about students' characteristics via class rosters, including demographics, socioeconomic status (i.e., free/reduced price lunch participation), Title I accommodation status, special education status, English learner status, English Language Arts (ELA) class grades, and most recent state test scores[1] for ELA and Science (z-transformed within each school district for comparison across sites; see Table 1).

In the high-needs district, students generally had low ELA grades, were mainly Hispanic/Latino (64%) or Black/African American (24%), and most (76%) were students with (one or more) disabilities; all students participated in the free/reduced price lunch program. Among the suburban schools, one was situated within a small city and racially and economically diverse (41% white, 39% Black/African American, 13%

[1] Testing year varied depending on school site and subject. In most cases, test scores reported were from the preceding school year (2021-2022 school year). Several test scores reported reflected administrations as early as 2018 or 2019, likely reflecting disruptions to standardized testing schedules due to the COVID-19 pandemic.

Hispanic/Latino; 50% free/reduced lunch), while the other was situated in a more rural area and somewhat less racially diverse (74% white, 23% Black/African American; 47% free/reduced lunch). The rural Southern school sample was mainly white (89%) low-income students (99% free/reduced lunch). Thus, our overall sample was diverse in terms of geography, urbanicity, race/ethnicity, socioeconomic status, and academic achievement, suggesting that personalized and culturally responsive assessment approaches would be appropriate.

Measures

Pre-Assessment Surveys

Students completed several survey questionnaires prior to completing the SBA. These surveys were designed to collect background information about student characteristics that could inform an integrated learner model and thus ultimately be used to incorporate adaptations to the SBA to support personalization and cultural responsiveness. In addition to the knowledge-related items that are our current focus, students also completed measures of other social-emotional characteristics (e.g., growth mindset, cognitive flexibility, reading motivation, reading self-efficacy) during the pre-survey which are not reported here. All survey measures were administered using an online survey platform that collected responses and timing data.

Self-reported knowledge and experience measures. We administered a series of self-report measures to assess students' funds of knowledge and lived experiences related to organic farming (see Table 2). Students completed four measures, described in turn.

Self-report topic knowledge. Students completed one multiple-choice item asking them to report their level of knowledge about organic farming (nothing, a little, some, a lot).

Self-report sources of knowledge. Students completed one select-all-that-apply item asking them to identify the sources from which they have learned about organic farming (e.g., classes, family, community members). Students were also given the option to report if they had never learned about this topic.

Funds of knowledge for organic farming. Students were presented with a block of eight items measuring their funds of knowledge for organic farming in a grid format with a 5-point Likert-type response scale ranging from Strongly Disagree to Strongly Agree. The midpoint was labeled "Neither Agree nor Disagree", and students were also given the option to respond IDK to any item (IDK responses were treated as missing for subsequent analyses; see Ober et al., 2024). Confirmatory factor analysis (CFA) was conducted to test a two-correlated factors model, indicating reasonably good model fit [$\chi^2(df = 19) = 119.92$, $CFI = .96$, $TLI = .94$, $RMSEA = .08$, $SRMR = .04$] based on meeting two of three recommended benchmarks of CFI $\geq .95$, RMSEA $\leq .06$, and SRMR $\leq .08$ (Hu & Bentler, 1999). As such, items were divided into two subscales ($\varrho = .595$); Table 3 presents reliability statistics (α) for each subscale and the overall scale. The first subscale (3 items; overall $\alpha = .779$) measured prior learning about organic farming, specifically whether students had learned about this topic in school, outside of school on their own, or outside of school with family or community members, and was labeled *prior learning*. The second subscale (5 items; overall $\alpha = .822$) asked students to rate the importance of organic farming for their school, themselves, their family, and their community; students were also asked to report if there were organic farms in their community. This subscale was labeled *local relevance*. These items enabled us to assess students' experiences around the topic of organic farming, including school-, home-, and community-based experiences.

Funds of knowledge for organic food, food preparation, and gardening. Students were presented with a block of 12 items measuring their funds of knowledge for organic food, food preparation, and gardening in a grid format with a 5-point Likert-type response scale ranging from Strongly Disagree to Strongly Agree. As with the previous survey, the midpoint was labeled "Neither Agree nor Disagree", and students were also given the option to respond IDK to any item. Items were divided

Table 1. Sample Demographics and State Test Performance

	Overall Sample (N=1,347)	**Midsize City (n=72)**	**Suburban/ Small City (n=739)**	**Suburban/ Rural (n=395)**	**Rural (n=141)**
	f (%)	*f* (%)	*f* (%)	*f* (%)	*f* (%)
Grade Level					
6	434 (32.2%)	17 (23.6%)	239 (32.3%)	140 (35.4%)	38 (27.0%)
7	460 (34.1%)	31 (43.1%)	260 (35.2%)	121 (30.6%)	48 (34.0%)
8	448 (33.3%)	24 (33.3%)	239 (32.3%)	134 (33.9%)	51 (36.2%)
Gender					
Male	742 (55.1%)	54 (75.0%)	402 (54.4%)	212 (53.7%)	74 (52.5%)
Female	599 (44.5%)	17 (23.6%)	336 (45.5%)	183 (46.3%)	63 (44.7%)
Nonbinary	1 (0.1%)	1 (1.4%)	0 (0%)	0 (0%)	0 (0%)
Race/Ethnicity					
American Indian or Alaskan Native	3 (0.2%)	0 (0%)	2 (0.3%)	1 (0.3%)	0 (0%)
Asian or Asian American	24 (1.8%)	0 (0%)	15 (2.0%)	9 (2.3%)	0 (0%)
Black or African American	401 (29.8%)	17 (23.6%)	285 (38.6%)	92 (23.3%)	7 (5.0%)
Hispanic/Latino	147 (10.9%)	46 (63.9%)	99 (13.4%)	1 (0.3%)	1 (0.7%)
Native Hawaiian or Pacific Islander	4 (0.3%)	0 (0%)	3 (0.4%)	1 (0.3%)	0 (0%)
White	732 (54.3%)	9 (12.5%)	306 (41.4%)	291 (73.7%)	126 (89.4%)
Two or more races	28 (2.1%)	0 (0%)	28 (3.8%)	0 (0%)	0 (0%)
Free/Reduced Price Lunch Participation	766 (56.9%)	72 (100%)	369 (49.9%)	186 (47.1%)	139 (98.6%)
Title I Accommodation: Yes	185 (13.7%)	62 (86.1%)	113 (15.3%)	0 (0%)	10 (7.1%)
Special Education Status: Yes	162 (12.0%)	55 (76.4%)	67 (9.1%)	34 (8.6%)	6 (4.3%)
English Learner (EL) Status					
Current or Former EL	37 (2.7%)	9 (13.9%)	25 (3.4%)	2 (0.5%)	0 (0%)
Non-EL	1,307 (97.0%)	63 (86.1%)	713 (96.5%)	393 (99.5%)	139 (98.6%)

Current ELA Grade					
A+/A/A-	556 (41.3%)	1 (1.4%)	276 (37.3%)	210 (53.2%)	69 (48.9%)
B+/B/B-	368 (27.3%)	14 (19.4%)	212 (28.7%)	98 (24.8%)	44 (31.2%)
C+/C/C-	215 (16.0%)	21 (29.2%)	131 (17.1%)	47 (11.9%)	16 (11.3%)
D+/D/D-	126 (9.4%)	27 (37.5%)	79 (10.7%)	18 (4.6%)	2 (1.4%)
F	70 (5.2%)	9 (12.5%)	40 (5.4%)	21 (5.3%)	0 (0%)
	M (SD), Min-Max	*M (SD), Min-Max*	*M (SD), Min-Max*	*M (SD), Min-Max*	*M (SD), Min-Max*
ELA State test score (Z-score transformed)	0.01 (1.00), -3.45 - 2.84	0.01 (1.04), -1.39 - 2.48	0.02 (0.99), -2.77 - 2.84	0.00 (1.00), -3.41 - 2.61	0.02 (1.01), -3.45 - 2.59
Science State test score (Z-score transformed)	0.00 (1.00), -3.89 - 3.01	-0.06 (1.10), -2.13 - 2.01	0.02 (0.99), -2.28 - 3.01	-0.00 (1.00), -2.76 - 2.32	-0.03 (1.02), -3.89 - 2.80

Table 2. Self-Reported Knowledge and Experience Measures

Measure	**Items**	**Response Options**
Self-report Topic Knowledge	How much do you know about the topic of ***organic farming***? Choose one answer choice.	Nothing, A little, Some, A lot
Self-report Sources of Knowledge	I have learned about ***organic farming*** from these sources (select all that apply):	Classes; Friends; Family; Community members; Museums; School field trips; Visiting a farm; TV shows, movies; Books; Internet (websites); Internet (social media); Games; Other (please specify); I have never learned about this topic
Funds of Knowledge for Organic Farming	The following questions ask about your prior experiences with the topic of ***organic farming***. Please rate your agreement with each statement by selecting **one** answer choice for each row. If you don't know or aren't sure, you can choose "I don't know".	Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree, I don't know
Prior Learning	I have learned about ***organic farming*** in school.	
	I have learned about ***organic farming*** outside of school, learning on my own.	

	I have learned about ***organic farming*** outside of school, with my family at home or with people in my community.	
Local Relevance	***Organic farming*** is important to my school.	
	Organic farming is important to me.	
	Organic farming is important to my family.	
	Organic farming is important to my community.	
	There are ***organic farms*** in my community.	
Funds of Knowledge for Organic Food and Gardening	The following questions ask about your prior experiences with ***organic foods and gardening***. Please rate your agreement with each statement by selecting **one** answer choice for each row. If you don't know or aren't sure, you can choose "I don't know".	Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree, I don't know
Organic Food Access and Value	My family can easily access organic food in our community.	
	Organic food is affordable for my family.	
	My family eats organic food often.	
	My family thinks it is important to buy and eat organic food.	
	My community thinks it is important to buy and eat organic food.	
Food Preparation and Gardening	I am involved in buying food for my home/family (for example, going to the store/market to get groceries).	
	I am involved in deciding what foods my home/family eats (for example, deciding what groceries to buy, what foods to eat).	
	I am involved in preparing meals at home (for example, cooking food).	
	My family has a home garden (for example, a backyard or container garden).	
	My community has a community garden.	
	I have helped work in a home or community garden (for example, planting, watering, harvesting).	
	I have helped work in a garden that uses organic farming methods.	

into two subscales. A CFA was conducted to test a two-correlated factors model, and while model fit was not ideal[2] [$\chi^2(df = 53) = 511.30$, *CFI* = .88, *TLI* = .84, *RMSEA* = .10, *SRMR* = .06], we proceeded with treating the two dimensions as separate scales (ϱ = .507) in light of the items' conceptual alignment and acceptable internal consistency (see Table 3). The first subscale (5 items; overall α = .819) measured students' access to and perceived value of organic food (e.g., "my family thinks it is important to buy and eat organic food"). The second subscale (7 items; overall α = .778) asked students to report their experiences with food preparation and gardening (e.g., "I am involved in preparing meals at home"; "My community has a community garden"). These items enabled us to assess students' experiences around organic food, preparing food, and gardening, which may be considered more familiar "everyday" topics with potentially greater personal relevance to students than organic farming per se. Students' experiences with these everyday practices might meaningfully be leveraged to support them in making connections to the potentially less familiar topic of organic farming.

Organic Farming SBA

The Organic Farming SBA (Sabatini et al., 2014) measures applied reading comprehension skills. This SBA included reading passages introducing students to information about organic farming methods, differences from conventional methods, and pros and cons. Students' goal was to work with virtual peers to develop a website about the topic. The task comprised a variety of comprehension tasks, including summarization, evaluating web search results, identifying pros and cons, correcting misconceptions, and integrating multiple texts and perspectives. Students could earn up to 30 points on the assessment. Total scores on the SBA were used as our measure of reading comprehension in this study (overall α = .878, see Table 3).

Topical vocabulary knowledge measures. Topical vocabulary knowledge of organic farming was assessed within the SBA using selected-response items (O'Reilly, Sabatini, et al., 2019; Wang et al., 2025). At the beginning of the SBA, students completed two types of vocabulary items. The first measured knowledge of topical vocabulary (TV) by asking students to identify whether 46 vocabulary words were *related or unrelated* to the topic of organic farming (overall α = .872, see Table 3). The second item type included five vocabulary-in-context (VIC) items that presented students with an underlined term in a sentence context and asked them to *choose the closest meaning* from a list of three one-word options (overall α = .536). For each item type, students were also given an IDK option. Students could earn up to 46 points on TV items and 5 points on VIC items, with IDK responses scored as incorrect. Total prior knowledge scores were computed by combining the two item types (ϱ = .449), for a total of 51 points (overall α = .879). While the TV items showed sufficient reliability, the VIC items did not, likely due to the small number of items. Thus, we analyzed these two topical vocabulary knowledge measures separately, rather than using the combined score based on all 51 items.

Test-taking engagement. Test-taking engagement was measured by leveraging response times to develop a proxy measure for cognitive engagement for each screen that participants interacted with on the SBA. The Normative Threshold Method (Wise & Ma, 2012) was used to convert response times for each

[2] For the two-factor CFA of this measure, modification indices suggest that the largest sources of misfit come from correlated residuals between certain item pairs, particularly "I am involved in deciding what foods my home/family eats (for example, deciding what groceries to buy, what foods to eat)." with "I am involved in preparing meals at home (for example, cooking food)." (MI ≈ 21) and "I have helped work in a home or community garden (for example, planting, watering, harvesting)." with "I have helped work in a garden that uses organic farming methods." (MI ≈ 19), indicating these items share variance not explained by the two-factor model, possibly due to overlap in item wording. Standardized residuals show the biggest discrepancies between observed and model-implied covariances for "My community has a community garden." and "My family can easily access organic food in our community." (≈ 1.94), suggesting these items may not align well with the assumed factor structure, although no values for standardized residuals exceed 2, indicating that the model reasonably reproduces the observed relationships among items.

screen for each participant to designate each screen as reflecting an engaged or disengaged interaction (i.e., whether sufficient time was spent on the screen for an engaged interaction to occur). This method involves

Table 3. Reliability of SBA and Knowledge Measures (Coefficient Alpha)

	Overall Sample (N=1,347)	**Midsize City (n=72)**	**Suburban/ Small City (n=739)**	**Suburban /Rural (n=395)**	**Rural (n=141)**
Prior Topic Vocabulary Knowledge Score Overall (max score = 51)	.879	.944	.866	.858	.896
Topical Vocab (TV) Score (max score = 46)	.872	.945	.857	.849	.893
Vocabulary-in-Context (VIC) Score (max score = 5)	.536	.594	.530	.513	.492
SBA Total Score (max score = 30)	.878	.729	.872	.894	.856
Funds of Knowledge for Organic Farming	.865	.881	.871	.858	.843
Prior Learning	.779	.795	.782	.804	.679
Local Relevance	.822	.837	.830	.816	.801
Funds of Knowledge for Organic Food and Gardening	.854	.913	.860	.828	.854
Organic Food Access and Value	.819	.863	.833	.784	.808
Food Preparation and Gardening	.778	.862	.787	.739	.779

taking the average response time for each item (or screen in the present context) across all participants, taking a percentage of the average response time to set a threshold, and then comparing each student's response time to that threshold. If a response time was the same as or greater than the threshold, then the interaction was deemed to be engaged. If a response time was less than the threshold, the interaction was deemed to be disengaged. The threshold can vary from 10% to 30% of average response time based on the task and the goals of the engagement measure (Wise & Kuhfeld, 2021).

In the present research, we used a 30% threshold for each screen as prior research has shown this percentage to be related to key construct-relevant variables (Lehman, Sparks, & Steinberg, 2024) and to be useful for post-test investigations of student engagement (Lehman et al., 2025; Wise & Kuhfeld, 2021). After screen-level engagement was determined for each participant, we computed response time engagement (RTE) for each participant (Wise & Kong, 2005). RTE is the proportion of items (screens) out of the total number of items (screens) that were determined to be an engaged interaction. For example, if there are 10 screens and a participant is engaged on eight of those screens, their RTE would be .8 (8 ÷ 10 = .8). Each participant's RTE for the entire SBA was used as our measure of engagement in this study.

Post-Assessment Surveys

Students completed several additional survey measures after completing the SBA. These post-assessment surveys focused on self-reports of test-taking effort, achievement emotions, and perceptions of various features experienced during the SBA. For the present analysis, we focused on students' perceptions of SBA

features. Students were presented with two blocks of 23 items in a grid format, where each item described a specific design feature in the SBA (e.g., the topic of organic farming, the option to respond "I don't know" in the glossary questions at the beginning of the test, using the texts to answer questions). Each item included a 7-point Likert-type scale with response options ranging from Strongly Disagree to Strongly Agree; the midpoint was labeled as "Neither Agree nor Disagree" and there was no IDK option provided. In the first block, students were asked to rate each aspect of the SBA in terms of whether the feature helped motivate them to complete the test (i.e., supported engagement). In the second block, students were asked to rate the same set of 23 features in terms of whether the feature helped them do better on the test (i.e., supported performance). These items allowed us to determine if novel SBA features such as the IDK option in the topical vocabulary knowledge measures were perceived by students to be motivating and beneficial for their performance.

Procedure

Students participated in the study during the school day across multiple sessions. In the first session, students completed pre-assessment surveys. In the second session, students completed the SBA. If time allowed, students completed post-assessment surveys in the same class period. Otherwise, students completed post-assessment surveys in a third session. Teachers completed a roster with details on each participant and removed student names from the file before submitting it to the research team. These participant characteristics were linked to the response data via an anonymized ID.

Data Analyses

Preliminary analyses conducted to compute descriptive statistics, scale reliabilities, and zero-order correlations (overall, and by school site) were performed using SPSS version 29, while CFA and regression analyses were conducted using the *lavaan* (Rosseel, 2012) and *lm.beta* (Behrendt, 2014) packages in R. Hierarchical linear regression was used to model students' reading comprehension scores (SBA performance) and engagement on the task (RTE; Wise & Ma, 2012). Predictors were entered in sequential blocks to assess the added variance explained by self-report measures over and above baseline and vocabulary predictors. These analyses examined whether topical vocabulary knowledge and self-report funds of knowledge measures predicted SBA performance (RQ1) and engagement (RQ2) after controlling for student background characteristics. A third model combined knowledge measures and engagement to predict SBA performance (RQ3), examining the additional variance explained by incorporating engagement as a predictor of comprehension, controlling for student background characteristics and knowledge measures. The degree to which students' use of the IDK option in the topical vocabulary knowledge and the self-report knowledge and experience measures predicted performance and engagement (RQ4) was explored within the context of the preceding regression models. To further explore impact of the IDK option, additional analyses were conducted examining the degree to which students' perceptions of this option in the topical vocabulary knowledge measures (based on the post-assessment survey) interacted with their use of that option, their test-taking engagement, and their SBA performance. Students' post-test ratings of the IDK feature and associated interaction terms were entered into another set of regression models (RQ5). Finally, to address the degree to which students' prior knowledge and experiences interact with other factors (i.e., demographics, engagement) to predict SBA performance (RQ6), we examined subgroups of students based on their levels of topical vocabulary knowledge and self-reported funds of knowledge, in relation to other characteristics.

Results

Before discussing the detailed results addressing each research question, we present some preliminary descriptive statistics for key variables and zero-order Spearman correlations among these variables and students' performance on the SBA.

Descriptive Statistics

Descriptive statistics (means, standard deviations, minimums, and maximums) for the SBA scores and knowledge measures appear in Table 4. Overall, students demonstrated modest performance on the SBA, with an average score of 12.17 of 30 possible points (approximately 41% correct). SBA performance varied across school sites, with the urban school showing lower performance (~27% correct) compared to the suburban schools (~40% and 44% correct) and the rural school (~43%). Students attending schools in more rural districts showed higher performance than students in the small and midsize city districts.

Performance on the topical vocabulary knowledge measures embedded in the SBA were similarly modest, with an overall mean of 24.55 out of 46 for TV items, and an overall mean of 2.31 out of 5 for VIC items (~53% and 46% correct, respectively). Across sites, these prior knowledge measures showed a similar pattern as with SBA scores, with the urban school showing the lowest scores, and the more rural suburban school showing the highest scores. Examination of students' use of the IDK option in these measures indicated that, on average, students did make use of this option on ~28% of TV items (about 13 out of 46) and 21% of VIC items (about 1 out of 5). Students in the midsize city school were slightly more likely to use IDK on both item types than students at other schools.

In general, the self-report knowledge measures indicated that students reported modest levels of knowledge about organic farming (overall $M = 1.95$). Examination of response selections for the standalone self-report item indicated that approximately 35% of students reported they knew nothing about organic farming ("none"), while 40% reported "a little" knowledge, 20% reported "some" knowledge, and 5% reported "a lot" of knowledge on the topic. On average, students reported learning about organic farming from about 3.5 sources (median = 3.0), with the most common sources selected including classes (45%), TV shows and movies (41%), family (37%), Internet websites (34%), school field trips (29%), social media (28%), visiting a farm (27%), and books (26%); 31% of students reported that they had never learned about organic farming.

On average, students tended to slightly disagree that they had previously learned about organic farming (overall $M = 2.62$, 1 = Strongly Disagree, 5= Strongly Agree) and that organic farming was relevant to themselves and their communities (overall $M = 2.83$). Students tended to slightly agree when asked about their access to and perceived value of organic foods (overall $M = 3.39$), with urban students being more neutral ($M_{MidsizeCity} = 2.90$). Students tended to be neutral on their involvement in food preparation and gardening (overall $M = 2.91$), with urban students reporting slightly lower average agreement on this subscale ($M_{MidsizeCity} = 2.74$) than the other three school sites ($M_{Suburban/SmallCity} = 2.92$; $M_{Suburban/Rural} = 2.94$; $M_{Rural} = 2.88$). Students rarely used the IDK option when responding to prior learning items (overall $M = 9\%$) and were more likely to use this option in response to measures of local relevance (overall $M = 18\%$), organic food access (overall $M = 20\%$), and food preparation and gardening experience (overall $M = 13\%$).

Zero-Order Correlations

We next examined zero-order Spearman correlations among students' overall SBA scores, student background variables as reported in the teacher-submitted rosters, scores on embedded topical vocabulary knowledge measures (TV, VIC, and use of the IDK option in these measures), SBA engagement, and self-report measures of prior knowledge and experiences (see Table 5). Spearman correlations are reported because measures other than SBA scores were not assumed to be normally distributed.

Table 4. Descriptive Statistics for SBA Scores and Knowledge Measures

	Overall Sample (N=1,347)	**Midsize City (n=72)**	**Suburban/ Small City (n=739)**	**Suburban/ Rural (n=395)**	**Rural (n=141)**
	M (SD), Min-Max	*M (SD), Min-Max*	*M (SD), Min-Max*	*M (SD), Min-Max*	*M (SD), Min-Max*
SBA: Topical Vocabulary (TV) Score (max score = 46)	24.55 (8.13), 0-43	20.10 (12.05), 0-41	24.18 (7.76), 0-43	25.94 (7.35), 5-41	24.91 (8.69), 0-41
SBA: Vocabulary-in-Context (VIC) Score (max score = 5)	2.31 (1.40), 0-5	1.54 (1.42), 0-5	2.22 (1.39), 0-5	2.60 (1.34), 0-5	2.38 (1.37), 0-5
SBA: Use of IDK Overall (max = 51)	13.89 (10.71), 0-51	17.14 (18.21), 0-51	13.71 (10.34), 0-47	13.17 (9.32), 0-46	15.16 (10.85), 0-47
SBA: Use of IDK in TV Items (max = 46)	12.83 (10.10), 0-46	15.58 (17.16), 0-46	12.65 (9.74), 0-46	12.25 (8.85), 0-41	14.03 (10.19), 0-46
SBA: Use of IDK in VIC Items (max = 5)	1.05, (1.30), 0-5	1.56 (1.96), 0-5	1.06 (1.33), 0-5	.93 (1.05), 0-5	1.13 (1.33), 0-5
SBA Total Score (max score = 30)	12.17 (6.47), 0-29	8.22 (4.26), 2-24	11.85 (6.30), 0-29	13.26 (6.99), 0-29	12.78 (5.87), 3-28
Self-report Topic Knowledge	1.95 (0.87), 1-4	1.94 (0.93), 1-4	1.96 (0.84), 1-4	1.96 (0.89), 1-4	1.90 (0.88), 1-4
Self-report Sources of Knowledge	3.51 (3.24), 0-13	3.92 (3.89), 0-13	3.58 (3.22), 0-13	3.31 (3.08), 0-13	3.56 (3.42), 0-13
Funds of Knowledge for Organic Farming	2.72 (1.01), 1-5	2.46 (1.15), 1-5	2.70 (1.00), 1-5	2.83 (0.99), 1-5	2.69 (1.01), 1-4.88
Prior Learning	2.62 (1.15), 1-5	2.34 (1.22), 1-5	2.60 (1.13), 1-5	2.71 (1.16), 1-5	2.65 (1.12), 1-5
Local Relevance	2.83 (1.08), 1-5	2.54 (1.21), 1-5	2.81 (1.07), 1-5	2.95 (1.05), 1-5	2.76 (1.08), 1-5
Use of IDK in Prior Learning	0.09 (0.24), 0-1	0.06 (0.19), 0-1	0.08 (0.24), 0-1	0.11 (0.27), 0-1	0.08 (0.22), 0-1
Use of IDK in Local Relevance	0.18 (0.31), 0-1	0.12 (0.26), 0-1	0.18 (0.31), 0-1	0.21 (0.33), 0-1	0.17 (0.28), 0-1
Funds of Knowledge for Organic Food and Gardening	3.10 (0.83), 1-5	2.81 (1.11),1-4.92	3.10 (0.85), 1-5	3.15 (0.71), 1-5	3.06 (0.85), 1-4.67
Organic Food Access and Value	3.39 (1.00), 1-5	2.90 (1.30), 1-5	3.40 (1.01), 1-5	3.48 (0.90), 1-5	3.40 (1.00), 1-5
Food Preparation and Gardening	2.91 (0.90), 1-5	2.74 (1.08), 1-4.86	2.92 (0.92), 1-5	2.94 (0.80), 1-5	2.88 (0.97), 1-4.67
Use of IDK in Organic Food Access/Value	0.20 (0.31), 0-1	0.14 (0.28), 0-1	0.18 (0.30), 0-1	0.22 (0.32), 0-1	0.22 (0.32), 0-1
Use of IDK in Food Prep/Gardening	0.13 (0.23), 0-1	0.12 (0.26), 0-1	0.12 (0.23), 0-1	0.15 (0.25), 0-1	0.13 (0.21), 0-1

SBA scores showed statistically significant correlations with all demographic variables included in the study, suggesting differences in performance as a function of grade level (ϱ = .066), gender (ϱ = .085), and race/ethnicity (ϱ = .226). Post-hoc tests comparing average SBA scores indicated that grade 8 students outperformed those in grades 6 and 7 (both p's < .05), female students outperformed male students (p < .001), and students who were Asian American or White outperformed Black/African American and Hispanic/Latino students (all p's < .05). Free/reduced price lunch participation (ϱ = -.234), special education status (ϱ = -.190), and English learner status (ϱ = -.109) were all negatively correlated with SBA scores. Follow-up t-tests confirmed that lower SBA scores were obtained by those who participated in free/reduced lunch, received Title I accommodations during state assessments, were designated as special education students, or were designated as (current or former) English learners (all p's < .001). Students' academic performance was moderately correlated with SBA scores, based on current ELA grades (ϱ = .449) and state standardized test scores in both ELA (ϱ = .450) and science (ϱ = .460).

SBA scores also showed statistically significant correlations with all topical vocabulary knowledge measures used in this study. As expected, scores on topical vocabulary knowledge items were significantly and moderately correlated with SBA total scores, with VIC items (ϱ = .476) showing somewhat stronger correlations than TV items (ϱ = .386), and VIC items showing consistently significant relationships across all schools. Overall use of the IDK response option in the topical vocabulary knowledge measures showed small but significant negative correlations with SBA scores (ϱ = -.064), with a slightly stronger correlation for VIC items (ϱ = -.085) than TV items (ϱ = -.054). Use of IDK in TV items only showed significant relationships for the suburban/rural school site, and non-significant trends in VIC items for the two suburban schools and the rural school. Notably, SBA scores and engagement in the SBA (based on the 30% RTE threshold) were strongly correlated (ϱ = .762), across schools.

Interestingly, only some of the self-report knowledge and experience measures showed significant correlations with SBA scores. Although self-report topic knowledge was not significantly correlated with SBA scores, a non-significant trend was observed for the full sample (p = .064). Sources of knowledge (i.e., count of sources of knowledge about organic farming) showed a small but statistically significant negative correlation (ϱ = -.058), driven by a non-significant trend for the suburban/rural school (p = .066). The two subscales for funds of knowledge for organic farming showed near-zero correlations and thus were not significantly correlated with SBA scores; however, students' use of the IDK option in the local relevance subscale showed a small but significant positive correlation with SBA performance (ϱ = .103), driven by significant correlations for the two suburban schools. Among the funds of knowledge for organic food and gardening measures, we observed statistically significant, positive correlations for organic food access and value (ϱ = .214) and use of the IDK option in this subscale (ϱ = .155); significant correlations (or non-significant trends) were observed for all schools except the rural school. Food preparation and gardening experience showed a near-zero correlation and thus was not correlated with SBA scores, while use of the IDK option on this subscale showed a small but significant correlation with performance (ϱ = .116), driven by significant correlations for the two suburban schools.

RQ1. Do measures of self-reported prior knowledge and experiences predict reading comprehension as measured by SBA scores, over and above topical vocabulary measures?

To answer this question, we examined a series of hierarchical regression models with SBA score as the outcome (see Appendix Table A1). Results for a baseline model with only student demographics (Model 0) indicated that these characteristics accounted for 12.7% variance in SBA scores, with grade level (β = .068) and gender (β = .070) as statistically significant positive predictors, and with Black/African American race/ethnicity (β = -.164), free/reduced lunch participation (β = -.171), special education status (β = -.101), and English learner status (β = -.066) as significant negative predictors of SBA scores. Model 1a included embedded topical vocabulary knowledge measures, which explained 39.6% of the variance in SBA scores

Table 5. Spearman Correlations with SBA Score

Variable	**Overall Sample (N=1,347)**		**Midsize City (n=72)**		**Suburban/ Small City (n=739)**		**Suburban/ Rural (n=395)**		**Rural (n=141)**	
	ϱ	*p*	*ϱ*	*p*	*ϱ*	*p*	*ϱ*	*p*	*ϱ*	*p*
Grade Level	**.066**	**.016**	-.082	.492	.061	.101	**.114**	**.024**	.080	.350
Gender	**.085**	**.002**	.116	.331	.063	.088	.060	.236	.112	.193
Race/Ethnicity	**.226**	**.001**	-.147	.217	**.217**	**< .001**	**.200**	**< .001**	.001	.989
Free/Reduced Price Lunch Participation: Yes	**-.234**	**< .001**	[a]		**-.248**	**< .001**	**-.230**	**< .001**	[a]	
Title I Accommodation: Yes	**-.118**	**< .001**	.049	.685	-.014	.709	[a]		***-.161***	***.059***
Special Education Status: Yes	**-.190**	**< .001**	.002	.989	**-.122**	**< .001**	**-.165**	**< .001**	**-.169**	**.048**
English Learner Status: Yes	**-.109**	**< .001**	-.126	.290	**-.093**	**.012**	-.067	.186	[a]	
Current ELA Grade	**.449**	**< .001**	.186	.118	**.447**	**< .001**	**.415**	**< .001**	**.298**	**< .001**
ELA State test score (Z-score)	**.450**	**< .001**	.197	.116	**.440**	**< .001**	**.542**	**< .001**	**.525**	**< .001**
Science State test score (Z-score)	**.460**	**< .001**	-.169	.409	**.446**	**< .001**	**.546**	**< .001**	**.483**	**< .001**
SBA: Topical Vocabulary (TV) Score (max score = 46)	**.386**	**< .001**	.142	.235	**.369**	**< .001**	**.440**	**< .001**	**.338**	**< .001**
SBA: Vocabulary-in-Context (VIC) Score (max score = 5)	**.476**	**< .001**	**.304**	**.010**	**.473**	**< .001**	**.465**	**< .001**	**.420**	**< .001**
SBA: Use of IDK Overall	**-.064**	**.020**	.086	.475	-.056	.128	**-.116**	**.022**	-.108	.203
SBA: Use of IDK in TV Items	**-.054**	**.048**	.113	.345	-.048	.192	**-.110**	**.030**	-.087	.307
SBA: Use of IDK in VIC Items	**-.085**	**.002**	-.119	.320	***-.069***	***.061***	***-.089***	***.080***	***-.140***	***.098***

Engagement in SBA (30% RTE)	**.762**	**< .001**	**.509**	**< .001**	**.756**	**< .001**	**.773**	**< .001**	**.701**	**< .001**
Self-report Topic Knowledge	***.051***	***.064***	.042	.727	.058	.119	.034	.508	.101	.234
Self-report Sources of Knowledge	**-.058**	**.033**	.084	.485	-.046	.215	***-.093***	***.066***	-.035	.683
Funds of Knowledge for Organic Farming	.008	.769	***.213***	***.076***	-.029	.448	.002	.964	-.030	.734
Prior Learning	.028	.311	***.202***	***.093***	.023	.543	-.026	.625	.009	.916
Local Relevance	-.006	.845	***.227***	***.061***	-.057	.139	.022	.685	-.068	.437
Use of IDK in Prior Learning	.004	.879	-.016	.896	-.021	.570	-.017	.736	.079	.354
Use of IDK in Local Relevance	**.103**	**< .001**	.127	.289	**.076**	**.040**	**.138**	**.006**	.025	.767
Funds of Knowledge for Organic Food and Gardening	**.086**	**.002**	***.210***	***.083***	***.067***	***.075***	**.107**	**.038**	-.008	.927
Organic Food Access and Value	**.214**	**< .001**	**.261**	**.031**	**.195**	**< .001**	**.36**	**< .001**	.135	.127
Food Preparation and Gardening	-.002	.937	.145	.239	-.018	.640	-.019	.723	.013	.882
Use of IDK in Organic Food Access/Value	**.155**	**< .001**	***.212***	***.074***	**.112**	**.002**	**.198**	**< .001**	.129	.128
Use of IDK in Food Preparation/Gardening	**.116**	**< .001**	.038	.750	**.079**	**.032**	**.162**	**.001**	.014	.873

Note. Cells in boldface reflect statistically significant effects. Cells in bold-italics reflect effects with *p* values between .05 and .10.
[a] Could not be computed due to lack of within-cell variance

after controlling for student demographics ($\Delta R^2 = .269$, $p < .001$). With the addition of these topical vocabulary knowledge variables ($\beta_{TV} = .645$, $\beta_{VIC} = .327$), students' grade level, special education status, and English learner status were no longer significant.

Model 1b evaluated whether the self-report knowledge and experience measures explained additional variance beyond student demographics. This model explained 18.6% of variance in SBA scores after controlling for student demographics ($\Delta R^2 = .059$, $p < .001$). In this model, students' funds of knowledge for organic food access and value was a significant positive predictor of SBA scores ($\beta = .219$), while experience with food preparation and gardening ($\beta = -.122$) and local relevance of organic farming ($\beta = -.087$) were significant negative predictors.

Model 2 incorporated self-reported knowledge variables, after controlling for student demographics and topical knowledge measures from the SBA. The inclusion of self-reported knowledge variables in this model slightly but significantly improved the model fit over Model 1a by about 2%, increasing the explained variance to 41.5% ($\Delta R^2 = .019$, $p < .001$). In the full model addressing RQ1, all topical knowledge variables remained statistically significant predictors of SBA scores ($\beta_{TV} = .600$, $\beta_{VIC} = .318$). In addition, students' funds of knowledge for organic food access and value remained a positive predictor ($\beta = .121$), and funds of knowledge for food preparation and gardening remained a significant negative predictor ($\beta = -.072$). Thus, findings indicate that self-reported funds of knowledge contribute a modest but statistically significant additional amount of explanatory power beyond topical vocabulary knowledge measures.

RQ2. To what extent do measures of students' prior knowledge and experiences predict task engagement in the SBA?

Regression models with engagement as the outcome examined the degree to which prior knowledge and experience measures were significant predictors of engagement in the SBA, as measured by RTE. In Appendix Table A2, Model 0 indicated that student demographics explained 19.6% of the variance in engagement. Among the demographic variables, female students ($\beta = .082$) and students attending the rural school ($\beta = .070$) showed higher engagement compared to males and students in the suburban/small city school (reference groups), while students located in the midsize city school ($\beta = -.247$), those of Black/African American race/ethnicity ($\beta = -.109$), those participating in free/reduced lunch ($\beta = -.140$), and those with special education ($\beta = -.102$) or English learner status ($\beta = -.072$) showed significantly lower engagement compared to respective reference groups.

When SBA topical vocabulary knowledge measures were added in Model 1a, controlling for student background characteristics, these variables were statistically significant predictors ($\beta_{TV} = .473$, $\beta_{VIC} = .264$) and the model fit improved to 35.3% variance explained ($\Delta R^2 = .157$, $p < .001$). Effects for female students, midsize city school students, and free/reduced lunch participation remained significant; a statistically significant negative effect also emerged for grade level ($\beta = -.071$) such that students in lower grades were more engaged than those in higher grades.

Model 1b evaluated whether the self-report knowledge and experience measures added any predictive value for engagement over and above student demographics. This model explained 24.2% of variance in engagement ($\Delta R^2 = .045$, $p < .001$). Funds of knowledge for organic food access and value was a significant positive predictor of SBA engagement in this model ($\beta = .199$), while funds of knowledge for food preparation and gardening was a significant negative predictor ($\beta = -.069$).

Model 2 combined the self-report knowledge measures with the topical vocabulary knowledge measures as predictors of SBA engagement, controlling for student demographics. This full model explained 37.0% of variance in students' SBA engagement, showing a small but statistically significant improvement in model fit compared to Model 1a ($\Delta R^2 = .018$, $p = .001$). Based on this model, funds of knowledge for organic food access and value significantly predicted engagement ($\beta = .126$) after controlling for topical vocabulary

knowledge measures (β_{TV} = .429, β_{VIC} = .255) and student demographics. Gender remained a positive predictor of engagement (β = .073), while attending the midsize city school (β = -.245), grade level (β = -.088), and free/reduced lunch participation (β = -.062) remained negative predictors. Thus, this analysis indicates that both self-report knowledge measures and topical vocabulary knowledge measures embedded in the SBA explain additional variance in students' engagement during the SBA, as measured by RTE.

RQ3. To what extent does task engagement predict reading comprehension, over and above measures of students' prior knowledge and experiences?

In subsequent models, we investigated whether engagement predicted reading comprehension over and above measures of students' prior knowledge and experiences. As shown in Appendix Table A1, Model 2 indicates that student backgrounds and prior knowledge and experiences explained 41.5% of the variance in SBA scores. Appendix Table A3 shows the results for Model 3 that entered SBA engagement (measured by RTE) as an additional predictor of SBA scores, controlling for student background, topical vocabulary knowledge, and self-report knowledge and experience measures. This model explained 59.9% of variance in SBA scores, a significant improvement in model fit beyond Table A1 Model 2 (ΔR^2 = .184, $p < .001$). Engagement was the strongest predictor of SBA scores (β = .541), indicating that students' behavioral engagement during the task significantly contributed to performance, over and above student demographics and knowledge measures. This is perhaps not surprising given the strong correlations observed between RTE and SBA scores (see Table 5). With engagement included in the model, students in the midsize city (β = .116) showed stronger SBA performance compared to suburban/small city school students (reference group). Performance also increased with students' grade level (β = .045). Students' topical vocabulary knowledge scores remained significant predictors of SBA scores (β_{TV} = .368, β_{VIC} = .180). Students' self-reports of the number of sources where they had learned about organic farming (β = .052) emerged as a significant predictor. Students' funds of knowledge for access to and value for organic food (β = .053) and food preparation and gardening (β = -.055) remained significant predictors after engagement was added into the model. Thus, behavioral engagement during the assessment also emerged as a strong predictor of SBA performance, controlling for demographics and knowledge measures.

RQ4. Does students' use of the IDK option within the prior knowledge measures predict reading comprehension or task engagement?

We estimated effects of students' reliance on the IDK option in topical vocabulary knowledge measures and in self-report survey measures on their reading comprehension and their engagement in the SBA. In models with SBA scores as the outcome (Appendix Table A1), use of IDK within both types of topical vocabulary knowledge measure was a statistically significant predictor of SBA scores (Model 1a: β_{TV} = .458, β_{VIC} = .095); Model 1b indicated that use of IDK in self-reports of local relevance (β = .073) and organic farming access and value (β = .084) were positive predictors, while use of IDK in prior learning items was a negative predictor (β = -.087). With all knowledge measures entered as predictors (Model 2), use of IDK in topical vocabulary knowledge measures remained significant (β_{TV} = .429, β_{VIC} = .088), but use of IDK in the self-report measures did not significantly predict SBA scores (*p*'s > .05). Students who were more likely to use IDK in the topical vocabulary knowledge measures tended to have higher SBA scores, after controlling for demographics and scores on all knowledge measures. Interestingly, when engagement was entered as a predictor of SBA scores (Appendix Table A3), the effects of IDK use on topical vocabulary knowledge measures remained significant for TV items, but were no longer statistically significant for VIC items, suggesting that when engagement captured a large amount of variance in comprehension performance, use of the IDK option in the VIC items no longer contributed to predict SBA scores.

In models with engagement as the outcome (Appendix Table A2), use of IDK in the topical vocabulary knowledge measures was also a statistically significant predictor of engagement as measured by RTE (Model 1a: β_{TV} = .331, β_{VIC} = .095). Use of IDK in the organic food access and value subscale was a significant

predictor (Model 1b: β = .092) of engagement. In the full model with all knowledge measures entered as predictors (Model 2), use of IDK in topical vocabulary knowledge measures remained significant (β_{TV} = .297, β_{VIC} = .093) as did use of IDK in self-reports of organic food access and value (β = .063). Students who were more likely to use IDK in these measures were more likely to demonstrate higher behavioral engagement during the SBA, controlling for demographics and scores on all knowledge measures. Thus, results from both reading comprehension and engagement models highlighted the predictive role of IDK use, although the magnitude of IDK effects was generally smaller compared to the size of the effects of performance on the respective scales. This suggests that measures of metacognition (as evidenced by use of the IDK option) predict SBA performance and engagement, but to a lesser extent than do measures of prior knowledge.

RQ5. Do students' perceptions of the IDK option within the prior knowledge measures interact with the relationship among use of IDK, performance, and engagement?

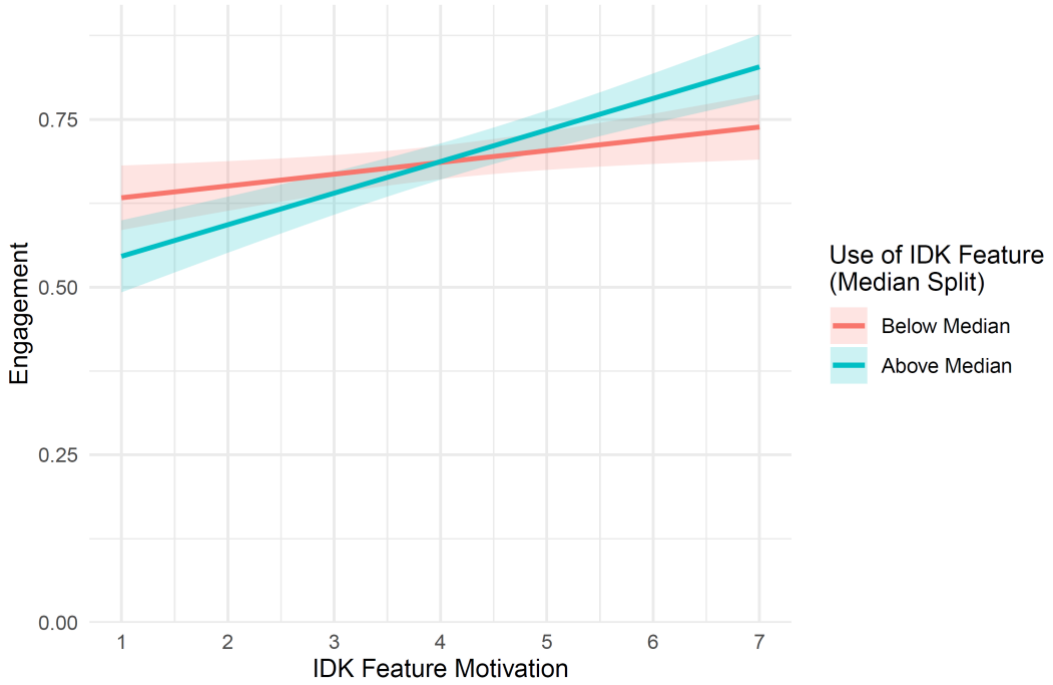
The extent to which motivation moderated IDK use was explored through additional regression models incorporating students' post-test ratings of the degree to which the IDK option in the topical vocabulary knowledge items affected their performance and motivation to complete the SBA, as well as interaction terms between this post-test rating and students' actual use of the feature during these items. Appendix Table A4 summarizes the results of these regression models. Results of Model 4 show that students' perceptions of the degree to which the IDK feature helped them do better on the SBA was a small but statistically significant predictor of SBA scores (β = .058), controlling for student background characteristics, all knowledge measures, and engagement ($\Delta R^2 < .001$, $p = .638$ vs. Model 3, Table A3). This model explained 60.2% of the variance in SBA scores. Following the socioculturally responsive assessment principle of characterizing assessment results as an interaction (Bennett, 2023), we also examined whether students' perceptions of the IDK feature interacted with their use of the feature. When interaction terms were added into the model (Model 5); the results were largely identical to Model 4, with a slightly larger effect for the post-test rating (β = .067), and no significant interaction terms (60.3% variance explained; $\Delta R^2 = .001$, $p = .683$ vs. Model 4). In general, students who gave higher ratings as to whether the IDK feature supported their SBA performance tended to score higher on the SBA.

In parallel models predicting engagement (see Appendix Table A5), Model 4 indicates that students' perceptions of the degree to which the IDK feature helped motivate them to complete the SBA was a small but statistically significant predictor of SBA engagement (β = .088), controlling for student background and all knowledge measures (37.7% variance explained; $\Delta R^2 = .007$, $p < .001$ vs. Model 2, Table A2). In general, students who gave higher ratings as to whether the IDK option was motivating showed higher engagement during the SBA. Interaction terms were entered in Model 5 (Table A5), and a significant positive interaction between students' use of IDK in TV Items and their ratings of the degree to which the IDK feature helped motivate them to complete the SBA was observed (β = .163), although this model did not explain a significant amount of additional variance (38.1% variance explained; $\Delta R^2 = .003$, $p = .065$ vs. Model 4, *n.s.*).

A post-hoc analysis of this interaction effect indicates that the relationship between students' perceptions of the IDK feature and their SBA engagement varied as a function of their use of the IDK option in the 46-item TV measure (see Figure 1). Specifically, those students who were more likely to use the IDK option in TV items (above-median) showed a stronger relationship between their perceptions of how motivating the IDK option was and their actual test-taking engagement; those who less often used the IDK option in the TV items (below-median) showed more consistent levels of engagement regardless of how they rated the feature. This interaction indicates that greater use of IDK in the topical vocabulary knowledge items enhances the motivational effect of the IDK option on students' SBA engagement—highlighting that enabling students to express uncertainty may enhance perceived motivation and observed behavioral

engagement compared to students who are less likely to express uncertainty. It is notable that this interaction effect was only observed for engagement, and not SBA performance.

Figure 1. Plot showing interaction between students' perceptions of how motivating the IDK feature is and students' use of the IDK option in the 46-item topical vocabulary items on SBA engagement (proportion engaged).



RQ6. To what extent do students' levels of prior knowledge and experience interact with one another or with other factors (i.e., demographics, engagement) to predict SBA performance?

Additional analyses were conducted to investigate interactions between topical vocabulary knowledge and self-reported funds of knowledge and experience measures, and potential interactions between knowledge variables and other factors, namely demographic variables and test-taking engagement. Preliminary analyses investigated the interactions between TV item scores and the self-report subscales for organic food access and value and for food preparation and gardening experience. These analyses indicated that both interaction terms displayed non-significant trends ($\beta_{TV \times Access/Value} = .065$, $p = .065$; $\beta_{TV \times Food\ Prep/Gardening} = -.061$, $p = .058$), suggesting a potentially compounding effect of topical vocabulary knowledge and lived experience on SBA performance. To further explore these relationships, we created subgroups using median splits for each variable and examined subgroup means. Students scoring above the median (>= 54.3% correct) on TV items ($M = 66.7\%$ correct) showed significantly higher access to and value for organic food ($t = 4.01$, $p < .001$, Cohen's $d = .23$) and significantly higher engagement in the SBA ($t = 9.73$, $p < .001$, Cohen's $d = .53$) compared to students scoring below the median ($M = 35.4\%$ correct); these subgroups did not differ in their food preparation and gardening experience ($t = 0.29$, $p = .77$, Cohen's $d = .02$).

Using the median splits for all variables, a chi-square test of independence indicated that there was a significant relationship between TV item scores and self-reported access to and value for organic food [$\chi^2(df = 1) = 11.01$, $p < .001$], but there was no such relationship for experience with food preparation and gardening [$\chi^2(df = 1) = 0.54$, $p = .46$]. Further examination of the relationship between topical vocabulary scores and organic food access and value indicated that students with high topical vocabulary knowledge

and high access/value demonstrated the highest SBA scores (M = 49.9% correct) and engagement (M = .81), and students with high topic knowledge but limited experience had the second highest levels of SBA performance (M = 43.1% correct) and engagement (M = .72). Among students with low topical vocabulary knowledge, those reporting higher organic food access had higher SBA performance (M = 36.2% correct) and engagement (M = .68) than those with minimal relevant prior experiences (M = 30.2% correct, M = .55, respectively). Not only do these results underscore the important role of topical vocabulary knowledge for comprehension, but they also suggest potential compensatory effects of prior experience with and value for organic foods, in the relative absence of topical vocabulary knowledge. Students having both high topical vocabulary knowledge and experience demonstrated the strongest performance, but students having either high topical vocabulary knowledge or prior experience showed benefits over students relatively lacking in relevant knowledge and experiences.

More formally, we tested a series of interaction effects by incorporating them into models predicting SBA scores, building from Model 3 (Table A3) which included engagement as a predictor in addition to student demographics and all knowledge variables. We only considered interaction terms for variables that were demonstrated to be significant predictors of SBA score in previous models (Table A3, Model 3). Specifically, we included interaction terms crossing both types of topical vocabulary knowledge measure (TV, VIC) with organic food access/value and food preparation and gardening experience subscales, as well as interaction terms crossing each of these four knowledge variables with students' school location, grade level, and engagement in the SBA (Table A4, Model 6). With these 24 interaction terms added into the model, 66.3% of variance in SBA scores was explained (ΔR^2 = .064, p < .001 vs. Model 3, Table A3).

The baseline interactions between topic knowledge and lived experience measures were not statistically significant. However, we observed significant interaction effects between students' school location and their funds of knowledge for both organic food access and value and food preparation and gardening for students in the midsize city school ($\beta_{Access/Value}$ = .214; $\beta_{Food\ Prep/Gardening}$ = -.159) and the rural school ($\beta_{Access/Value}$ = -.234; $\beta_{Food\ Prep/Gardening}$ = .159). Figure 2 illustrates the nature of this interaction for organic food access and value, while Figure 3 illustrates the nature of the interaction for food preparation and gardening experience. Relationships between organic food access and value and SBA scores were generally positive, with rural students showing a weaker (flatter) relationship compared to students in other schools; a similar pattern can be observed for the relationship between food preparation and gardening experience and SBA scores (with a negative relationship observed for the two suburban school sites). Thus, students' self-reported experiences with organic food and food preparation and gardening appeared to vary depending on their geographic location (urban vs. rural). Midsize city school students' lived experience with respect to organic foods and food preparation and gardening seemed to have a larger impact on performance than the experiences of the rural school students.

We also observed significant interaction effects between students' levels of prior knowledge and lived experiences and their engagement in the SBA. Students' engagement on the SBA showed interactions with their scores on TV items (β = .461), VIC items (β = .634), and self-reports of organic food access and value (β = .266). Figure 4 illustrates this interaction effect for TV scores, Figure 5 illustrates this interaction for VIC scores, and Figure 6 illustrates this interaction for self-reported access to and value of organic foods. Students who scored above the median on both topical vocabulary measures and were engaged on more than 50% of the SBA items tended to have higher SBA scores than students scoring below the median on the topic vocabulary measures. However, at low levels of engagement, those with lower prior knowledge tended to show slightly higher SBA scores than those with above median knowledge but low engagement. Put differently, it appears that there may be thresholds for minimum levels of knowledge and engagement that enable students to perform well on the SBA. Effects for the self-report measure show a different pattern; it appears that students' self-reports of organic food access and value have a larger effect on SBA scores at lower levels of engagement (i.e., below 50% engagement, students below median access and value score

slightly higher), while at higher levels of engagement there was a minimal difference in scores as a function of access and value.

Figure 2. Plot showing the interaction effect between students' self-reports of access to and value for organic foods and their school location on SBA scores (percent correct).

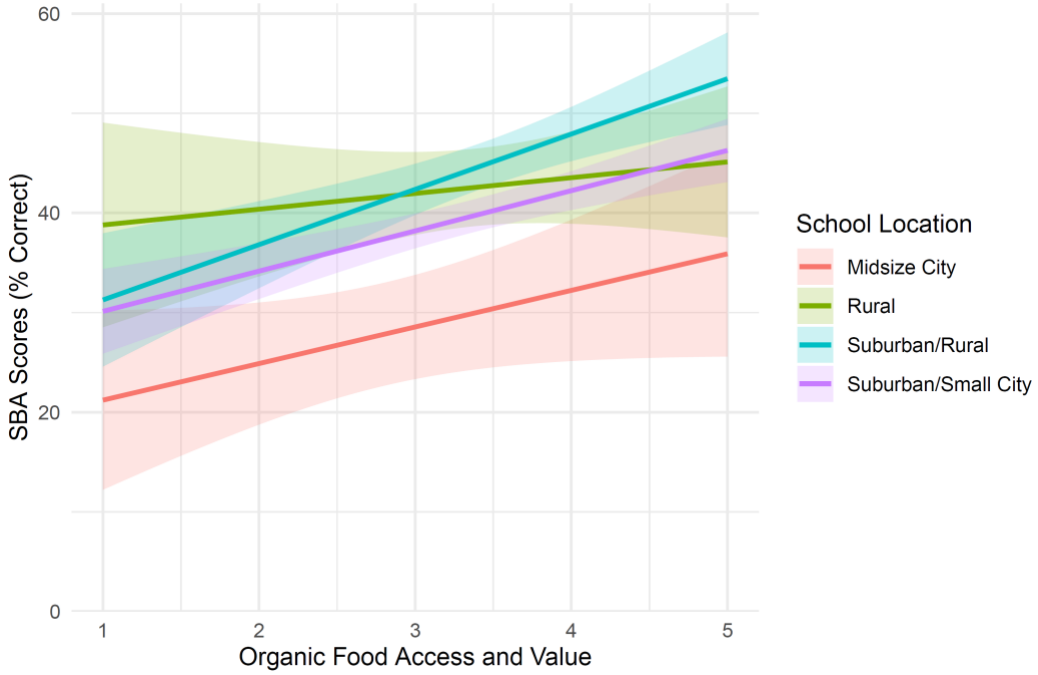


Figure 3. Plot showing the interaction effect between students' self-reported experiences with food preparation and gardening and their school location on SBA scores (percent correct).

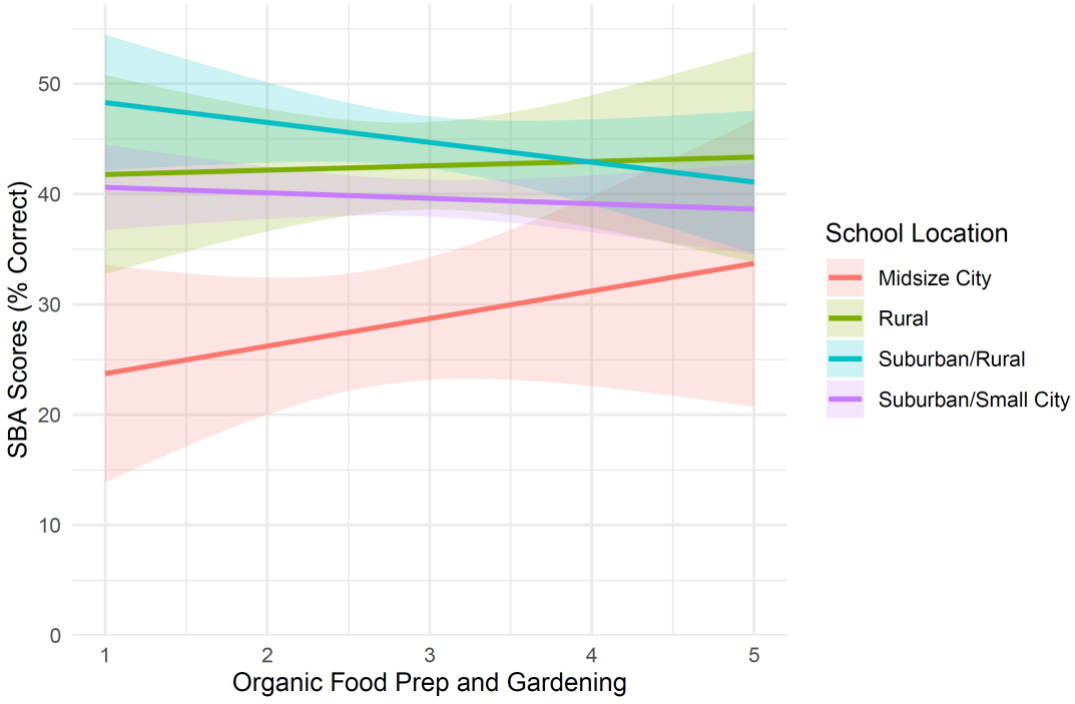


Figure 4. Plot showing the interaction effect between students' engagement (percent engaged) and topical vocabulary scores (median-split) on SBA scores (percent correct).

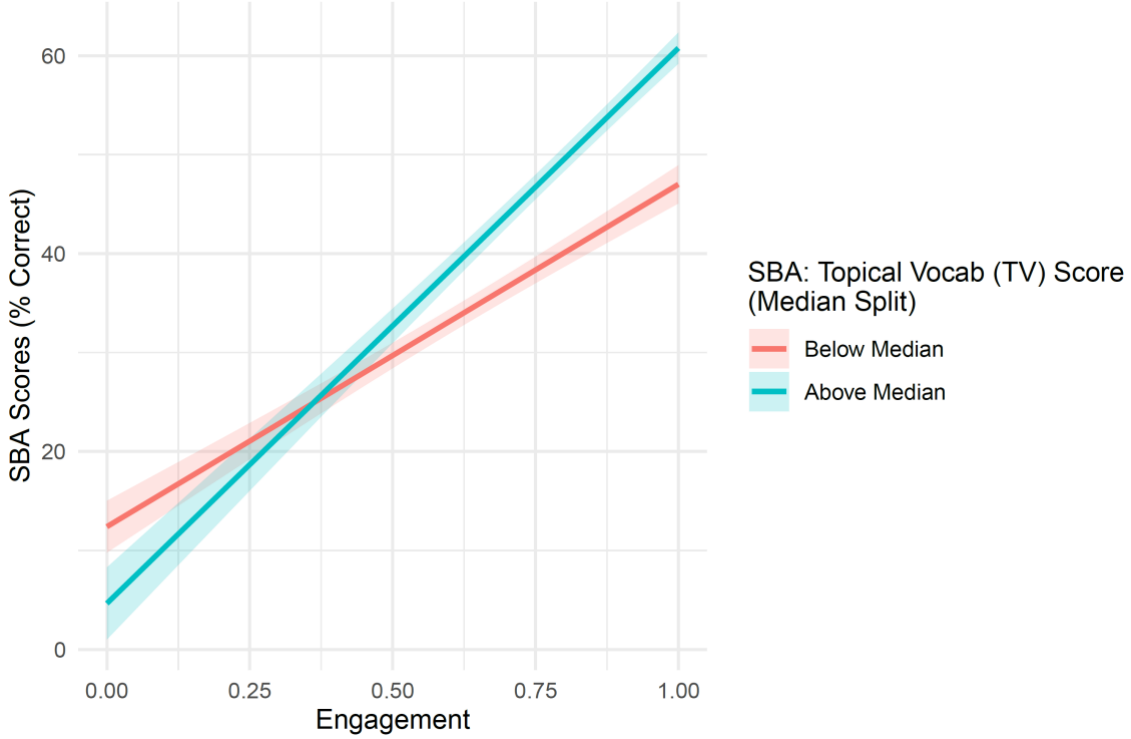


Figure 5. Plot showing the interaction effect between students' engagement (percent engaged) and vocabulary-in-context scores (median-split) on SBA scores (percent correct).

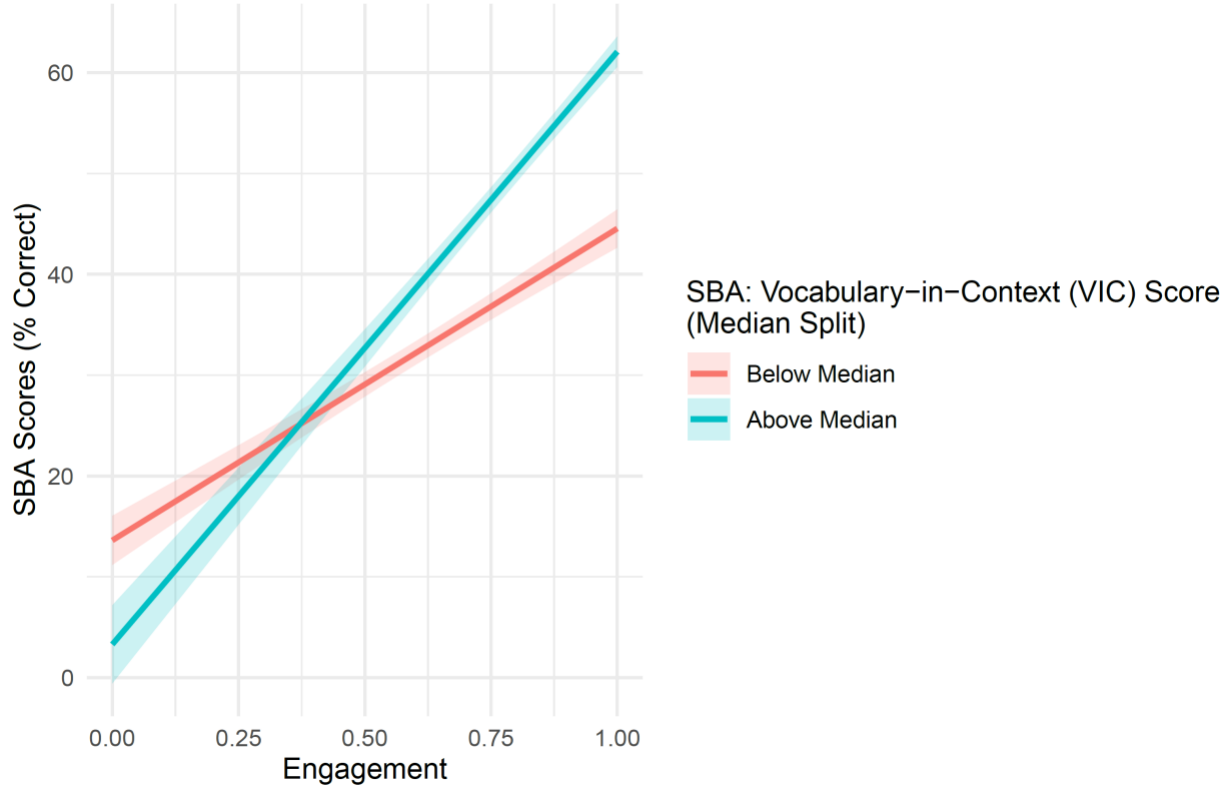
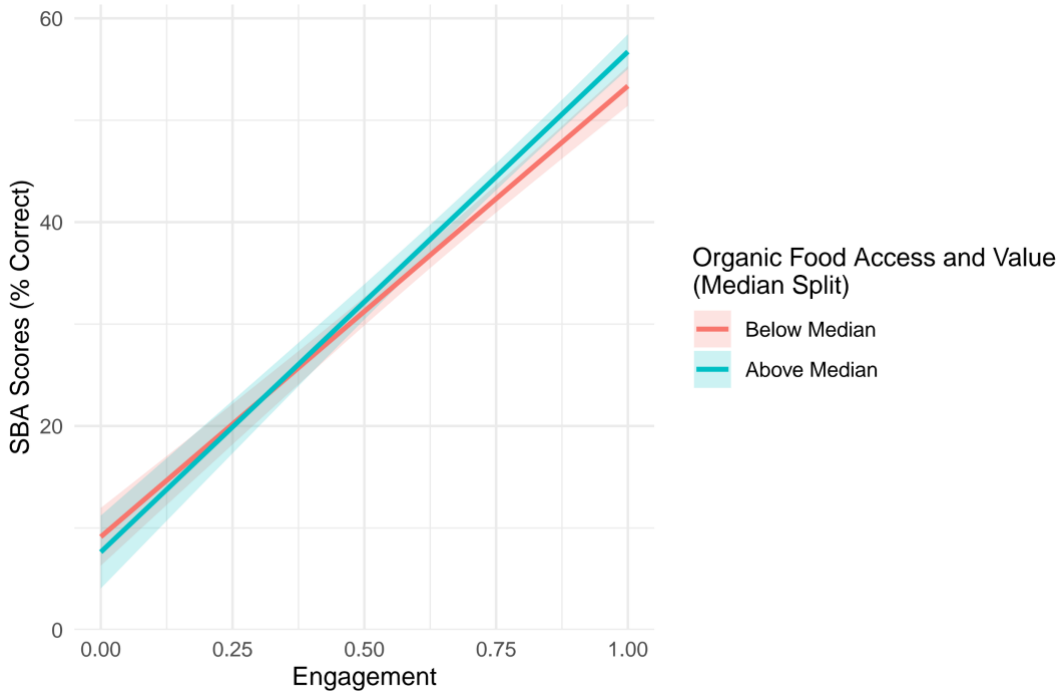


Figure 6. Plot showing the interaction effect between students' engagement (percent engaged) and self-reports of organic food access and value (median-split) on SBA scores (percent correct).



Discussion

This study examined the effects of multiple dimensions of students' prior knowledge, assessed through a combination of topical vocabulary knowledge measures and self-reports of prior knowledge and lived experiences, on students' performance on and engagement during a reading comprehension SBA. Results indicated that prior knowledge of topical vocabulary was strongly associated with SBA performance and engagement, while a slightly different mix of self-reported funds of knowledge measures were found to predict performance or engagement, depending on model specifications, explaining small but statistically significant additional variance beyond topical vocabulary knowledge measures. Overall, students' experiences with organic food significantly predicted both performance and task engagement. Students' use of the IDK option on this measure uniquely predicted engagement, while experiences with food preparation and gardening negatively predicted performance but not engagement. The findings offer implications for the evidence-based design of personalized, equitable, and socioculturally responsive assessments, as well as several future directions for research and development of such assessments. We summarize the key findings aligned to each research question (see Table 6) and then discuss implications and future directions in turn.

Summary of Key Findings

RQ1 considered whether the self-report knowledge and experience measures we developed demonstrated additional predictive value for SBA performance, beyond topical vocabulary knowledge measures that have previously been demonstrated to be effective in explaining variance in SBA scores. We observed that after controlling for topical knowledge measures, which were strong predictors of performance, students' funds of knowledge in terms of their experiences with and value for organic foods positively predicted SBA scores, while students' experiences with food preparation and gardening negatively predicted scores. The self-report instruments explained a small but significant additional amount of variance in SBA scores.

RQ2 considered whether the self-report knowledge measures added predictive value for SBA engagement, as measured by RTE, beyond topical vocabulary knowledge measures. We found that after controlling for topical vocabulary knowledge measures, which were also strong predictors of engagement, funds of knowledge in terms of students' experiences with and value for organic foods also positively predicted SBA engagement. The self-report instruments again explained a small but significant additional amount of variance in engagement.

RQ3 investigated whether engagement explained additional variance in SBA scores over and above the topical vocabulary-based and self-report knowledge measures. Results indicated that engagement (i.e., RTE) was a stronger predictor of performance than any of the knowledge measures. Effects for the topical vocabulary knowledge measures were largely consistent with the addition of engagement (except for use of the IDK option in VIC items, which was no longer a significant predictor). Similarly, effects for self-report knowledge measures were largely consistent, with funds of knowledge for access to and value for organic foods as a small but significant positive predictor, and funds of knowledge for food preparation and gardening as a small but significant negative predictor. In addition, once RTE was taken into account, the total number of sources from which students reported having learned about organic farming emerged as a small but significant positive predictor.

RQ4 examined effects of reliance on the IDK option in various measures on students' performance and engagement. Even after RTE was entered into the model, use of this option in topical vocabulary knowledge measures showed significant effects for both performance and engagement, remaining a significant predictor of SBA scores. Use of the IDK option in self-report funds of knowledge measures showed differential effects depending on model specifications. Use of IDK in self-report measures did not predict SBA scores after controlling for student demographics and topical vocabulary knowledge measures, and remained non-significant after RTE was entered into the model. Models predicting engagement, however, indicated that students' reliance on the IDK option in reporting their access to and value for organic foods was found to be a significant predictor of RTE, controlling for demographics and topical vocabulary knowledge measures. Thus, incorporating use of IDK on the self-report measures in the models explained a small but statistically significant amount of additional variance in engagement.

Given that prior results indicated significant effects of IDK use on both SBA scores and RTE, RQ5 further examined the role of students' perceptions of the degree to which this SBA design feature supported their SBA performance and motivated them to complete the SBA. Students' rating of performance support for the IDK option was found to be a significant predictor of SBA scores, controlling for student demographics, all knowledge measures, and engagement. The more positively students perceived this feature, the better they tended to perform on the assessment. Similarly, students' ratings of motivation support for the IDK option were found to be a significant predictor of RTE, controlling for student background and all knowledge measures. The more positively students perceived the IDK feature, the more they tended to engage with the assessment. A significant interaction was also observed such that the relationship between students' perceptions of motivational support and their actual engagement during the task varied depending on how often they used the IDK option. Those who relied more heavily on this option while answering topical vocabulary knowledge items showed stronger relationships between RTE and perceived motivational support of the IDK option. No interaction was observed for SBA scores.

Finally, RQ6 examined interactions between knowledge and experience measures, and between these measures, students' demographics, and engagement. While preliminary analyses suggested a trend toward TV scores interacting with self-report funds of knowledge for organic food access and value and for food preparation and gardening, with a potential compensatory effect of prior experiences on SBA scores in cases of low topical vocabulary knowledge, a more formal test via regression analysis indicated no such

Table 6. Summary of Main Findings

RQ	Models	Predictors	Outcomes	Significant Predictors	Key Findings
RQ1	Table A1 Model 2	- Demographics - Topical Vocabulary Measures - Self-report Knowledge Measures	Comprehension (SBA Scores)	- Female (+) - Black or African American (-) - Free/Reduced Lunch (-) - TV Score (+) - VIC Score (+) - IDK in TV (+) - IDK in VIC (+) - Organic Food Access/Value (+) - Food Preparation & Gardening Experience (-)	Measures of self-reported prior knowledge and experiences predicted a modest amount of variance in reading comprehension scores, over and above topical vocabulary knowledge measures.
RQ2	Table A2 Model 2	- Demographics - Topical Vocabulary Measures - Self-report Knowledge Measures	Engagement (RTE)	- Midsize City School (-) - Female (+) - Grade Level (-) - Free/Reduced Lunch (-) - TV Score (+) - VIC Score (+) - IDK in TV (+) - IDK in VIC (+) - Organic Food Access/Value (+) - IDK in Organic Food Access/Value (+)	Self-report knowledge measures and topical vocabulary measures predicted SBA engagement, explaining a considerable amount of the variance in RTE.
RQ3	Table A3 Model 3	- Demographics - Topical Vocabulary Measures - Self-report Knowledge Measures	Comprehension (SBA Scores)	- Midsize City School (+) - Grade Level (+) - TV Score (+) - VIC Score (+) - IDK in TV (+)	Engagement (RTE) emerged as a strong predictor of comprehension (SBA Scores), after controlling for measures of topical vocabulary knowledge and

		- Engagement (RTE)		- Self-report Sources of Knowledge (+) - Organic Food Access/Value (+) - IDK in Organic Food Access/Value (+) - Food Preparation & Gardening Experience (-) - Engagement (+)	self-report knowledge and experiences.
RQ4	Table A1 Model 2	- Demographics - Topical Vocabulary Measures - Self-report Knowledge Measures	Comprehension (SBA Scores)	*See RQ1*	Use of IDK in topical vocabulary knowledge items predicted comprehension (SBA Scores).
	Table A2 Model 2	- Demographics - Topical Vocabulary Measures - Self-report Knowledge Measures	Engagement (RTE)	*See RQ2*	Use of IDK in topical vocabulary knowledge items and self-reports of organic food access and value predicted engagement (RTE).
RQ5	Table A4 Model 4 Table A4 Model 5	- Demographics - Topical Vocabulary Measures - Self-report Knowledge Measures - Engagement (RTE) - Post-test Ratings of IDK Feature - Interaction Term: Use of IDK X IDK Feature Rating	Comprehension (SBA Scores)	- Midsize City School (+) - Grade Level (+) - Free/Reduced Lunch (-) - TV Score (+) - VIC Score (+) - IDK in TV (+) - Organic Food Access/Value (+) - Food Preparation & Gardening Experience (-) - Engagement (+) - Post-test Rating: IDK Helped Performance (+)	Students' perceptions of the degree to which the IDK option helped their SBA performance predicted comprehension (SBA Scores) but did not interact with IDK use.

	Table A5 Model 4 Table A5 Model 5	- Demographics - Topical Vocabulary Measures - Self-report Knowledge Measures - Post-test Ratings of IDK Feature - Interaction Term: Use of IDK X IDK Feature Rating	Engagement (RTE)	- Midsize City School (-) - Female (+) - Grade Level (-) - Free/Reduced Lunch (-) - TV Score (+) - VIC Score (+) - IDK in TV (+) - IDK in VIC (+) - Organic Food Access/Value (+) - IDK in TV Items X IDK Helped Motivation (+)	Students' perceptions of the degree to which the IDK option helped motivate them to complete the SBA predicted engagement (RTE), showing a significant interaction with IDK use in TV items.
RQ6	Table A4 Model 6	- Demographics - Topical Vocabulary Measures - Self-report Knowledge Measures - Engagement (RTE) - Interaction Terms	Comprehension (SBA Scores)	- IDK in TV (+) - IDK in VIC (+) - Organic Food Access/Value X Midsize City School (+) - Organic Food Access/Value X Rural School (-) - Food Preparation & Gardening X Midsize City School (-) - Food Preparation & Gardening X Rural School (+) - TV Score X Engagement (+) - VIC Score X Engagement (+) - Organic Food Access/Value X Engagement (+)	Several interactions were observed among predictors of comprehension (SBA Scores). Self-report measures of organic food access and value and food preparation and gardening experience showed significant interactions with school location (city vs. rural), in opposite directions. Scores on topical vocabulary measures and self-reports of organic food access and value also showed significant interactions with engagement (RTE).

interactions. Instead, this analysis indicated that self-reported experiences with organic food and with food preparation and gardening had different effects on SBA scores depending on the students' geographic location (i.e., urban vs. rural schools). Topical vocabulary knowledge measures and self-reported organic food access and value also showed significant interactions with RTE, such that the knowledge measures had different relationships to SBA scores at different levels of engagement. No interactions were observed with students' grade levels, suggesting that these relationships may be relatively stable across age groups, or that effects of knowledge and engagement may matter more for students' SBA performance than do their grade levels.

It is notable that after controlling for demographics, topical vocabulary knowledge, and engagement, a positive interaction effect was observed between self-reported experiences with organic food and SBA scores for students in the midsize city school, while a negative interaction effect was observed for students in the rural school. The opposite pattern was observed for self-reported experiences with food preparation and gardening. Examination of students' responses to the self-report scales indicated disparities in self-reported access to and availability of organic food, with students from the rural school tending to agree that their family could access and afford organic food, while students from the midsize city school tended to strongly disagree or give neutral responses to all items in the subscale. Students from the midsize city school also tended to report stronger disagreement with items in the food preparation and gardening experience subscale, also suggesting disparities in involvement in these practices as a function of students' location. Given this variation across schools and locales, it is apparent that students who participated in this study do not have uniform experiences when it comes to exposure to organic food or participation in food preparation, decisions about what foods are purchased and consumed, and helping in home or community gardens. We can only speculate as to the underlying causes of these differences (e.g., whether indeed rural students have greater access to organic produce from local farms or farmers markets, or whether the city students might live in neighborhoods classified as food deserts), although they do not appear to be explained by socioeconomic status as measured by free/reduced lunch participation, which did not differ between the city and rural schools. Further investigation is needed to understand the nature of these experiential disparities across intersections of student characteristics; however, these findings underscore the potential for personalization based on prior experiences to support students' comprehension assessment performance.

Implications for Designing Personalized SBAs

The current findings have potential to support the development of more personalized socioculturally responsive SBAs. The findings suggest that knowledge measures should include both more traditional prior knowledge measures (e.g., topical vocabulary) and measures of students' relevant lived experiences to guide the adaptation of assessment content. Further, the findings suggest that knowledge measures should also include an IDK response option to improve the quality of inferences that can be drawn from these measures and provide students with a more positive and motivating test-taking experience. Test-takers' engagement in the assessment also emerged as an important consideration that not only accounted for a great deal of variation in SBA scores, but also interacted with students' levels of prior knowledge and lived experiences to affect performance. Effective measures of these student characteristics are necessary to implement personalized socioculturally relevant assessment design principles, such as the incorporation of community-based learning, negotiation of familiar and novel contexts, and reflection on students' multiple identities, into SBAs (O'Dwyer et al., 2023). Next, we describe three adaptations based on the present findings that are organized based on when they could occur in the test-taking experience.

To support personalization, prior knowledge and lived experience measures would need to be administered prior to the assessment. Thus, the first example adaptation we will discuss would occur immediately after the completion of these measures. We propose to leverage pedagogical agents (PAs) to

engage students in conversations to gain a more nuanced understanding of their connections to current assessment topics, and to identify students' interests to inform selection of other relevant topics, connecting with two design principles for socioculturally relevant SBAs (O'Dwyer et al., 2023): emphasizing empowerment through student-led actions and reflecting the multiple identities students bring to assessments. It may be the case that students do not always realize the ways in which particular topics are relevant to their daily lives (e.g., connecting their experiences with organic food at the grocery store to the topic of organic farming). In addition, the exact ways in which these topics might be relevant may vary depending on the regions from which the students come (e.g., urban vs. rural school students). The PAs can engage students in conversations that facilitate making these connections to increase the relevance of assessment topics to students, which in turn is likely to increase their motivation to complete the assessments, their engagement on individual assessment items, and ultimately assessment performance. These PA conversations can also be leveraged to identify relevant topics for students, either from a preset list of topics that have associated assessment versions (O'Dwyer et al., 2023) or to dynamically adapt assessment shells to new topics, an approach that is promising in mathematics assessment (Arslan et al., 2025) but needs further study in the context of literacy SBAs to evaluate its feasibility. PA conversations that seek to better understand students and the larger contexts of their day-to-day experiences offer another opportunity to demonstrate to students that who they are outside of the present assessment—their unique intersectional identity—matters in their educational experience.

The results of the prior knowledge and lived experience measures can also be leveraged to identify specific types of support students may need to ensure that they are given their best opportunity to succeed in assessments. Based on the present work, we hypothesized adaptations that would support two key groups of students: those with low topical vocabulary knowledge but high funds of knowledge and those with low topical vocabulary and low funds of knowledge. In the first case, the SBA content can be adapted to provide supplemental materials that help to build the necessary topical vocabulary to enable students to access the reading materials, while also helping them recognize how their funds of knowledge relate to the topic to support task engagement. In the second case, it may be appropriate to identify alternative SBA contexts, so as not to burden students with an assessment that may be difficult for them to access given limited prior knowledge and experiences. Evidence from recent empirical studies suggests that providing students with assessment forms that are more personally and culturally relevant to them can help to reduce performance gaps between historically underserved students and those from the dominant culture (e.g., assigning an SBA about African-American culture and history during the Harlem Renaissance to African-American students; Wang et al., 2025), as compared to less-culturally relevant topics (e.g., European immigration to the U.S.) or to traditional assessments that do not take cultural relevance into account (see Sinharay, Johnson, et al., 2025). Such reduced gaps in performance appear to be driven in part by higher rates of engagement on the culturally relevant test forms (Wang et al., 2025). Applying personalized adaptations that provide students with assessments focused on highly relevant topics and content, and set in familiar contexts and scenarios, can thus better support students' engagement and performance, giving them fairer opportunities to show what they know and can do, and reducing disparities in performance among subgroups of interest.

However, selection of familiar contexts should be balanced with opportunities to learn about unfamiliar topics and diverse cultural experiences outside students' own lived experiences, to help develop their cultural competence (Ladson-Billings, 1995b) and balance the trade-offs of familiar and novel contexts (O'Dwyer et al., 2023). Notably, Wang et al. (2025) found that culturally relevant test forms showed comparable reliability and standard errors of measurement for both the targeted and non-targeted groups. Thus, SBAs can provide "windows" into novel cultural contexts as well as "mirrors" that reflect students' prior experiences (Bishop, 1990) without sacrificing technical quality for groups who may find the topics less personally relevant. As noted earlier, context selection can vary based on who is making the selection (students, teachers, ML algorithms). We recommend including students as much as possible in the selection process to promote

feelings of agency and ownership in their educational experiences, which can positively impact motivation, engagement, and performance (Nakkula & Toshalis, 2025; O'Dwyer et al., 2023; Ryan & Deci, 2000; Walker et al., 2023). Future research is needed to determine the full implications and appropriate balance of assessments that are culturally "matched" and "mismatched" to student characteristics (Bennett, 2024; Sinharay & Johnson, 2024), as well as the impact of who is making the selection, on engagement and performance.

The last example adaptation we will discuss involves the inclusion of a culminating task in the SBA that not only connects with the various tasks completed as part of the SBA but also connects with students' prior experiences and the communities in which they participate. For example, students could be provided with the prompt, "What information is most important to communicate to members of your community?" as opposed to one that includes a more generic audience when developing a final summarization of what was learned from the SBA materials. A small modification to the prompt for a culminating task can support students to incorporate community-based learnings and situate the problem-solving process in the local context (O'Dwyer et al., 2023). This example does not require personalization of SBA tasks but rather would involve the personalization (or adaptation) of the ways in which student responses are evaluated, such as in the context of the communities to which they are communicating. This approach requires an understanding of the multiple, diverse perspectives, past experiences, and current goals of those communities, for both the student and those evaluating their responses. With the growing presence of generative artificial intelligence (AI) as a tool for both developing assessments and evaluating students' performance, it is critical for future research to investigate the ways in which generative AI can be leveraged to promote personalized socioculturally relevant assessments and the areas that should remain in the control of humans that have received training to develop their own cultural competence (Ladson-Billings, 1995b).

Limitations and Future Directions

While the findings summarized here serve as a foundation to support future work on personalization, the present work has several notable limitations. First, we used a self-report methodology to elicit information about students' knowledge and lived experiences as relevant to the topic of organic farming and other related topics. These measures are vulnerable to validity issues facing all such self-report measures (i.e., social desirability bias). Middle school students may also have limited insights into their own prior experiences, those of their families, and especially those of their communities. The IDK response option was in part intended to mitigate these response biases (e.g., encouraging students to report IDK rather than take a guess), and evidence suggests that students tended to make greater use of this option in self-report items asking about their communities, compared to items concerning family or individual experiences. Future work might explore whether relevant prior knowledge and experiences can be elicited through other means (e.g., conversation-based assessments; Zapata-Rivera et al., 2023). Second, we utilized a single SBA focused on a topic which may not be of equal interest to all students. It is likely that for some students, no amount of personalization would be sufficient to elicit interest in the topic necessary for effective engagement in and performance on the SBA. The SBA approach also requires longer administration time compared to self-report or discrete multiple choice items, which may have resulted in fatigue for some students. Applying personalization to route students to alternate topics (as suggested in the Implications section) or to use adaptive testing approaches that reduce the number of items required for valid estimation of performance may be a useful strategy to mitigate such concerns going forward. Finally, in our regression analyses, the inclusion of a rich set of predictors may have introduced potential suppression effects, complicating interpretation of some of the findings, especially with respect to interaction effects. Future work might evaluate more targeted models focused on the variables found in this study to have strong evidence of predictive validity, whether for SBA scores or RTE.

As noted previously, the current work is a first step in the development of personalized socioculturally responsive assessments. We included a large, diverse sample in terms of race, urbanicity, socioeconomic status, academic achievement, and prior knowledge. We have provided evidence that students' funds of knowledge and experiences and topic vocabulary knowledge are both meaningful predictors of their performance and engagement on an SBA. We have also hypothesized potential adaptations based on these findings that must be developed and evaluated by future research. It is critical that future research adopt co-design partnerships with teachers and students when developing such adaptations to ensure that the good intentions of this approach are realized into personalized socioculturally responsive assessment designs that meet the needs of students and can effectively be leveraged by teachers to support instruction (Ober et al., in press).

The present work builds upon prior research that investigated the critical role of prior knowledge on SBA performance by expanding the conception of prior knowledge to also include students' prior lived experiences and funds of knowledge. However, this expansion does not fully encompass the holistic student representation promoted by the CECA framework (Lehman, Sparks, Zapata-Rivera et al., 2024) and it is very likely that other student characteristics and other dimensions of prior knowledge (McCarthy & McNamara, 2021) are similarly important when determining when and how to provide caring support. This work identified several interactions among student characteristics (i.e., school location) and lived experiences, as well as between knowledge measures and test-taking engagement. As mentioned in the Methodology section, students also completed additional background surveys tapping constructs like reading self-efficacy, reading motivation, growth mindset, and cognitive flexibility, among others. Preliminary analyses suggested that these constructs play an important role in understanding the student characteristics contributing to engagement and SBA performance (Ober & Tenison, 2023). Further exploration of how such social-emotional factors may interact with prior knowledge, experiences, and engagement to affect assessment performance remains a fruitful area for future research.

Finally, the present work focused on SBA performance and engagement as key outcome variables. However, the CECA framework and caring educational systems more broadly (du Boulay et al., in press; Self, 1999) propose that caring support can have positive impacts beyond the immediate educational activity. Future research should explore the relationships among different types of caring support and various short- and long-term outcomes. For example, can caring support that leverages valuable insights from culturally relevant, responsive, and sustaining pedagogy help students from historically underrepresented groups feel a greater sense of belonging in academics in general, as well as in specific disciplines? It is important for future research on personalized socioculturally responsive assessments to identify the target outcomes for caring support to guide development and facilitate the evaluation of each support's effectiveness at providing students with safe, supportive, and engaging opportunities to show what they know and can do in ways that are best suited for them.

Declaration
The research reported here was supported in part by a grant from the Institute of Education Sciences (R305F100005) awarded to Educational Testing Service. The opinions expressed are those of the authors and do not represent the views of the U.S. Department of Education.

Received: 6/30/2025. **Accepted:** 1/2/2026. **Published:** 1/28/2026.

Citation: Sparks, J. R., Ober, T., & Lehman, B. (2026). Leveraging measures of students' prior knowledge and lived experiences toward more personalized culturally responsive assessment. *Practical Assessment, Research, & Evaluation*, 30(2)(6). Available online: https://doi.org/10.7275/pare.3372

Corresponding Author: Jesse R. Sparks, ETS Research Institute.
Email: jsparks@ets.org

References

Alrawashdeh, G. S., Fyffe, S., Azevedo, R. F. L., & Castillo, N. M. (2024). Exploring the impact of personalized and adaptive learning technologies on reading literacy: A global meta-analysis. *Educational Research Review, 42*, 100587. https://doi.org/10.1016/j.edurev.2023.100587

Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education: A synthesis of research across content areas. *Review of Educational Research, 86*(1), 163–206. https://doi.org/10.3102/0034654315582066

Arslan, B., Lehman, B., Tenison, C., Sparks, J.R., López, A.A., Gu, L., & Zapata-Rivera, D. (2024). Opportunities and challenges of using generative AI to personalize educational assessment. *Frontiers in Artificial Intelligence,* 7. 1460651. https://doi.org/10.3389/frai.2024.1460651

Arslan, B., Tenison, C., & Gooch, R. M. (2025). *On-the-fly context personalization with generative AI in mathematics assessments: Co-designing with teachers.* Paper presented at the annual meeting of the American Educational Research Association. Denver, CO.

Bennett, R. E. (2023). Toward a theory of socioculturally responsive assessment. *Educational Assessment, 28*(2), 83–104. https://doi.org/10.1080/10627197.2023.2202312

Bennett, R. E. (2024). Personalizing assessment: Dream or nightmare? *Educational Measurement: Issues and Practice, 43*(4), 119–125. https://doi.org/10.1111/emip.12652

Bennett, R. E. (2025). A descriptive review of culturally responsive, socioculturally responsive, and related assessment conceptions. In R. E. Bennett, L. Darling-Hammond, & A. Badrinarayan (Eds.), *Socioculturally responsive assessment: Implications for theory, measurement, and systems-level policy* (pp. 11–28). Routledge.

Bernacki, M. L., Greene, M. J., Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose(s)? *Educational Psychology Review, 33,* 1675–1715. https://doi.org/10.1007/s10648-021-09615-8

Behrendt, S. (2025). lm.beta: Add standardized regression coefficients to lm objects. *R package version 1.7-3*. https://cran.r-project.org/web/packages/lm.beta/lm.beta.pdf

Bishop, R. S. (1990). Mirrors, windows, and sliding glass doors. *Perspectives: Choosing and using books for the classroom, 6*(3). https://scenicregional.org/wp-content/uploads/2017/08/Mirrors-Windows-and-Sliding-Glass-Doors.pdf

Buchin, Z. L., & Mulligan, N. W. (2025). Prior knowledge and new learning: An experimental study of domain-specific knowledge. *Journal of Experimental Psychology: Applied, 31*(2), 84–98. https://doi.org/10.1037/xap0000520

du Boulay, B., Lehman, B., & Zapata-Rivera, D. (in press). Caring systems. In O. C. Santos, D. Zapata-Rivera, & M. Mavrikis (Eds.) *Encyclopedia of AI in Education*.

Evans, C., & Taylor, C. S. (Eds.). (2025). *Culturally Responsive Assessment in Classrooms and Large Scale Contexts*. New York: Routledge. https://doi.org/10.4324/9781003392217

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education, 53*(2), 106–116. https://doi.org/10.1177/0022487102053002003

Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice* (2nd ed.). New York: Teachers College Press.

Gay, G. (2013). Teaching to and through cultural diversity. *Curriculum Inquiry, 43*(1), 48–70. https://doi.org/10.1111/curi.12002

González, N., Moll, L. C., & Amanti, C. (Eds.). (2005). *Funds of knowledge: Theorizing practices in households, communities, and classrooms.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hattan, C., Alexander, P. A., & Lupo, S. M. (2024). Leveraging what students know to make sense of texts: What the research says about prior knowledge activation. *Review of Educational Research, 94*(1). https://doi.org/10.3102/00346543221148478

Higgs, K. P., Santuzzi, A. M., Gibson, C., Kopatich, R. D., Feller, D. P., & Magliano, J. P. (2023). Relationships between task awareness, comprehension strategies, and literacy outcomes. *Frontiers in Psychology, 14,* 1056457. https://doi.org/10.3389/fpsyg.2023.1056457

Hood, S. (1998). Culturally responsive performance-based assessment: Conceptual and psychometric considerations. *Journal of Negro Education, 67*(3), 187–196. https://doi.org/10.2307/2668188

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. http://dx.doi.org/10.1080/10705519909540118

Hwang, H., McMaster, K. J., & Kendeou, P. (2022). A longitudinal investigation of directional relations between domain knowledge and reading in the elementary years. *Reading Research Quarterly, 58*(1), 59–77. https://doi.org/10.1002/rrq.481

Kendeou, P., Rapp, D. N., & van den Broek, P. (2004). The influence of readers' prior knowledge on text comprehension and learning from text. In R. Nata (Ed.), *Progress in education* (Vol. 13, pp. 189–209). New York: Nova Science.

Ladson-Billings, G. (1995a). But that's just good teaching! The case for culturally relevant pedagogy. *Theory Into Practice, 34*(3), 159–165. https://doi.org/10.1080/00405849509543675

Ladson-Billings, G. (1995b). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465–491. https://doi.org/10.3102/00028312032003465

Ladson-Billings, G. (2009). *The dreamkeepers: Successful teachers of African American children* (2nd ed.). San Francisco, CA: Jossey-Bass.

Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: a.k.a. the Remix. *Harvard Educational Review, 84*(1), 74–84. https://doi.org/10.17763/haer.84.1.p2rj131485484751

Lau, P. N. K., Lau, S. H., Hong, K. S., & Usop, H. (2011). Guessing, partial knowledge, and misconceptions in multiple-choice tests. *Journal of Educational Technology & Society*, *14*(4), 99–110. https://www.jstor.org/stable/jeductechsoci.14.4.99

Lee, C. D. (1998). Culturally responsive pedagogy and performance-based assessment. *Journal of Negro Education, 67*(3), 268-279. https://doi.org/10.2307/2668195

Lehman, B., Sparks, J.R., & Bailey, J. (2025, April). *Engagement across grade levels on a low-stakes, large-scale assessment of mathematics.* Paper in symposium session presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.

Lehman, B., Sparks, J.R., Steinberg, J. (2024). *Detecting test-taking engagement in changing test contexts.* ETS Research Report Series. Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12384

Lehman, B., Sparks, J.R., Zapata-Rivera, D., Steinberg, J., & Forsyth, C. (2024). A culturally enhanced framework of caring assessments for diverse learners. *Practical Assessment, Research, & Evaluation, 29*(1): 9. https://doi.org/10.7275/pare.2102

McCarthy, K. S., Guerrero, T. A., Kent, K. M., Allen, L. K., McNamara, D. S., Chao, S.-F., Steinberg, J., O'Reilly, T., & Sabatini, J. (2018). Comprehension in a scenario-based assessment: Domain and topic-specific knowledge. *Discourse Processes, 55*(5-6), 510–524. https://doi.org/10.1080/0163853X.2018.1460159

McCarthy, K. S., & McNamara, D. S. (2021). The Multidimensional Knowledge in Text Comprehension framework. *Educational Psychologist, 56*(3), 196–214. https://doi.org/10.1080/00461520.2021.1872379

McCarthy, K. S., Steinberg, J., Dreier, K., O'Reilly, T., Sabatini, J., Butterfuss, R., & McNamara, D. S. (2023). The effects of prior knowledge in a scenario-based comprehension assessment: A multidimensional approach. *Learning and Individual Differences, 103,* 102283. https://doi.org/10.1016/j.lindif.2023.102283

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement.* New York: Routledge.

Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory Into Practice, 31*(2), 132–141. https://doi.org/10.1080/00405849209543534

Nakkula, M., & Toshalis, E. (2025). Student engagement, motivation, and assessment. In C. M. Evans & C. S. Taylor (Eds.), *Culturally Responsive Assessment in Classrooms and Large Scale Contexts* (pp. 34–51). Routledge. https://doi.org/10.4324/9781003392217-4

National Center for Education Statistics. (2024). Racial/Ethnic Enrollment in Public Schools. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved June 13, 2025. https://nces.ed.gov/programs/coe/indicator/cge

O'Dwyer, E., Sparks, J.R., & Nabors Oláh, L. (2023). Enacting a process for developing culturally relevant classroom assessments. *Applied Measurement in Education, 36*(3), 286-303. Special issue on Advancing Racial Justice in Educational Assessment: Perspectives and Practices. https://doi.org/10.1080/08957347.2023.2214652

O'Reilly, T., & Sabatini, J. (2013). *Reading for Understanding: How Performance Moderators and Scenarios Impact Assessment Design* (Research Report No. RR-13-31). Princeton, NJ: Educational Testing Service.

O'Reilly, T., Sabatini, J., & Wang, Z. (2019). What you don't know won't hurt you, unless you don't know you're wrong. *Reading Psychology*, *40*(7), 638–677. https://psycnet.apa.org/doi/10.1080/02702711.2019.1658668

O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. *Psychological Science*, *30*(9), 1344–1351. https://doi.org/10.1177/0956797619862276

Ober, T. M., Lehman, B., & Gooch, R. (in press). Teacher perspectives on implementing culturally responsive personalized learning. In M. Bernacki, C. Walkington, A. Emery, & L. Zhang (Eds.), *Handbook of Personalized Learning.*

Ober, T., & Tenison, C. (2023). Modeling students' reading comprehension in a theoretically grounded and culturally contextualized way. Paper presented in coordinated symposium session at the 7th Culturally Responsive Evaluation and Assessment conference, Chicago, IL.

Ober, T., Tenison, C., & Sparks, J.R. (2024, April). *"Don't Know About 'I Don't Know'"?: Handling Such Responses in Likert Scales.* Paper presented at the annual meeting of the National Council for Measurement in Education, Philadelphia, PA.

Paquette, L., & Baker, R. S. (2018). Comparing machine learning to knowledge engineering for student behavior modeling: A case study in gaming the system. *Interactive Learning Environments*, *27*, 585–597.

Paquette, L., de Carvalho, A. M. J. A., Baker, R. S., & Ocumpaugh, J. (2014). Reengineering the feature distillation process: a case study in the detection of gaming the system. In *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 284–287).

Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher, 41*(3), 93–97. https://doi.org/10.3102/0013189X12441244

Pearson, P. D., Palincsar, A. S., Biancarosa, G., & Berman, A. I. (2021). *Reaping the rewards of the Reading for Understanding initiative.* National Academy of Education. https://naeducation.org/reaping-the-rewards-of-reading-for-understanding-initiative/

Qualls, A. L. (1998). Culturally responsive assessment: Development strategies and validity issues. *Journal of Negro Education, 67*(3), 296–301. https://doi.org/10.2307/2668197

Ramasubramanian, S., Riewestahl, E., & Landmark, S. (2021). The Trauma-informed Equity-minded Asset-based Model (TEAM): The six R's for social justice-oriented educators. *Journal of Media Literacy Education, 13*(2), 29–42. https://doi.org/10.23860/JMLE-2021-13-2-3

RAND Reading Study Group. (2002). *Reading for understanding: Towards an R&D program in reading comprehension.* Santa Monica, CA: RAND.

Randall, J. (2021). "Color-neutral" is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues and Practice, 40*(4), 82–90. https://doi.org/10.1111/emip.12429

Randall, J., Poe, M., & Slomp, D. (2021). *Ain't* oughta be in the dictionary: Getting to justice by dismantling anti-black literacy assessment practices. *Journal of Adolescent and Adult Literacy, 64*(5), 594–599. https://doi.org/10.1002/jaal.1142

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

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78.

Sabatini, J. P., O'Reilly, T., Halderman, L. K., & Bruce, K. (2014). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities Research & Practice*, *29*(1), 36–43.

Sabatini, J., O'Reilly, T., & Doorey, N. A. (2018). *Retooling literacy education for the 21st Century: Key findings of the reading for understanding initiative and their implications.* ETS Center for Research on Human Capital and Education. https://www.ets.org/Media/research/pdf/Retooling_Literacy_Education.pdf

Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2020). Engineering a 21st Century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing, 20*(1), 1–23. https://doi.org/10.1080/15305058.2018.1551224

Scoboria, A., & Fisico, S. (2013). Encouraging and clarifying "don't know" responses enhances interview quality. *Journal of Experimental Psychology: Applied*, *19*(1), 72.

Self, J. (1999). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International Journal of Artificial Intelligence in Education, 10*, 350–364.

Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach (Ed.) *Adaptive Technologies for Training and Education* (pp. 7–27). New York, NY: Cambridge University.

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist, 57*(1), 31–54. https://doi.org/10.1080/00461520.2021.1939700

Sinharay, S., & Johnson, M. S. (2024). Computation and accuracy evaluation of comparable scores on culturally responsive assessments. *Journal of Educational Measurement, 61,* 5–46. https://doi.org/10.1111/jedm.12381

Sinharay, S., Johnson, M. S., Bennett, R. E., Lopez, R., Guzman-Orth, D., Sparks, J.R., & Pillarisetti, S. (2025). Culturally responsive assessments in the context of the National Assessment of Educational Progress: An initial look. *Educational Measurement: Issues and Practice.* http://doi.org/10.1111/emip.70008

Sinharay, S., Kane, M., Bennett, R., & Sparks, J.R. (2025). Validation for personalized assessments: A threats-to-validity approach. *Journal of Educational Measurement.* https://doi.org/10.1111/jedm.12434

Sparks, J.R., Lehman, B., & Zapata-Rivera, D. (2024). Caring assessments: Challenges and opportunities. *Frontiers in Education – Assessment, Testing, and Applied Measurement, 9*. https://doi.org/10.3389/feduc.2024.1216481

Sparks, J.R., Wang, Z., O'Reilly, T., & O'Dwyer, E. P. (in press). Designing personalized socioculturally responsive assessments of digital literacies. In J. Castek, J. Coiro, E. Forzani, M. S. Hagerman, C. Kiili, & J.R. Sparks (Eds.), *International Handbook of Research in Digital Literacies*. Routledge.

van der Linden, W. J., & Glas, C. A. W. (Eds.). (2010). *Elements of adaptive testing.* New York: Springer.

Wakabayashi, T., & Guskin, K. (2010). The effect of an "unsure" option on early childhood professionals' pre-and post-training knowledge assessments. *American Journal of Evaluation*, *31*(4), 486–498.

Walker, M. E., Olivera-Aguilar, M., Lehman, B., Laitusis, C., Guzman-Orth, D., & Gholson, M. (2023). *Culturally responsive assessment: Provisional principles.* (ETS Research Report No. RR-23-11). Princeton, NJ: ETS. https://doi.org/10.1002/ets2.12374

Walkington, C. & Bernacki, M. L. (2020). Appraising research on personalized learning: Definitions, theoretical alignment, advancements, and future directions. *Journal of Research on Technology in Education, 52*(3), 235–252. https://doi.org/10.1080/15391523.2020.1747757

Wang, Z., O'Reilly, T., Sabatini, J., McCarthy, K. S., & McNamara, D. S. (2021). A tale of two tests: The role of topic and general academic knowledge in traditional versus contemporary scenario-based reading. *Learning and Instruction, 73,* 1001462. https://doi.org/10.1016/j.learninstruc.2021.101462

Wang, Z., Sparks, J.R., Walker, M., O'Reilly, T., & Bruce, K. (2025). Group differences across scenario-based reading assessments: Examining the effects of culturally relevant test content. In R. Bennett, L.

Darling-Hammond, & A. Badrinarayan (Eds.), *Socioculturally responsive assessment: Implications for theory, measurement, and systems-level policy* (pp. 369-398). New York: Routledge.

Wise, S. L., Bhola, D., & Yang, S. (2006). Taking the time to improve the validity of low-stakes tests: The effort-monitoring CBT. *Educational Measurement: Issues and Practice*, *25*(2), 21–30.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, *18*(2), 163–183.

Wise, S. L., & Kuhfeld, M. R. (2021). Using retest data to evaluate and improve effort-moderated scoring. *Journal of Educational Measurement*, *58*(1), 130–149.

Wise, S. L., Kuhfeld, M. R., & Soland, J. (2019). The effects of effort monitoring with proctor notification on test-taking engagement, test performance, and validity. *Applied Measurement in Education*, *32*(2), 183–192.

Wise, S., & Ma, L. (2012). Setting response time thresholds for a CAT item pool: The normative threshold method. *Paper presented at the annual meeting of the National Council on Measurement in Education*. Vancouver, Canada.

Wise, S. L., & Smith, L. F. (2016). The validity of assessment when students don't give good effort. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 204–220). New York, NY: Routledge.

Zapata-Rivera, D. (2017). Toward caring assessment systems. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (UMAP '17). ACM, New York, NY, USA, 97–100.

Zapata-Rivera, D., & Arslan, B. (2021). Enhancing personalization by integrating top-down and bottom-up approaches to learner modeling. In R. A. Sottilare & J. Schwarz (Eds.) *Adaptive Instructional Systems. Adaptation Strategies and Methods.* HCII 2021. Lecture Notes in Computer Science (12793). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-77873-6_17

Zapata-Rivera, D., Lehman, B., & Sparks, J. R. (2020). Learner modeling in the context of caring assessments. In R. A. Sottilare & J. Schwarz (Eds.), *Proceedings of the second international conference on Adaptive Instructional Systems, held as part of HCI International Conference 2020, LNCS 12214*, (pp. 422-431). Cham, Switzerland: Springer Nature.

Zapata-Rivera, D., Sparks, J.R., Forsyth, C.M., & Lehman, B. (2023). Conversation-based assessment: Current findings and future work. In R. J. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International Encyclopedia of Education 4th Edition, Quantitative Research and Educational Measurement* (pp. 504-518). Elsevier. https://www.elsevier.com/books/international-encyclopedia-of-education/tierney/978-0-12-818629-9

Appendix

Table A1. Regression models predicting SBA scores for RQ1, RQ4 with standardized estimates (β) shown.

	Model 0		Model 1a		Model 1b		Model 2	
	β	*SE(β)*	*β*	*SE(β)*	*β*	*SE(β)*	*β*	*SE(β)*
School (Reference Group: Suburban/Small City)								
Midsize City	-0.015	0.035	-0.020	0.029	-0.007	0.034	-0.016	0.029
Suburban/Rural	0.061ǂ	0.034	0.001	0.028	0.059	0.033	0.004	0.028
Rural	0.038	0.031	-0.009	0.026	0.028	0.031	-0.012	0.026
Gender (Reference Group: Male)								
Female	**0.070***	0.029	**0.053***	0.024	**0.080****	0.028	**0.059***	0.024
Race/Ethnicity (Reference Group: White)								
Asian or Asian American	0.031	0.029	0.024	0.024	0.017	0.028	0.016	0.024
Black or African American	**-0.164*****	0.033	**-0.063***	0.028	**-0.154*****	0.032	**-0.064***	0.028
Hispanic/Latino	-0.055	0.037	0.005	0.031	-0.052	0.036	0.003	0.031
Other	-0.007	0.029	-0.013	0.024	-0.008	0.028	-0.012	0.024
Grade Level	**0.068***	0.028	-0.013	0.024	**0.070***	0.028	-0.003	0.024
Free/Reduced Price Lunch Participation (Reference Group: No)								
Yes	**-0.171*****	0.033	**-0.081****	0.028	**-0.148*****	0.032	**-0.077****	0.024
Special Education Status (Reference Group: No)								
Yes	**-0.101****	0.033	-0.016	0.028	**-0.080***	0.032	-0.011	0.028
English Learner Status (Reference Group: Never Classified as EL)								
Current or Reclassified/Former EL	**-0.066***	0.032	-0.018	0.027	-0.049	0.031	-0.009	0.027
SBA: Topical Vocabulary (TV) Score			**0.645*****	0.050			**0.600*****	0.051
SBA: Vocabulary-in-Context (VIC) Score			**0.327*****	0.032			**0.318*****	0.031

SBA: Use of IDK in TV Items			**0.458*** **	0.048		**0.429*** **	0.048
SBA: Use of IDK in VIC Items			**0.095** **	0.030		**0.088** **	0.030
Self-Report Sources of Knowledge					0.01	0.033	0.046	0.028
Funds of Knowledge for OF: Prior Learning					0.053	0.039	-0.022	0.033
Funds of Knowledge for OF: Local Relevance					**-0.087***	0.038	-0.023	0.033
Use of IDK in Prior Learning					**-0.061***	0.029	-0.032	0.025
Use of IDK in Local Relevance					**0.073***	0.032	0.042	0.027
Funds of Knowledge for Organic Food and Gardening: Organic Food Access and Value					**0.219*** **	0.033	**0.121*** **	0.029
Funds of Knowledge for Organic Food and Gardening: Food Preparation and Gardening					**-0.122*** **	0.035	**-0.072***	0.030
Use of IDK in Organic Food Access/Value					**0.084***	0.033	0.046	0.028
Use of IDK in Food Preparation/Gardening					0	0.032	0.008	0.027
Model R^2	.127		.396		.186		.415	
Model *F*	13.211		44.576		11.754		30.573	
Model *df*	(12, 1090)		(16, 1086)		(21, 1081)		(25, 1077)	

Note. ǂ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$. Intercept and its standard error (SE) are not reported because all variables were standardized, making the intercept approximately zero.

Table A2. Regression models predicting SBA engagement (RTE) for RQ2, RQ4 with standardized estimates (β) shown.

Variables	Model 0		Model 1a		Model 1b		Model 2	
	β	*SE(β)*	*β*	*SE(β)*	*β*	*SE(β)*	*β*	*SE(β)*
School (Reference Group: Suburban/Small City)								
Midsize City	**-0.247*** **	0.030	**-0.252*** **	0.030	**-0.238*** **	0.033	**-0.245*** **	0.030
Suburban/Rural	0.052‡	0.032	0.007	0.027	0.049	0.030	0.008	0.027
Rural	**0.070***	0.032	0.035	0.029	0.061‡	0.032	0.032	0.029
Gender (Reference Group: Male)								
Female	**0.082****	0.028	**0.070****	0.025	**0.087****	0.027	**0.073****	0.025
Race/Ethnicity (Reference Group: White)								
Asian or Asian American	0.033	0.028	0.028	0.025	0.021	0.027	0.021	0.025
Black or African American	**-0.109*** **	0.032	-0.033	0.029	**-0.102****	0.031	-0.035	0.029
Hispanic/Latino	0.006	0.036	0.052	0.032	0.012	0.035	0.054‡	0.032
Other	0.020	0.028	0.014	0.025	0.018	0.027	0.014	0.025
Grade Level	-0.035	0.027	**-0.097*** **	0.025	-0.034	0.027	**-0.088*** **	0.025
Free/Reduced Price Lunch Participation (Reference Group: No)								
Yes	**-0.140*** **	0.031	**-0.071***	0.028	**-0.116*** **	0.031	**-0.062***	0.028
Special Education Status (Reference Group: No)								
Yes	**-0.102****	0.032	-0.038	0.029	**-0.083****	0.031	-0.032	0.029
English Learner Status (Reference Group: Never Classified as EL)								
Current or Reclassified/Former EL	**-0.072***	0.031	-0.036	0.028	**-0.060***	0.030	-0.032	0.028
SBA: Topical Vocabulary (TV) Score			**0.473*** **	0.052			**0.429*** **	0.053
SBA: Vocabulary-in-Context (VIC) Score			**0.264*** **	0.033			**0.255*** **	0.033
SBA: Use of IDK in TV Items			**0.331*** **	0.049			**0.297*** **	0.050

SBA: Use of IDK in VIC Items			**0.095****	0.031		**0.093****	0.031
Self-Report Sources of Knowledge					-0.039	0.032	-0.012	0.029
Funds of Knowledge for OF: Prior Learning					0.046	0.038	-0.009	0.035
Funds of Knowledge for OF: Local Relevance					-0.039	0.037	0.009	0.034
Use of IDK in Prior Learning					-0.044	0.028	-0.021	0.026
Use of IDK in Local Relevance					0.045	0.031	0.023	0.028
Funds of Knowledge for Organic Food and Gardening: Organic Food Access and Value					**0.199*****	0.032	**0.126*****	0.030
Funds of Knowledge for Organic Food and Gardening: Food Preparation and Gardening					**-0.069***	0.034	-0.031	0.031
Use of IDK in Organic Food Access/Value					**0.092****	0.032	**0.063***	0.029
Use of IDK in Food Preparation/Gardening					-0.034	0.027	-0.030	0.028
Model R^2	.196		.353		.242		.370	
Model *F*	22.158		37.002		16.391		25.345	
Model *df*	(12, 1090)		(16, 1086)		(21, 1081)		(25, 1077)	

Note. ǂ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$. Intercept and its standard error (SE) are not reported because all variables were standardized, making the intercept approximately zero.

Table A3. Regression model with SBA engagement (RTE) as a predictor of SBA scores, controlling for student background and all knowledge measures for RQ3, RQ4 with standardized estimates (β) shown.

	Model 3	
	β	*SE (β)*
School (Reference Group: Suburban/Small City)		
Midsize City	**0.116*** **	0.025
Suburban/Rural	<.001	0.022
Rural	-0.029	0.023
Gender (Reference Group: Male)		
Female	0.020	0.020
Race/Ethnicity (Reference Group: White)		
Asian or Asian American	0.005	0.020
Black or African American	-0.045‡	0.023
Hispanic/Latino	-0.026	0.026
Other	-0.020	0.020
Grade Level	**0.045***	0.020
Free/Reduced Price Lunch Participation (Reference Group: No)		
Yes	-0.043‡	0.020
Special Education Status (Reference Group: No)		
Yes	0.006	0.023
English Learner Status (Reference Group: Never Classified as EL)		
Current or Reclassified/Former EL	0.008	0.022
SBA: Topical Vocabulary (TV) Score	**0.368*** **	0.043
SBA: Vocabulary-in-Context (VIC) Score	**0.180*** **	0.027

SBA: Use of IDK in TV Items	**0.268*** **	0.040
SBA: Use of IDK in VIC Items	0.038	0.025
Self-Report Sources of Knowledge	**0.052***	0.023
Funds of Knowledge for OF: Prior Learning	0.012	0.023
Funds of Knowledge for OF: Local Relevance	0.024	0.023
Use of IDK in Prior Learning	-0.020	0.021
Use of IDK in Local Relevance	0.030	0.022
Funds of Knowledge for Organic Food and Gardening: Organic Food Access and Value	**0.053***	0.024
Funds of Knowledge for Organic Food and Gardening: Food Preparation and Gardening	**-0.055***	0.025
Use of IDK in Organic Food Access/Value	-0.017	0.028
Use of IDK in Food Preparation/Gardening	-0.028	0.027
Engagement in SBA (30% RTE)	**0.541*** **	0.024
Model R^2	0.599	
Model *F*	61.843	
Model *df*	(26, 1076)	

Note. ǂ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$. Intercept and its standard error (SE) are not reported because all variables were standardized, making the intercept approximately zero.

Table A4. Additional models predicting SBA scores for RQ5, RQ6 with standardized estimates (β) shown.

	Model 4		Model 5		Model 6	
	β	*SE (β)*	*β*	*SE (β)*	*β*	*SE (β)*
School (Reference Group: Suburban/Small City)						
Midsize City	**0.114*** **	0.025	**0.114*** **	0.025	0.022	0.073
Suburban/Rural	-0.001	0.022	-0.001	0.022	0.020	0.106
Rural	-0.028	0.023	-0.027	0.023	-0.052	0.098
Gender (Reference Group: Male)						
Female	0.014	0.020	0.014	0.020	0.028	0.019
Race/Ethnicity (Reference Group: White)						
Asian or Asian American	0.003	0.020	0.003	0.020	0.014	0.018
Black or African American	-0.045‡	0.023	-0.045‡	0.023	-0.039	0.021
Hispanic/Latino	-0.025	0.026	-0.026	0.026	-0.019	0.024
Other	-0.018	0.020	-0.019	0.020	-0.018	0.019
Grade Level	**0.049***	0.020	**0.050***	0.020	0.068	0.093
Free/Reduced Price Lunch Participation (Reference Group: No)						
Yes	**-0.046***	0.023	**-0.046***	0.023	-0.035	0.021
Special Education Status (Reference Group: No)						
Yes	0.007	0.023	0.007	0.023	0.007	0.022
English Learner Status (Reference Group: Never Classified as EL)						
Current or Reclassified/Former EL	0.011	0.022	0.011	0.022	0.006	0.021
SBA: Topical Vocabulary (TV) Score	**0.363*****	0.043	**0.365*****	0.043	-0.059	0.217
SBA: Vocabulary-in-Context (VIC) Score	**0.185*****	0.027	**0.185*****	0.027	-0.080	0.214
SBA: Use of IDK in TV Items	**0.266*****	0.040	**0.254*****	0.058	**0.245*****	0.038

SBA: Use of IDK in VIC Items	0.043‡	0.025	0.083‡	0.049	0.044‡	0.023
Self-Report Sources of Knowledge	**0.057***	0.023	**0.056***	0.023	**0.047***	0.022
Funds of Knowledge for OF: Prior Learning	-0.023	0.028	-0.023	0.028	-0.043	0.026
Funds of Knowledge for OF: Local Relevance	-0.035	0.027	-0.035	0.027	0.001	0.025
Use of IDK in Prior Learning	-0.019	0.020	-0.019	0.020	-0.016	0.019
Use of IDK in Local Relevance	0.028	0.022	0.028	0.022	0.024	0.021
Funds of Knowledge for Organic Food and Gardening: Organic Food Access and Value	**0.054***	0.024	**0.054***	0.024	0.190	0.203
Funds of Knowledge for Organic Food and Gardening: Food Preparation and Gardening	**-0.055***	0.025	**-0.054***	0.025	0.112	0.196
Use of IDK in Organic Food Access/Value	0.013	0.023	0.013	0.023	-0.002	0.022
Use of IDK in Food Prep/Gardening	0.021	0.023	0.021	0.023	0.023	0.021
Engagement in SBA (30% RTE)	**0.534*****	0.024	**0.534*****	0.024	-0.208‡	0.108
Post-test Rating: IDK Feature Helped Performance	**0.058****	0.020	**0.067***	0.032		
Use of IDK in TV Items x IDK Feature Helped Performance			0.016	0.053		
Use of IDK in VIC Items x IDK Feature Helped Performance			-0.046	0.049		
TV x Organic Food Access/Value					0.102	0.124
TV x Food Preparation and Gardening					-0.136	0.108
VIC x Organic Food Access/Value					-0.140	0.103
VIC x Food Preparation and Gardening					-0.052	0.095
TV x School = Midsize City					0.040	0.048
TV x School = Suburban/Rural					0.003	0.078
TV x School = Rural					0.057	0.070
VIC x School = Midsize City					0.043	0.032

VIC x School = Suburban/Rural			0.040	0.048
VIC x School = Rural			-0.026	0.044
Organic Food Access/Value x School = Midsize City			**0.214***	0.086
Organic Food Access/Value x School = Suburban/Rural			-0.081	0.084
Organic Food Access/Value x School = Rural			**-0.159***	0.070
Food Preparation and Gardening x School = Midsize City			**-0.234***	0.092
Food Preparation and Gardening x School = Suburban/Rural			0.020	0.077
Food Preparation and Gardening x School = Rural			**0.159***	0.070
TV x Grade Level			0.245	0.204
VIC x Grade Level			-0.026	0.193
Organic Food Access/Value x Grade Level			-0.280	0.220
Food Preparation and Gardening x Grade Level			-0.003	0.204
TV x Engagement in SBA (30% RTE)			**0.461*****	0.093
VIC x Engagement in SBA (30% RTE)			**0.634*****	0.075
Organic Food Access/Value x Engagement in SBA (30% RTE)			**0.266***	0.117
Food Preparation and Gardening x Engagement in SBA (30% RTE)			-0.094	0.103
Model R^2	.602	.603	.663	
Model *F*	60.271	56.088	41.474	
Model *df*	(27, 1075)	(29, 1073)	(50, 1052)	

Note. ǂ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$. Intercept and its standard error (SE) are not reported because all variables were standardized, making the intercept approximately zero.

Table A5. Additional models predicting SBA engagement (RTE) for RQ5 with standardized estimates (β) shown.

Variables	Model 4		Model 5	
	β	*SE (β)*	*β*	*SE (β)*
School (Reference Group: Suburban/Small City)				
Midsize City	**-0.244*** **	0.030	**-0.249*** **	0.044
Suburban/Rural	0.007	0.027	0.006	0.027
Rural	0.035	0.029	0.036	0.029
Gender (Reference Group: Male)				
Female	**0.064***	0.025	**0.064***	0.025
Race/Ethnicity (Reference Group: White)				
Asian or Asian American	0.015	0.025	0.016	0.025
Black or African American	-0.031	0.029	-0.031	0.029
Hispanic/Latino	0.058‡	0.032	0.061‡	0.032
Other	0.014	0.025	0.015	0.025
Grade Level	**-0.083** **	0.025	**-0.081** **	0.025
Free/Reduced Price Lunch Participation (Reference Group: No)				
Yes	**-0.066***	0.028	**-0.068***	0.028
Special Education Status (Reference Group: No)				
Yes	-0.033	0.028	-0.031	0.028
English Learner Status (Reference Group: Never Classified as EL)				
Current or Reclassified/Former EL	-0.030	0.027	-0.031	0.027
SBA: Topical Vocabulary (TV) Score	**0.421*** **	0.052	**0.429*** **	0.052
SBA: Vocabulary-in-Context (VIC) Score	**0.252*** **	0.033	**0.247*** **	0.033
SBA: Use of IDK in TV Items	**0.290*** **	0.049	**0.167***	0.072

SBA: Use of IDK in VIC Items	**0.093****	0.031	**0.126***	0.062
Self-Report Sources of Knowledge	-0.006	0.029	-0.001	0.029
Funds of Knowledge for OF: Prior Learning	-0.011	0.034	-0.013	0.034
Funds of Knowledge for OF: Local Relevance	-0.002	0.034	-0.005	0.034
Use of IDK in Prior Learning	-0.022	0.026	-0.02	0.026
Use of IDK in Local Relevance	0.015	0.028	0.014	0.028
Funds of Knowledge for Organic Food and Gardening: Organic Food Access and Value	**0.122*****	0.030	**0.120*****	0.030
Funds of Knowledge for Organic Food and Gardening: Food Preparation and Gardening	-0.030	0.031	-0.027	0.031
Use of IDK in Organic Food Access/Value	**0.057***	0.029	0.055ǂ	0.029
Use of IDK in Food Preparation/Gardening	-0.026	0.028	-0.027	0.031
Post-test Rating: IDK Feature Helped Motivation	**0.088*****	0.025	0.025	0.041
Use of IDK in TV Items x IDK Feature Helped Motivation			**0.163***	0
Use of IDK in VIC Items x IDK Feature Helped Motivation			-0.040	0.004
Model R^2	.377		.381	
Model *F*	25.090		23.569	
Model *df*	(26, 1076)		(28, 1074)	

Note. ǂ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$. Intercept and its standard error (SE) are not reported because all variables were standardized, making the intercept approximately zero.