



Cultural Validity as Foundational to Assessment Development: An Indigenous Example

Pōhai Kūkea Shultz¹ and Kerry Englert^{2*}

¹University of Hawai'i at Mānoa, Honolulu, HI, United States, ²Seneca Consulting, Golden, CO, United States

The state of Hawai'i has a linguistically and culturally diverse population that recognizes Hawaiian and English as official languages. Working with the community, the state established the Hawaiian Language Immersion Program, Ka Papahana Kaiapuni Hawai'i (Kaiapuni), to support and promote the study of Hawaiian language, culture, and history. Kaiapuni students are historically marginalized test-takers and had been assessed using instruments that were culturally and linguistically insensitive, contained construct irrelevant variance, or had inadequate psychometric properties (U. S. Department of Education, 2006; Kaawaloa, 2014). In response, the Hawai'i State Department of Education and the University of Hawai'i developed the Kaiapuni Assessment of Educational Outcomes (KĀ'EO), which engages Kaiapuni students in technically rigorous, Native language assessments. This article details the theoretical framework of the KĀ'EO program, which includes traditional validity studies to build content and construct validity that support the assessment's use for accountability. However, the KĀ'EO team recognized that additional evidence was needed because the KĀ'EO theory of action is grounded in principles of community use of assessment scores to advance cultural and language revitalization. The article provides an example of one of the validity studies that the team conducted to build evidence in support of cultural and content validity.

Keywords: assessment, indigenous, cultural validity, community, psychometrics, theory of action, cognitive interview, native language revitalization

OPEN ACCESS

Edited by:

Janice Huber,
University of Alberta, Canada

Reviewed by:

Stavros Stavrou,
University of Saskatchewan, Canada
Dilma Mello,
Federal University of Uberlandia, Brazil

*Correspondence:

Kerry Englert
kenglert@senecaconsulting.org

Specialty section:

This article was submitted to
Assessment, Testing and Applied
Measurement,
a section of the journal
Frontiers in Education

Received: 28 April 2021

Accepted: 17 June 2021

Published: 01 July 2021

Citation:

Kūkea Shultz P and Englert K (2021)
Cultural Validity as Foundational to
Assessment Development: An
Indigenous Example.
Front. Educ. 6:701973.
doi: 10.3389/feduc.2021.701973

LANGUAGE REVITALIZATION AND ASSESSMENT

To ensure that students can access the material in a test to demonstrate their knowledge, test developers must consider the most salient characteristics of the student population. However, this can be a concern when testing diverse populations because assessment practices and priorities derive from the culture in which they are developed and are based on cultural and contextual assumptions (Solano-Flores et al., 2002; Solano-Flores, 2006; Nelson-Barber and Trumbull, 2007; Trumbull and Nelson-Barber, 2019). Assessment also necessitates academic knowledge of the written or spoken language (Solano-Flores, 2012). As Trumbull and Nelson-Barber (2019) state, "Nowhere is the disconnection between Native ways of knowing and Western ways of teaching more evident than in the arena of student assessment, most egregiously in the realm of large-scale tests" (p. 2). These implications must be acknowledged when developing a test to ensure valid measurement of the construct. If language and culture are not considered, a test could ultimately measure other domains, resulting in a biased assessment (Keegan et al., 2013). Therefore, test developers should consider

Brayboy's (2005) notion that, because racism and colonization are endemic to a society and therefore often invisible, these issues of cultural validity must be intentionally and systematically addressed within the test development process. This work helps ensure an equitable assessment guided by Cronbach's (1989) assertion: "Tests that impinge on the rights and life chances of individuals are inherently disputable" (p. 6).

In this article, we outline the theoretical framework that cultural validity should be the foundation for building validity evidence of an assessment program. Then, we provide an overview of the Kaiapuni Assessment of Educational Outcomes (KĀ'EO) program to contextualize the importance of including cultural validity evidence in an assessment program. Thus, in making our argument, we situate the idea of cultural validity within the community and discuss how community involvement in the KĀ'EO program is an integral part of building assessments for Hawaiian Language Immersion Program, or Kaiapuni, schools. We discuss how the community has been involved at critical junctures in the development of the assessment, including the formation of a theory of action. Finally, we provide an example of a cognitive interview study to illustrate how KĀ'EO developers use cultural validity in test development.

OVERVIEW OF THE KAIAPUNI ASSESSMENT OF EDUCATIONAL OUTCOMES PROGRAM

The state of Hawai'i has a linguistically and culturally diverse population. This diversity is highlighted in the state constitution, which names Hawaiian and English as official languages and ensures traditional and customary rights of Native Hawaiians (art. XV, § 4).¹ Established through years of struggle and activism by the Hawaiian community, these rights include the ability to use the Hawaiian language in home, school, and business settings (Haw. Const. art. X, § 4²; Lucas, 2000; Walk, 2007). Working with parents and Hawaiian leaders, the state established Ka Papahana Kaiapuni Hawai'i (Kaiapuni) program to support and promote the study of Hawaiian language, culture, and history. The Kaiapuni program currently consists of 25 public schools across five islands, with approximately 3,200 students enrolled (Warner, 1999; Wilson and Kamanā, 2001; Kawaiaea et al., 2007).

Because Kaiapuni schools are part of the state education system, they must comply with student testing requirements under the Every Student Succeeds Act (ESSA), which mandates annual testing. Kaiapuni schools face a unique challenge in administering statewide summative assessments because academic content is taught in Hawaiian. In the past, the Hawai'i State Department of Education (HIDOE) implemented two assessments for Kaiapuni students: a translation of the Hawai'i State Assessment (HSA) and the

Hawaiian Aligned Portfolio Assessment (HAPA). However, there were concerns with both.

The first assessment, the translation of the HSA from English to Hawaiian, lacked community credibility due to cultural and linguistic issues. For example, the underlying assumption in administering the HSA was that the translated versions of the summative assessments provided Hawaiian-language speakers with the opportunity to demonstrate their knowledge of the construct being measured (e.g., science) without changing the meaning of the construct. Thus, the translated versions were assumed to function like any other accommodation by leveling the playing field and improving score comparability between groups of students. However, the score comparability between the English and Hawaiian versions of the test was not necessarily well founded. Feedback from stakeholders and the state Technical Advisory Committee suggested that translated forms might not measure the same construct and might unduly disadvantage the students they are supposed to help (Kaawaloa, 2014; Englert et al., 2015).

The second assessment, the HAPA, provided a linguistically and culturally appropriate measure because it was developed in Hawaiian and specifically for Kaiapuni students. Although the HAPA was a positive shift to a more inclusive assessment that appropriately assessed students in their language of instruction (Abedi et al., 2004; Kieffer et al., 2009), technical quality issues hindered the use of the assessment for federal accountability (Kaawaloa, 2014; U.S. Department of Education, 2006). Subsequently, HIDOE returned to a Hawaiian translation of the English-language state assessment, much like the previous version of the HSA, which suffered translation issues and lacked community buy-in.

Because neither the translated HSA nor the HAPA provided an acceptable measure for use in accountability or for cultural appropriateness, the Kaiapuni community advocated for a fair and equitable assessment. In 2014, HIDOE contracted with the University of Hawai'i to develop the KĀ'EO. As a result, Kaiapuni students in grades 3–8 now engage in culturally appropriate Native language assessments in Hawaiian language arts, math, and science that are of sufficient technical quality to meet ESSA requirements (University of Hawai'i, 2020).

The KĀ'EO program is uniquely grounded in the language, culture, and worldview of the Kaiapuni community. The Foundational and Administrative Framework for Kaiapuni Education specifies the central role that assessment plays in the Kaiapuni schools: "It guides and binds us to our goals and values. It drives our curriculum and defines our teaching practices" (Ke Ke'ena Kaiapuni, Office of Hawaiian Education, 2015, p. 27). These factors play intricate and integral roles in the assessment process, thus necessitating a broader approach to the examination of validity. Because assessment practices and priorities are based on cultural and contextual assumptions, all aspects of test development reflect an underlying consideration of the learner, the learning process and context, and the content being measured (Keegan et al., 2013). These considerations place the onus on test developers to account for culture and language because they have a direct impact on the construct on which a test is based and which it will measure. Thus, KĀ'EO development

¹Haw Const. art. X, § 4.

²Haw Const, Haw Const. art. XV, § 4.

included widespread participation of the Hawaiian community in advisory groups, in writing and reviewing test items and student learning objectives, and in scoring the assessment.

CULTURE AND LANGUAGE IN ASSESSMENT VALIDITY

Good assessment practices dictate an explicit consideration of culture. Assessments often reflect the values, beliefs, and priorities of the dominant culture (Padilla and Borsato, 2008), which can create potential bias for underrepresented students. According to Klenowski (2009), learning and knowing are grounded in a sociocultural perspective because “differences in what is viewed as valued knowledge and the way individuals connect with previous generations and draw on cultural legacies (are) often mediated by the cultural tools that they inherit” (p. 90). This pluralistic perspective allows for improved relevancy of the assessment material as well as student access to and engagement with the material.

Culture and language need to be understood and examined on an ongoing basis and in multiple ways throughout the assessment development and administration process (He and van de Vijver, 2012). For example, Padilla and Borsato (2008) have recommended that community members be involved in assessment development and that test developers build their knowledge of customs and communication styles. Research has supported the adoption of a “pluralist” approach to item writing whereby test developers explicitly create items for a cultural group to ensure greater sensitivity (Keegan et al., 2013). In doing so, test developers can build assessments that strive to reduce bias through increased sensitivity, knowledge, and understanding.

Furthermore, test developers need to build comprehensive validity arguments that reflect their priorities for using data (Kane, 2012; 2006). Although culture has been considered in assessment literature (American Educational Research Association, American Psychological Association National Council on Measurement in Education, 2014), it has often been viewed as a threat to validity rather than an intrinsic consideration as an aspect of validity (Solano-Flores, 2011). Yet, as Kirkhart (2016) has argued, validity is culturally situated and should be central to any validity study to ensure sensitive and accurate measurement. If test developers fail to think broadly about validity and the degree to which an assessment embodies cultural priorities, their actions can result in the marginalization of some participants.

Thus, a validity argument that considers culture and language should be centered on two priorities. The first is ensuring the results of a Native language instrument can be used to draw similar conclusions to those from comparable English-language state assessments. In other words, the assessment must meet the technical requirements for accountability specified in U.S. Department of Education (2018) peer review. This necessitates that the assessment use many of the same validity methods used in assessment programs such as the Smarter Balanced Assessment Consortium (2010, 2020). Many of the KĀ’EO validity studies

were conducted to ensure rigorous methods were applied in a similar manner as in the Smarter Balanced Assessment Consortium (2020). The KĀ’EO technical manuals provide additional information on the range of validity and reliability studies supporting the program, including content, construct, and cognitive evidence (University of Hawai’i, 2020).

At the same time, an assessment that accounts for culture and language needs to exceed basic ideas of validity, which leads to the second and more crucial priority. To truly account for language and culture, the assessment results must be sensitive and responsive to the needs of a diverse community (Trumbull and Nelson-Barber, 2019). Cultural validity processes should be integrated into the traditional validity methodology and considered in the interpretation of those results. Cultural validity reinforces the need to follow traditional psychometric methods but also pushes the development and analysis to look beyond the familiar.

To the degree possible, test developers who consider culture and language should look to researchers who are also walking this path (Kaomea, 2003; Lawrence-Lightfoot, 2005). Those researchers articulate the challenge of looking beyond the familiar in how data are examined and interpreted in order to find more complex and nuanced narratives. By using “defamiliarizing tools, anti-oppressive researchers working in historically marginalized communities can begin to ask very different kinds of questions that will enable us to excavate layers of silences and erasures and peel back familiar hegemonic maskings” (Kaomea, 2003, p. 24). This orientation provides test developers with insight into a community’s priorities for an assessment while developing a theory of action, interpreting students’ responses during cognitive interviews, and even interpreting and reporting statistical data. This article represents an invitation for others to join in unpacking the complex narrative of inclusion and equity.

Cultural validity, as the foundation of an appropriate assessment for Native students, builds on broader theories such as critical race theory (Ladson-Billings and Tate, 1995), TribalCrit (Brayboy, 2005), culturally sustaining/revitalizing pedagogy (Ladson-Billings, 2014; McCarty and Lee, 2014), culturally responsive schooling (Castagno and Brayboy, 2008), and culturally relevant education (Aronson and Laughter, 2016), to name a few. Whereas these broader theories agree on the basic premise that racism and colonization are an inherent feature of our society and schooling and are “pedagogies of opposition committed to collective empowerment and social justice” (Aronson and Laughter, 2016, p. 164), cultural validity provides a particularly useful lens to examine how these theories can serve as critical underpinnings in a discussion of Native language assessments. Discussing culturally responsive schooling, Castagno and Brayboy (2008) describe the justified reservations of Native communities about an increased focus on standardized testing. But what if an assessment exhibits the “deep understanding of sovereignty and self-determination” that Castagno and Brayboy (2008, p. 969) advocate for, particularly as a part of a community’s effort to revitalize their Native language?

As Hermes et al. (2012) argue, the loss of Native languages has deep impacts on communities. Culturally responsive, sustaining, or revitalizing practice cannot be simply an add-on to address a failure of the American education system (Castagno and Brayboy, 2008; McCarty and Lee, 2014). Rather, we need approaches that “deepen insights for understanding how functioning in multiple discourses translates into strategies for language revitalization while also illuminating the role of Indigenous knowledge systems in learning” (Hermes et al., 2012, p. 382). Furthermore, Aronson and Laughter (2016) argue that if culturally relevant education is to have broad social justice impacts, educators need to “creatively play by the rules” while also fighting for change and educational sovereignty (p. 199). This is nothing new for Native educators, particularly those who use the culturally sustaining or revitalizing pedagogies that McCarty and Lee (2014) advocate for. Navigating between policies that prioritize monolingual and monocultural standards, while privileging the language, culture, and identity of Native students, is a balancing act and an everyday occurrence for educators of Native students.

This balancing act and the need for culturally sustaining pedagogy has been articulated by Native Hawaiian scholars as well (Benham and Heck, 1998; Warner, 1999; Wilson and Kamanā, 2001; Kawaiāea et al., 2007). For decades, schooling for Hawaiian students has served what Benham (2004) describes as “contested terrain” and represented a struggle over “content, values, instructional strategies, measures of accountability, and so on” (p. 36). In recognizing these systemic inequities, the Hawaiian community has been self-determining in creating culturally responsive schools, developing teacher education programs grounded in Hawaiian culture and language, and centering Hawaiian culture and language-based pedagogy (Kaomea, 2009). Utilizing this kind of self-determining approach has been the cornerstone of Hawaiian language revitalization and the development of Hawaiian immersion schools with notable success (Wilson and Kamanā, 2006). Furthermore, in alignment with notions of culturally sustaining pedagogy and cultural validity, Goodyear-Ka’ōpua (2013) has argued that as we navigate through a mainstream educational system that continues its history of inequality, we must take an approach grounded in survivance (Kūkea Shultz, 2014; Vizenor, 1999) and what she terms sovereign pedagogies because “education that celebrates Indigenous cultures without challenging dominant political and economic relations will not create futures in which the conditions of dispossession are alleviated” (Goodyear-Ka’ōpua, 2013, p. 6). Like bricoleurs (Kaomea, 2003; Berry, 2006), we must be savvy in our efforts to leverage ideals like culturally sustaining pedagogy and sovereign pedagogies in the development of assessments for marginalized communities. Cultural validity, as it relates to assessment development, is one way to do just that.

We propose that cultural validity is not a distinct type of validity; rather, it underpins the entire concept of validity. Thus, each time a validity study is developed as a part of the KĀ’EO program, the ways in which language and culture form a part of the validity argument are considered. Each validity study advances the thinking around the complexities involved in cultural validity, which are informed by worldview, learning

styles, and community (Solano-Flores and Nelson-Barber, 2001). As Solano-Flores and Nelson-Barber (2001) have suggested, “Ideally, if cultural validity issues were addressed properly at the inception of an assessment . . . there would be no cultural bias and providing accommodations for cultural minorities would not be necessary” (p. 557). True cultural validity goes beyond fairness and equity to consider culture and language through the purposeful involvement of the community. This process ensures transparency, buy-in, and ownership by the community, and it promotes a level of validity that cannot be achieved through traditional methods (Trumbull and Nelson-Barber, 2019).

When cultural validity is viewed as a critical component for all validity studies in a program, those studies become more actionable and focused. In the KĀ’EO validity studies, specific discussions about language and culture are embedded in the process and always include as many educators and community members as possible, with representation from all communities in the study context. As Nelson-Barber and Trumbull (2007) advocate, “Until assessment practices with Native students can be flexible enough to take into account the contexts of such students’ lives, they will not meet a standard of cultural validity” (p. 141). Across all KĀ’EO project tasks and validity studies, there is an intentional focus on integrating considerations of Native Hawaiian aspirations, wisdom, language, and worldview. Evidence is collected throughout the development, refinement, and analyses of the test cycle. The KĀ’EO developers built a foundation for the validity framework, using a theory of action that places Hawaiian culture at the center of the program. The theory of action provides a crucial foundation for all of the validity studies.

CULTURAL VALIDITY IN ACTION

In its broadest use, a theory of action provides a framework for evaluating the impacts of an initiative or program (Bennett, 2010; Lane, 2014). In the context of educational testing, a theory of action can be used to frame a validity argument (Kane, 2006) or, more simply, to evaluate whether the intended effects and benefits of an assessment have been achieved (Bennett, 2010). However, the development of theories of action in many assessment programs is built on a monolingual, English-based construct (Lane, 2014). Bennett et al. (2011) provided an interpretation and graphic representation of the theories of action of English summative assessments and found little to no focus was placed on students’ language and culture (beyond their achievement levels) and community. Validating an alternative approach, Haertel (2018) challenges evaluation specialists and researchers to “examine the ways testing practices have sometimes functioned to justify or support systemic social inequality” while also employing new and unfamiliar research methods and tools and collaborating with others outside of the field (p. 212).

The development of the KĀ’EO theory of action aligns with Haertel (2018) as well as the tenets of culturally responsive schooling espoused by Castagno and Brayboy (2008). The theory of action is grounded in community involvement and places student outcomes at the center of the work as well as

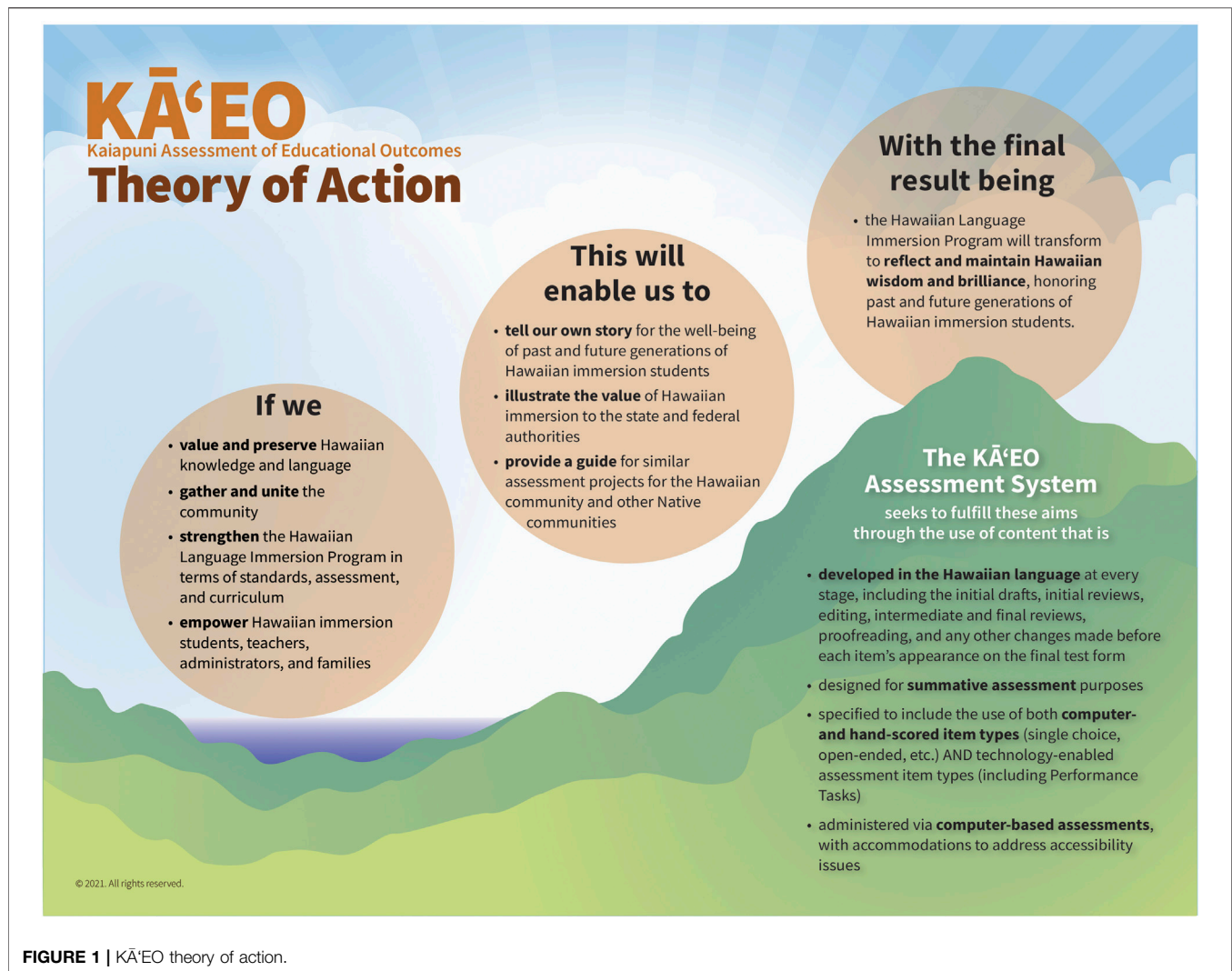


FIGURE 1 | KĀ'EO theory of action.

systems of Native epistemologies and interests (Paris and Alim, 2014). The development of the KĀ'EO theory of action focused on two important priorities: engaging with the stakeholders of the community and privileging Hawaiian knowledge, language, and culture throughout the development process. These two priorities helped to ensure that the community's aspirations for their children were respected. The KĀ'EO theory of action is one example of how the test developers and community stakeholders successfully balanced the tension between maintaining the technical requirements of a state assessment and serving the needs of the community as defined by the community (McCarty and Lee, 2014; Patton, 2011; **Figure 1**).

The KĀ'EO theory of action informed the validity work by highlighting several key considerations in terms of building validity evidence. First, the KĀ'EO program was integral to the preservation and revitalization of the Hawaiian language, and the assessment results would strengthen the Kaiapuni schools by providing key data to teachers, parents, and students. This could be done only by ensuring the assessment accurately measured key linguistic and cognitive attributes. Second,

engaging the community throughout the assessment process in different ways would ensure a unified vision for the assessment and the data use. Validity studies were intentionally structured with the goal of building the credibility and value of the assessment. The cognitive interviews described below are an example of carefully building evidence to ensure the assessment reinforces key linguistic and cognitive attributes.

A VALIDITY STUDY IN SUPPORT OF CULTURAL VALIDITY

A key to the KĀ'EO validity argument was understanding the specific linguistic and cognitive processes of the Kaiapuni program's bilingual students. An example of the program's validity studies, cognitive interviews seek to understand the cognitive and linguistic underpinnings of the assessment. Cognitive interviews should be a foremost concern in test development to ensure that students are interpreting the items as intended (Solano-Flores and Trumbull, 2003; Trumbull and

Nelson-Barber, 2019). Research has shown that emerging bilinguals possess “cognitive and linguistic practices that differ from monolinguals” (Menken et al., 2014, p. 602). These practices should be examined and evaluated independently to best understand their unique characteristics (Menken et al., 2014; Mislevy and Durán, 2014).

When developing validity evidence for tests such as the KĀ‘EO, there needs to be deep consideration of the complexity of linguistic, academic content, and contextual factors. Because the Hawaiian language is being revitalized (Warner, 1999; Wilson and Kamanā, 2019), students often possess different skills and abilities in grammar, text familiarity, and sociolinguistic knowledge (Weir, 2005). This is often driven by a student’s home language, which is sometimes Hawaiian but most often English. Because the KĀ‘EO is a new testing program for the Hawaiian language, there is an imperative to understand how students are making sense of items given their range of linguistic and academic abilities.

Cognitive Interviews

The central issue in assessment is accurately measuring students on the key domain or construct. To this end, test developers have implemented various methods to ensure that assessments allow students to demonstrate their knowledge in appropriate ways. One method of understanding students’ access to items is cognitive interviews (Zucker et al., 2004; Rabinowitz, 2008; Almond et al., 2009). This method uses structured interviews to ask students to discuss their mental processes and interpretations as they work through an item. This method can provide test developers with a greater understanding of how students are interpreting the item and how it corresponds to the intended construct. Cognitive interviews can specifically help identify confusing instructions, items that are unclear, and item choices (i.e., distractors) that are poorly worded. While cognitive testing can be resource intensive in terms of developing and administering protocols and coding the qualitative results, it is critical for test developers to use other methods (e.g., reliability analyses) in conjunction with cognitive interviews to build a complete picture of score reliability and validity. Cognitive interviews were particularly critical for the KĀ‘EO to build a validity argument that supported cultural validity.

Methods

Purpose. In considering the cognitive interviews, the KĀ‘EO team looked to the theory of action to guide the study. It was critical not only to understand how students interacted with the test but also to build validity evidence in support of linguistic and cognitive processes. By speaking directly with students, the KĀ‘EO team could gain deeper insights into the students’ linguistic processing and into interactions between their language proficiency and their access to the content of the assessment items. In addition, the team could better understand any bias in the items as well as the clarity of the items, which provided assurance that the items maintained an integrity to the Hawaiian language and cultural knowledge.

During a thorough item review, conducted after the analysis of items on the 2019 assessments, issues related to reliability emerged. First, despite efforts to improve the reliability data of all the assessments, the grade 7 math assessment continued to have lower levels of reliability. Second, the analyses revealed the low reliability of particular items for students in the IEP/504 subgroup. In addition, language proficiency issues were a recurring theme, and the KĀ‘EO team decided to explore those issues in the cognitive interviews. Kaiapuni students often have a range of exposure to and education in the Hawaiian language, and their varied degrees of fluency may affect their access to the assessment items (Ke Ke‘ena Kaiapuni, Office of Hawaiian Education, 2015). Finally, each Kaiapuni school might present academic content using different terminology, grammar, or structure. These differences needed to be evaluated to ensure that each student could access the material on the assessment.

Participants. To get a representative sample of students from across different islands and schools, the KĀ‘EO team invited 11 schools on five islands across Hawai‘i to have their students participate in the cognitive interviews. Potential participants were grade 8 students from Kaiapuni schools, who were administered the grade 7 KĀ‘EO math assessment in 2019. From the 11 schools invited to participate, 19 students were interviewed at four schools on three islands. Seven of these students had an IEP/504 plan, and 12 did not. The KĀ‘EO team asked classroom teachers to select students who were at or above proficiency based on their observations in the classroom and who were in the IEP/504 subgroup. The KĀ‘EO team worked closely with HIDOE on a data sharing agreement that would protect student anonymity and data. HIDOE also initiated all communications with school administration. In addition, clear communication with parents and students was critical to provide an understanding of the process as well as to allow them to ask questions or opt out of the interviews.

Interview and Analysis Protocol. Cognitive interviews use a one-on-one questioning approach whereby an interviewer (e.g., researcher) sits with a student and asks specific questions about how they solved the assessment items. There are two main methods for conducting cognitive interviews: concurrent and retrospective (Zucker et al., 2004). The concurrent method involves collecting data from students as they work through an item, whereas the retrospective method involves asking questions immediately after students work through an item. Both methods provide useful information, but the KĀ‘EO team used the retrospective method for these cognitive interviews because that method is less likely to interfere with a student’s performance (Zucker et al., 2004).

The KĀ‘EO team selected a set of three questions from the grade 7 math subject area to be presented to the grade 8 participants. Using these items and the operational test software, the team produced a testlet that replicated the appearance of the operational form. In addition, interviewers used digital booklets that included each reading passage, item, and associated distractors. The booklets also included interview scripts and prompts as well as places to type in all necessary documentation to ensure consistent information was collected for

each item. All of these documents were created in Hawaiian to align with the language of instruction and the assessment.

Each cognitive interview was conducted in Hawaiian by a team consisting of an interviewer and a note-taker. Both team members were fluent in Hawaiian and thoroughly trained on the protocols and scripts so that they would not influence the students' responses. The following list represents an interpretation of the general questions asked during the interview, although it may not reflect follow-up questions related to specific questions or issues:

- Can you explain how you found the answer? What was the first thing you did?
- Did you find this question easy, medium or difficult? Was the language of the question clear?
- Do you remember learning this content?
- Did you come across any vocabulary words that you did not understand? What did you do when that happened? Is there another word that you would use to describe this concept?

The interviewer worked one-on-one with students to ensure they were comfortable with and understood the process. The interviewer read the scripts and prompts and guided the timing of the interview. In addition, the interviewer was instructed to pause (e.g., 10 s) between a student's responses to allow the student to give complete answers. The note-taker documented the process and recorded notes. Each session lasted no more than 40 min, which ensured students were engaged in the interview process and did not become tired or frustrated.

After conducting each cognitive interview at the separate school sites, interviewers and note-takers met to debrief. This session was held immediately after the cognitive interviews in order to document initial impressions of how the students responded to the questions as well as improve the process for following interviews. After all cognitive interviews were completed, the notes from debriefings and the notes taken during the interviews in the digital booklets were collated and organized by question to make the analysis more seamless. The KĀ'EO team, who has knowledge and expertise of the items, Hawaiian language proficiency, and content knowledge, conducted a final analysis of the notes. During this analysis, the team identified salient themes and organized the initial analysis into two categories: 1) a summary of students' thoughts, opinions, and actions toward each question; and 2) a recommendation for future actions regarding each question.

An example of this process was in the questioning and analysis of student feedback to a grade 7 math question aligned to a grade 7 statistics and probability standard. The question itself had content-specific vocabulary in Hawaiian related to probability, random samples, and so forth, which can be challenging for emerging bilinguals or students with limited language proficiency skills. Students who were considered at or above proficiency as well as students with IEP/504 plans were interviewed about this question, and the results were similar. All students thought that the question was relatively easy and that, overall, the language was clear. One student suggested that some kind of graphic or visual representation might help in understanding the question. The

most interesting finding related to this question, however, was that students overwhelmingly did not understand the term "random sample," which was a central part of the question. This one term, when in Hawaiian, ended up being the part of the question that hindered students the most and prevented some of them from selecting the correct answer. These results were illuminating and led to recommendations for concrete actions specific to this question and other questions with challenging, content-specific vocabulary to maximize student access to question content. The next section includes a more in-depth discussion related to vocabulary, content, and language proficiency, but it is clear that the KĀ'EO team has merely scratched the surface in terms of the potential of cognitive interview analysis and its impact on the assessment development.

Findings

After completing the analysis of each individual question as summarized above, the KĀ'EO team analyzed the results for broad themes that could inform future item and test development as well as professional development activities initiated by HODOE.

Vocabulary and Content. As in years past, knowledge of content-specific vocabulary appeared to be a factor in how students understood and explained questions. This issue was apparent in most cognitive interviews but seemed to be more prominent in the higher grades, particularly regarding content-specific vocabulary in math items. As the content becomes more complex, so too does the Hawaiian language vocabulary associated with it. In addition, teaching math content in Hawaiian is not easy because it requires a high level of language proficiency and mastery of content-specific knowledge in two languages. If teaching this higher level of language is difficult, then learning it is just as difficult. Thus, the issue of content-specific vocabulary was evident in the cognitive interviews with the students. In addition, 2019 was the first year for students to be administered an operational test in Hawaiian, which may have affected the content taught during the year as well as students' familiarity with the Hawaiian vocabulary and content aligned with a Hawaiian worldview.

Acting on this finding, the KĀ'EO team continued to focus on the content-specific vocabulary to ensure that the Kaiapuni schools can properly prepare students for the words used on the assessment. The team also made recommendations to HODOE and its Office of Hawaiian Education about providing Hawaiian language professional development in schools to strengthen curriculum and instruction and make resources in the math content area available to teachers. In addition, because students and teachers are expected to become more familiar with the content-specific vocabulary and Kaiapuni student learning outcomes in future KĀ'EO administrations, this should become a less central issue in future years.

Language Proficiency. Another salient theme that emerged from the analysis of the cognitive interview data was the impact of students' language proficiency on their performance on the KĀ'EO. Although evidence of this impact had surfaced in previous cognitive interviews, it was clearly seen in the 2019 interviews with the grade 8 students.

As mentioned previously, in selecting participants for the cognitive interviews, the KĀ'EO team asked schools to select a group of students who had high proficiency in the Hawaiian language. During each interview, the interviewer and note-taker informally evaluated a student's language proficiency based on the conversation as well as whether the student reported that they consistently spoke Hawaiian at home with another family member. Although interviewers were not able to pinpoint an exact proficiency level for each student, they were able to gain a general sense of a student's language proficiency from their conversation.

Overwhelmingly, students who had higher proficiency in Hawaiian could not only better understand and correctly answer questions but also better articulate their reasoning behind their answers. Although this finding does not point to a direct correlation between language proficiency and performance on KĀ'EO, it is the first step in examining how language proficiency affects a student's performance.

In addition, a student's ability to clearly articulate their reasoning is a highly valued skill in all three content areas of the KĀ'EO. In extended response questions for Hawaiian language arts, for example, students need to make inferences about a text they read and thoroughly explain how those inferences are evident in the text. In extended response questions for science, students need to describe how natural phenomena are connected and what their impacts are. Finally, for math, many questions are aligned with Claim 3: Communicate Reasoning of the Smarter Balanced Assessment Consortium (2020): "Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others." In all of these cases, language proficiency in Hawaiian is vitally important to a student's ability to articulate an argument, evaluate an argument, tell a story, or explain their thinking. Not having high enough Hawaiian language proficiency can seriously inhibit a student's ability to show what they know.

What is clear from these findings is that language proficiency may affect more than just a student's right or wrong answers on the KĀ'EO. It may also point to item performance data, particularly for questions on the assessment that require students to explain or justify their answers. This finding may seem obvious, but less obvious is how the KĀ'EO team may go about addressing the impacts to item performance (Claim three questions in math, for example) and addressing the language proficiency issue on the assessment. As mentioned previously, Kaiapuni educators are "recalibrating" their instruction in the classroom because of the KĀ'EO, so the language proficiency issue is expected to become less pronounced and have less of an impact over time.

One addition to the research agenda of the KĀ'EO project is the collection of evidence for external validity. In the 2019 administration, the KĀ'EO team began surveying students about their language proficiency and using that survey data for both external validity and differential item functioning (DIF). DIF is a measure of bias that examines the degree to which individual assessment items have differential response patterns between demographic groups (e.g., boys vs. girls; Camilli and Shepard, 1994). The KĀ'EO team began an additional round of DIF analysis

in 2019 to start to build an understanding of DIF as it relates to students' self-reported language proficiency skills. Even though this DIF analysis is in its early stages and no valid and reliable conclusions can be drawn yet, using the results of the analysis, along with the cognitive interview findings, has tremendous potential. The KĀ'EO team will report these findings to HIDOE so that professional development can be developed that not only addresses the need for more knowledge around the Kaiapuni student learning outcomes but also targets the Hawaiian language proficiency of students and teachers.

Summary

Even though cognitive interviews are a typical part of the assessment development process, the approach and findings for the KĀ'EO program are far from typical. Rooted in the KĀ'EO theory of action and the tenets of cultural validity, these cognitive interviews privileged Hawaiian language, culture, and worldview while gathering essential data to inform future item development. The diversity of this population of students, who are mostly emerging bilinguals with varying levels of language proficiency, dictates that attention be paid to language and culture in every study, survey, and analysis. This is because cultural validity is reflected in "the effectiveness with which . . . assessment addresses the sociocultural influences that shape student thinking and the ways in which students make sense of . . . items and respond to them" (Solano-Flores and Nelson-Barber, 2001, p. 555).

CONCLUSION AND IMPLICATIONS

"It is a blatant understatement to say that approaches to the assessment of Indigenous students in the United States have fallen far short of an ideal of culturally-responsive, culturally-valid practice (Trumbull and Nelson-Barber, 2019, p. 8).

KĀ'EO developers continue to build an understanding of the importance of cultural validity. The efforts to improve assessment for the Kaiapuni schools is a journey toward improved relevancy and understanding. Cultural validity has provided a framework to ensure the KĀ'EO program provides opportunities for students to demonstrate their knowledge while also fulfilling a commitment to support the community on the path to language revitalization. The explicit acknowledgment of the centrality of Hawaiian language and culture emphasizes the importance of a close collaboration with the community to build an understanding of the KĀ'EO program and to ensure that this assessment does not fall short of what Kaiapuni students deserve.

Community involvement has been an essential aspect of the KĀ'EO development and a way to gather data to support the cultural validity of the assessment. The theory of action was developed with input from K-12 educators, higher education staff, and community members in the Hawaiian language. Cognitive interviews confirmed the integral role that language plays in every aspect of the assessment development process. Finally, the complexity of the population of Kaiapuni students,

who are mostly emerging bilinguals (Mislevy and Durán, 2014) and second language learners of Hawaiian with varying levels of proficiency, dictates that special attention be paid to cultural validity in every study conducted in relation to the KĀ'EO.

As the first Native language program of its kind, the KĀ'EO provides a unique opportunity to develop assessments that reflect the priorities and needs of the Hawaiian community. The test developers are ready and willing to engage in this significant work. Through the integration of psychometric theories of validity (e.g., construct and content validity) and cultural validity, the developers hope to continue to improve the KĀ'EO as it is aligned to and reflective of the Hawaiian worldview. This includes a responsibility to contribute not only to the field of assessment but also to the work of social justice as described by Castagno and Brayboy (2008): "As with the concepts of sovereignty and self-determination, racism, its manifestations, and its effects must be made a more explicit part of the discussion among scholars researching and writing about (culturally responsive schooling)" (pp. 950–951). It is in this spirit that the KĀ'EO contributes to the broader narrative in educational assessment and cultural validity and shows that a Native language assessment built with the community can have broad impacts.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the State of Hawai'i, Department of Education has ownership of data and should be contacted for any data requests. Requests to access the datasets should be directed

to <https://www.hawaiipublicschools.org/VisionForSuccess/SchoolDataAndReports/HawaiiEdData/Pages/Data-Requests.aspx>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the State of Hawai'i, Department of Education. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

PK and KE have contributed to the development of the assessment system and validity framework. The data collection was designed, led, and summarized by PK. Additional data analysis was conducted by KE. All authors contributed to the article draft.

FUNDING

The authors declare that this project is funded by the Hawai'i Department of Education in partnership with the University of Hawai'i beginning in 2014, contract #s: 2014.1 and CO-10834. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. Seneca Consulting, LLC provided funding for the publication costs.

REFERENCES

- Abedi, J., Hofstetter, C. H., and Lord, C. (2004). Assessment Accommodations for English Language Learners: Implications for Policy-Based Empirical Research. *Rev. Educ. Res.* 74 (1), 1–28. doi:10.3102/00346543074001001
- Almond, P. J., Cameto, R., Johnstone, C. J., Laitusis, C., Lazarus, S., and Nagle, K. (2009). *Cognitive Interview Methods in reading Test Design and Development for Alternate Assessments Based on Modified Academic Achievement Standards (AA-MAS)* [White Paper]. *Measured Prog. SRI Int.* 1–38.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Aronson, B., and Laughter, J. (2016). The Theory and Practice of Culturally Relevant Education. *Rev. Educ. Res.* 86 (1), 163–206. doi:10.3102/0034654315582066
- Benham, M. K. P., and Heck, R. H. (1998). *Culture and Educational Policy in Hawai'i: The Silencing of Native Voices*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Benham, M. K. P. (2004). Where Can We Collectively Be that Is Greater Than where We Are Now? *Hulili: Multidisciplinary Res. Hawaiian Well-Being* 1 (1), 35–48.
- Bennett, R. E. (2010). Cognitively Based Assessment of, for, and as Learning (CBAL): A Preliminary Theory of Action for Summative and Formative Assessment. *Meas. Interdiscip. Res. Perspective* 8 (2-3), 70–91. doi:10.1080/15366367.2010.508686
- Bennett, R. E., Kane, M., and Bridgeman, B. (2011). *Theory Of Action and Validity Argument in the Context of Through-Course Summative Assessment Educational Testing Service*. Available at: https://www.ets.org/research/policy_research_reports/publications/paper/2011/imxq (Accessed February 18, 2021).
- Berry, K. S. (2006). "Research as Bricolage: Embracing Relationality, Multiplicity and Complexity," in *Doing Educational Research: A Handbook*. Editors K. Tobin and J. Kincheloe (Rotterdam, Netherlands: Sense Publishers), 3–14.
- Brayboy, B. M. J. (2005). Toward a Tribal Critical Race Theory in Education. *Urban Rev.* 37 (5), 425–446. doi:10.1007/s11256-005-0018-y
- Camilli, G., and Shepard, L. A. (1994). *Methods for Identifying Biased Test Items*. Newberry Park, CA: Sage Publications.
- Castagno, A. E., and Brayboy, B. M. J. (2008). Culturally Responsive Schooling for Indigenous Youth: A Review of the Literature. *Rev. Educ. Res.* 78 (4), 941–993. doi:10.3102/0034654308323036
- Cronbach, L. J. (1989). "Construct Validation after Thirty Years," in *Intelligence: Measurement, Theory, and Public Policy*. Editor R. L. Linn (Champaign, IL: University of Illinois Press), 147–171.
- Englert, K., Kelly, J., and Malone, N. (2015). *Considerations for Hawaiian Language Immersion Program Assessments: A Review of the Literature*. Honolulu, HI: Office of Hawaiian Affairs.
- Goodyear-Ka'ōpua, N. (2013). *The Seeds We Planted: Portraits of a Native Hawaiian Charter School*. University of Minnesota Press.
- Haertel, E. H. (2018). Tests, Test Scores, and Constructs. *Educ. Psychol.* 53 (3), 203–216. doi:10.1080/00461520.2018.1476868
- He, J., and van de Vijver, F. (2012). Bias and Equivalence in Cross-Cultural Research. *Online Readings Psychol. Cult.* 2 (2). doi:10.9707/2307-0919.1111
- Hermes, M., Bang, M., and Marin, A. (2012). Designing Indigenous Language Revitalization. *Harv. Educ. Rev.* 82 (3), 381–402. doi:10.17763/haer.82.3.q8117w861241871j
- Kaawaloa, L. K. (2014). *Translation V. Tradition Fighting for Equal Standardized Testing Ma Ka 'ōlelo Hawai'i*, 36. Honolulu, HI: University of Hawai'i Law Review, 487–528.2.

- Kane, M. T. (2006). "Validation," in *Educational Measurement*. Editor R. L. Brennan. 4th ed. (Westport, CT: Praeger), 17–64.
- Kane, M. (2012). Validating Score Interpretations and Uses. *Lang. Test.* 29 (1), 3–17. doi:10.1177/0265532211417210
- Kaomea, J. (2009). Contemplating Kuleana: Reflections on the Rights and Responsibilities of Non-indigenous Participants in Programmes for Indigenous Education. *AlterNative: Int. J. Indigenous Peoples* 5 (2), 78–99. doi:10.1177/117718010900500205
- Kaomea, J. (2003). Reading Erasures and Making the Familiar Strange: Defamiliarizing Methods for Research in Formerly Colonized and Historically Oppressed Communities. *Educ. Res.* 32 (2), 14–23. doi:10.3102/0013189X032002014
- Kawaiaea, K., Housman, A. K., and Alencastre, M. (2007). Pūā I Ka 'ōlelo, Ola Ka 'ohana: Three Generations of Hawaiian Language Revitalization. *Hūlili: Multidisciplinary Res. Hawaiian Well-Being* 4 (1), 183–237.
- Keegan, P. J., Brown, G. T. L., and Hattie, J. A. C. (2013). "A Psychometric View of Sociocultural Factors in Test Validity: The Development of Standardized Test Materials for Māori-Medium Schools in New Zealand/Aotearoa," in *Constructing Educational Achievement: A Sociocultural Perspective*. Editors S. Phillipson, K. Y. L. Ku, and S. N. Phillipson (London, UK: Routledge), 42–54.
- Ke Ke'ena Kaiapuni, Office of Hawaiian Education (2015). *The Foundational & Administrative Framework for Kaiapuni Education: Hawaiian Language Immersion Program*. Honolulu, HI: Hawai'i State Department of Education. <https://www.hawaiipublicschools.org/DOE%20Forms/KaiapuniFrameworkFinal.pdf>.
- Kieffer, M. J., Lesaux, N. K., Rivera, M., and Francis, D. J. (2009). Accommodations for English Language Learners Taking Large-Scale Assessments: A Meta-Analysis on Effectiveness and Validity. *Rev. Educ. Res.* 79 (3), 1168–1201. doi:10.3102/0034654309332490
- Kirkhart, K. (2016). Validity: Through Whose Eyes? [Invited Workshop]. Denver, CO, United States: Meeting of the Colorado Evaluation Network. Available at: <https://coloradoevaluationnetwork49.wildapricot.org/event-2166562>.
- Klenowski, V. (2009). Australian Indigenous Students: Addressing Equity Issues in Assessment. *Teach. Educ.* 20 (1), 77–93. doi:10.1080/10476210802681741
- Kukea Shultz, P. L. (2014). *Kuleana 'ohana Kaiapuni: A story of agency and Hawaiian Immersion Families* [Doctoral Dissertation, University of Hawai'i at Mānoa]. ScholarSpace Available at: <http://scholarspace.manoa.hawaii.edu/handle/10125/100418>.
- Ladson-Billings, G. (2014). Culturally Relevant Pedagogy 2.0: A.k.a. The Remix. *Harv. Educ. Rev.* 84 (1), 74–84. doi:10.17763/haer.84.1.p2rj131485484751
- Ladson-Billings, G., and Tate, W. F. I. (1995). Toward a Critical Race Theory of Education. *Teach. Coll. Rec.* 97 (1), 47–68.
- Lane, S. (2014). Validity Evidence Based on Testing Consequences. *Psicothema* 26 (1), 127–135. doi:10.7334/psicothema2013.258
- Lawrence-Lightfoot, S. (2005). Reflections on Portraiture: A Dialogue between Art and Science. *Qual. Inq.* 11 (1), 3–15. doi:10.7334/psicothema2013.25810.1177/1077800404270955
- Lucas, P. F. N. (2000). E Ola Mau Kākou I Ka 'ōlelo Makuahine: Hawaiian Language Policy and the Courts. *Hawaiian J. Hist.* 34, 1–28.
- McCarty, T., and Lee, T. (2014). Critical Culturally Sustaining/revitalizing Pedagogy and Indigenous Education Sovereignty. *Harv. Educ. Rev.* 84 (1), 101–124. doi:10.17763/haer.84.1.q83746nl5pj34216
- Menken, K., Hudson, T., and Leung, C. (2014). Symposium: Language Assessment in Standards-Based Education Reform. *Tesol Q.* 48 (3), 586–614. doi:10.1002/tesq.180
- Mislevy, R. J., and Durán, R. P. (2014). A Sociocognitive Perspective on Assessing EL Students in the Age of Common Core and Next Generation Science Standards. *Tesol Q.* 48 (3), 560–585. doi:10.1002/tesq.177
- Nelson-Barber, S., and Trumbull, E. (2007). Making Assessment Practices Valid for Indigenous American Students. *J. Am. Indian Educ.* 46 (3), 132–147.
- Padilla, A. M., and Borsato, G. N. (2008). "Issues in Culturally Appropriate Psychoeducational Assessment," in *Handbook of Multicultural Assessment: Clinical, Psychological, and Educational Applications*. Editors L. A. Suzuki and J. G. Ponterotto. 3rd ed. (San Francisco, CA: Jossey-Bass), 5–21.
- Paris, D., and Alim, H. S. (2014). What Are We Seeking to Sustain through Culturally Sustaining Pedagogy? A loving Critique Forward. *Harv. Educ. Rev.* 84 (1), 85–100. doi:10.17763/haer.84.1.9821873k2ht16m77
- Patton, M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York, NY: Guilford Press. doi:10.14375/np.9782365833691
- Rabinowitz, S. (2008). *Assessment of English Language Learners under Title I and Title III: How One Testing Program Can Inform the Other*. Washington, DC, United States: LEP Partnership meeting. [Paper presentation] Available at: <https://nclca.edu.gov/meetings/oct08lepship>.
- Smarter Balanced Assessment Consortium. (2020). 2018–19 Summative Technical Report. Available at: https://technicalreports.smarterbalanced.org/2018-19_summative-report/_book/ (Accessed January 15, 2021).
- Smarter Balanced Assessment Consortium (2010). Smarter Balanced Assessment Consortium: Theory of Action; an Excerpt from the Smarter Balanced Race to the Top Application. <https://eric.ed.gov/?id=ED536956>. (Accessed August 5, 2017)
- Solano-Flores, G. (2011). "Assessing the Cultural Validity of Assessment Practices: An Introduction," in *Cultural Validity in Assessment: Addressing Linguistic and Cultural Diversity*. Editors M. d. R. Basterra, E. Trumbull, and G. Solano-Flores (Routledge), 3–21.
- Solano-Flores, G. (2012). Smarter Balanced Assessment Consortium: Translation Accommodations Framework for Testing English Language Learners in Mathematics Smarter Balanced Assessment Consortium. Available at: <https://portal.smarterbalanced.org/library/en/translation-accommodations-framework-for-testing-english-language-learners-in-mathematics.pdf> (Accessed July 17, 2017).
- Solano-Flores, G. (2006). Language, Dialect, and Register: Sociolinguistics and the Estimation of Measurement Error in the Testing of English Language Learners. *Teach. Coll. Rec.* 108 (11), 2354–2379. doi:10.1111/j.1467-9620.2006.00785.xhttps://www.tcrecord.org/Home.asp
- Solano-Flores, G., and Trumbull, E. (2003). Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-language Learners. *Educ. Res.* 32 (2), 3–13. doi:10.3102/0013189x032002003
- Solano-Flores, G., Trumbull, E., and Nelson-Barber, S. (2002). Concurrent Development of Dual Language Assessments: An Alternative to Translating Tests for Linguistic Minorities. *Int. J. Test.* 2 (2), 107–129. doi:10.1002/tea.101810.1207/s15327574ijt0202_2
- Solano-Flores, G., and Nelson-Barber, S. (2001). On the Cultural Validity of Science Assessments. *J. Res. Sci. Teach.* 38 (5), 553–573.
- Trumbull, E., and Nelson-Barber, S. (2019). The Ongoing Quest for Culturally-Responsive Assessment for Indigenous Students in the U.S. *Front. Educ.* 4, 1–11. doi:10.3389/feuc.2019.00040www.frontiersin.org/article/10.3389/feuc.2019.00040
- University of Hawai'i (2020). Kaiapuni Assessment of Educational Outcomes (KĀ'EO): 2019 Technical Report. Available at: www.hawaiipublicschools.org/Reports/KĀEO%20Technical%20Report_2019_Operational%20Tests.pdf (Accessed January 14, 2021).
- U.S. Department of Education (2018). A State's Guide to the U.S. Department of Education's Assessment Peer Review Process. Available at: <https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf> (Accessed October 30, 2018).
- U.S. Department of Education (2006). *Hawaii Assessment Letter*. Available at: <https://www2.ed.gov/admins/lead/account/nclbfinalassess/hi2.html> (Accessed March 1, 2021).
- Vizenor, G. (1999). *Manifest Manners: Narratives on Postindian Survivance*. Lincoln, NE: University of Nebraska Press.
- Walk, K. (2007). "Officially" what? the Legal Rights and Implications of 'ōlelo Hawai'i. 30. *University of Hawai'i Law Review*, 243–268. Available at: <https://hawaiilawreview.com/> (Accessed January 14, 2021).
- Warner, S. L. N. e. (1999). Kuleana: The Right, Responsibility, and Authority of Indigenous Peoples to Speak and Make Decisions for Themselves in Language and Cultural Revitalization. *Anthropol. Educ. Q.* 30 (1), 68–93. doi:10.1525/aeq.1999.30.1.68
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Palgrave MacMillan. doi:10.1057/9780230514577

- Wilson, W. H., and Kamanā, K. (2001). "Mai Loko Mai O Ka 'Īini: Proceeding from a Dream", in *The green Book of Language Revitalization in Practice*. Editors L. Hinton and K. Hale (Academic Press), 147–176. doi:10.1016/b978-012349353-8/50020-0
- Wilson, W. H., and Kamanā, K. (2019). 'Aha Pūnana Leo - Advancing from the Grassroots. *Linguapax Rev.* 7, 49–70. Available at: <https://www.linguapax.org/wp-content/uploads/2020/02/linguapax19-1-1.pdf> (Accessed February 9, 2021).
- Wilson, W. H., and Kamanā, K. (2006). "For the Interest of the Hawaiians Themselves": Reclaiming the Benefits of Hawaiian-Medium Education. *Hulili: Multidisciplinary Res. Hawaiian Well-Being* 3 (1), 153–181.
- Zucker, S., Sassman, C., and Case, B. J. (2004). *Cognitive Labs* [Technical Report]. Pearson. Available at: http://images.pearsonassessments.com/images/tmrs/tmrs_rg/CognitiveLabs.pdf (Accessed July 15, 2015).

Conflict of Interest: KE was employed by the company Seneca Consulting.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Kūkea Shultz and Englert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.