Caring assessments: challenges and opportunities

Jesse R. Sparks*, Blair Lehman and Diego Zapata-Rivera

ETS, Princeton, NJ, United States

Caring assessments is an assessment design framework that considers the learner as a whole and can be used to design assessment opportunities that learners find engaging and appropriate for demonstrating what they know and can do. This framework considers learners' cognitive, meta-cognitive, intra- and inter-personal skills, aspects of the learning context, and cultural and linguistic backgrounds as ways to adapt assessments. Extending previous work on intelligent tutoring systems that "care" from the field of artificial intelligence in education (AIEd), this framework can inform research and development of personalized and socioculturally responsive assessments that support students' needs. In this article, we (a) describe the caring assessment framework and its unique contributions to the field, (b) summarize current and emerging research on caring assessments related to students' emotions, individual differences, and cultural contexts, and (c) discuss challenges and opportunities for future research on caring assessments in the service of developing and implementing personalized and socioculturally responsive interactive digital assessments.

KEYWORDS

caring assessments, formative assessment, interactive digital assessment, caring systems, adaptive assessment and learning, socioculturally responsive assessment

1 Introduction

Personalization in the assessment context is an umbrella term that can include many different approaches. Most prior research and development has focused on adaptations based on students' prior knowledge or performance during the assessment (e.g., Shemshack et al., 2021). However, personalization may sometimes consider other intra-or interpersonal aspects of students' experience (Du Boulay, 2018). For example, student engagement has been utilized in effort-monitoring computer-based tests (Wise et al., 2006, 2019) and a wider range of student emotions have been used to enhance performance-based adaptation in several personalized learning systems (D'Mello et al., 2011; Forbes-Riley and Litman, 2011). Research in the field of artificial intelligence in education (AIEd) has increasingly emphasized a more holistic picture of learners which takes into account cognitive, metacognitive, and affective aspects of the learner to explain their behavior in learning environments (Grafsgaard et al., 2012; Kizilcec et al., 2017; Yadegaridehkordi et al., 2019), reflecting growing interest in integrating positive psychology into research within the AIEd community (Bittencourt et al., 2023).

The caring assessments (CA) framework provides an approach for designing adaptive assessments that learners find engaging and appropriate for demonstrating their knowledge, skills, and abilities (KSAs; Zapata-Rivera, 2017). This conceptual framework considers cognitive aspects of the learner as well as metacognitive, intra-and interpersonal skills, aspects of the learning context, cultural and linguistic backgrounds, and interaction behaviors within an integrated learner model and uses this model to personalize assessment to students' needs

(Zapata-Rivera et al., 2023). Multiple lines of research must be conducted to bring this vision for caring assessment to fruition. This Perspective article describes the CA framework and its unique contributions to the field (Section 2) and summarizes current and emerging research on the CA framework emphasizing students' emotions (Section 3), individual differences (Section 4), and cultural contexts (Section 5). Challenges and opportunities emerging from this literature are also discussed (Section 6), highlighting gaps and future directions for AIEd research that is most promising to advance the vision of CA.

2 The caring assessments framework

The CA framework (see Figure 1) is a *conceptual framework for adaptive assessment design* which proposes that assessments can provide a more engaging student experience while collecting more precise information about their KSAs by better understanding who students are and how they interact with the assessment (Zapata-Rivera, 2017). This better understanding of students can be leveraged to provide "caring" in terms of *adaptations* before, *caring support* during, and *feedback* after the assessment (Lehman et al., 2018).

Caring support before the assessment involves the development of *student profiles* that include a variety of information about the student, from their *personal characteristics* (e.g., interests, beliefs, linguistic background) to *contextual information* such as prior learning opportunities (Zapata-Rivera et al., 2020). These profiles can then be leveraged to provide students with an adapted version of the assessment that affords them the best opportunity to engage with the assessment and demonstrate what they know and can do. Alternative versions of the assessment could vary from the assessment format (e.g., multiple-choice items or game-based) to the language (e.g., toggle between English/Spanish) to the context of the assessment (e.g., using different texts while measuring the same underlying reading skills).

The student profiles that enable caring support before the assessment also serve as the start for providing caring support during

FIGURE 1
Caring assessment framework.

the assessment. Caring support during the assessment will require an *integrated learner model* (ILM) that considers both student and contextual characteristics (from the student profile) and the *interaction behaviors* students demonstrate during the assessment. This ILM is a more complex learner model than is typically employed in personalized learning and assessment tasks but draws on prior research on various types of learner models (Zapata-Rivera and Arslan, 2021; Bellarhmouch et al., 2023). This ILM can leverage information from the student profile and interactions to provide on-demand support. For example, a student might become disengaged during the assessment and the ILM could deploy a motivational message that has been personalized based on the student's interests or prior opportunity to learn within the domain (Kay et al., 2022).

Caring support after the assessment is primarily provided in the feedback report. The goal is to provide feedback to the student that will be easy to understand and motivate them to continue their learning journey. This necessitates feedback reports that utilize asset-based language (Gay, 2013; Ramasubramanian et al., 2021) and provide context for performance on the assessment by leveraging the information in the ILM (e.g., identifying learners' relevant prior knowledge and lived experiences and the strengths they demonstrated on the assessment along with areas for improvement). This contextualized reporting could, for example, identify if student responses were connected to specific behavioral patterns or could connect current performance to students' prior experiences or opportunities to learn to highlight progress. This contextualized reporting can also be utilized when providing feedback to teachers, which can then support teacher decision-making on the next appropriate steps to support student learning and continue caring support outside of the assessment.

The CA framework builds on several areas of prior research. The notion of an adaptive "caring" assessment system (Zapata-Rivera, 2017) builds on AIED research on adaptive intelligent tutoring systems that "care" as they support learning (Self, 1999; Kay and McCalla, 2003; Du Boulay et al., 2010; Weitekamp and Koedinger, 2023). Attending to a broader set of student characteristics, contexts, and behaviors also allows the CA framework to leverage findings from multiple learning theories when developing "caring" supports. Emphasis on using intra-and interpersonal characteristics and other contextual information to drive assessment adaptation is consistent with and can leverage models of self-regulated learning (e.g., Winne and Hadwin, 1998; Pintrich, 2000; Kay et al., 2022). The inclusion of a broader set of characteristics, contexts, and behaviors also extends the idea of "conditional fairness" in assessments that use contextual information about students' backgrounds to adapt assessment designs and scoring rules (Mislevy, 2018) and extends typical research on computer adaptive assessments driven by performance and item difficulty (van der Linden and Glas, 2010; Shemshack et al., 2021).

While the CA framework has relevance to both large-scale summative and classroom formative assessment contexts, there is greater potential flexibility in applying this framework to the design of tools to be used in formative contexts, due to the emphasis on providing on-demand "caring" support to help learners maximize their learning and engagement during assessment tasks (Zapata-Rivera, 2017). Efforts toward realizing this framework have investigated how students' emotions, individual differences, and cultural contexts can best be leveraged to provide personalized assessment experiences. Next, we summarize this current and emerging research.
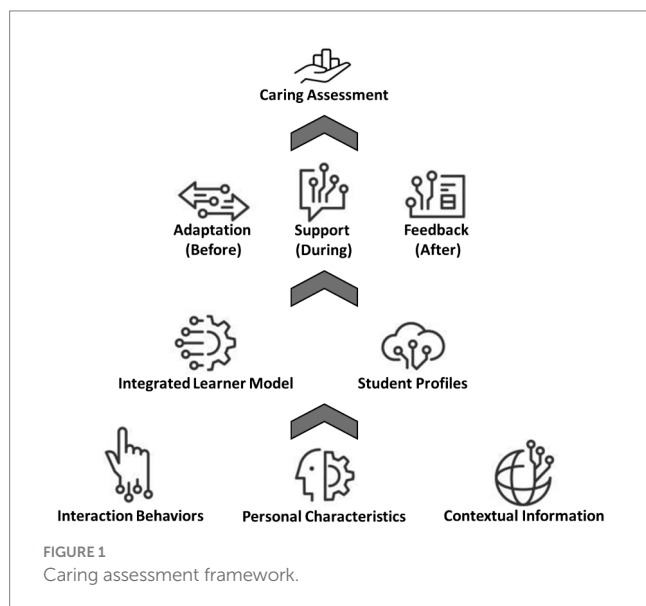
3 Student emotions

As anyone who has completed an assessment knows, it can be an emotional experience. However, very few assessments support students to remain in a productive emotional state (see Wise et al., 2006, 2019 for exceptions) or consider students' emotions when determining assessment outcomes (see Wise and DeMars, 2006 for an exception). Most research on student emotions during test taking has focused on documenting those experiences after test completion – and have shown that the experience of different emotions are differentially related to assessment outcomes (Spangler et al., 2002; Pekrun et al., 2004, 2011; Pekrun, 2006). Research on the impact of student emotions during learning activities has received far greater attention (see D'Mello, 2013 for a review) and there are multiple examples of personalized learning systems that leverage both student cognition and emotions to provide feedback and guide instructional decisions (e.g., D'Mello et al., 2011; Forbes-Riley and Litman, 2011).

In our own research on the emotional experiences of students during interactions with conversation-based assessments we build on prior work in both assessment and learning contexts by focusing on the intensity of discrete emotions (Lehman and Zapata-Rivera, 2018). When intensity was considered, we found the same pattern across boredom, frustration, and confusion: low intensity was positively correlated, medium intensity was not correlated, and high intensity was negatively correlated with performance, despite no overall relationship with performance. While it has been found that confusion has a more positive relationship with learning than boredom and frustration (e.g., Craig et al., 2004; D'Mello and Graesser, 2011, 2012, 2014; D'Mello et al., 2014) and frustration a more positive relationship than boredom (Baker et al., 2010), in assessment context it appears that the three emotions have a similar relationship with outcomes. However, the intensity findings for confusion, specifically, may relate to prior findings that the partial (Lee et al., 2011; Liu et al., 2013) or complete resolution of confusion (D'Mello and Graesser, 2014; Lehman and Graesser, 2015) is necessary for learning. Real-time tracking of students' emotional experiences (states and intensity) can be leveraged to provide caring support during the assessment as has been successfully implemented in personalized digital learning systems. However, integration of emotion detectors into the ILM will require going beyond prior research as both the experience of emotions and the ways in which those experiences are supported to promote learning will need to consider more factors (e.g., student interest, cultural background). In the assessment context, the use of student emotions can be expanded to provide caring support after the assessment by providing context for a student's performance to both the teacher and the student (e.g., student was confused while responding to items 2, 5, and 7), which can allow for more informed instructional decisions. In the CA framework, the ways in which student emotions are leveraged to support student learning will build upon prior learning research and will require new research efforts to ensure that emotions are productively integrated with other individual differences.

4 Individual differences

Students enter into test-taking experiences with a wide variety of interests, prior knowledge, experiences, attitudes, motivations, dispositions, or other intra-or interpersonal qualities that can affect their engagement with and performance on educational assessments and other academic outcomes (Braun et al., 2009; Lipnevich et al., 2013; Duckworth and Yaeger, 2015; West et al., 2016; Abrahams et al., 2019). For example, self-efficacy beliefs are strongly linked to academic achievement across domains (Guthrie and Wigfield, 2000; Richardson et al., 2012; Schneider and Preckel, 2017). Understanding how individual differences influence performance in interactive learning environments suggests directions for interventions or dynamic supports (Self, 1999) based on cognitive or motivational variables (Du Boulay et al., 2010) or prior knowledge (Khayi and Rus, 2019) that can be applied in assessments.

In previous work, we investigated student characteristics that predict performance on innovative conversation-based assessments of science inquiry and mathematical argumentation (Sparks et al., 2019, 2022). Students' science self-efficacy, growth mindset, cognitive flexibility, and test anxiety (with a negative coefficient) predicted performance on a science assessment (Sparks et al., 2019), while cognitive flexibility and perseverance (with a negative coefficient) predicted performance on mathematical argumentation (Sparks et al., 2022), controlling for student demographics and domain skills. Cluster analyses resulted in interpretable profiles with distinct relationships to student characteristics and performance, suggesting distinct paths for caring support within the CA framework (Sparks et al., 2020). For example, one profile represented students with average domain ability but relatively low cognitive flexibility, while another reflected motivated but test-anxious students. We hypothesize that these profiles would benefit from different supports (i.e., motivational messages vs. anxiety-reduction strategies; Arslan and Finn, 2023). However, the profiles and associated supports must be developed and validated in future research with students and teachers to ensure that the profiles reflect, and the adaptations address, the aspects most meaningful for instruction.

5 Cultural contexts

The prominence of social justice and anti-racist movements has resulted in increasing or renewed interest in (socio-)culturally responsive assessment (SCRA) practices (Hood, 1998; Lee, 1998; Qualls, 1998; Sireci, 2020; Bennett, 2022, 2023; Randall, 2021) which are themselves grounded in culturally relevant, responsive, and sustaining pedagogies (Paris, 2012; Gay, 2013; Ladson-Billings, 2014). Recent research reflects increasing attention to students' cultural characteristics when designing and evaluating AI-enabled instructional systems (Blanchard and Frasson, 2005; Mohammed and Watson, 2019; Talandron-Felipe, 2021); we can apply lessons from this work toward digital assessment design. As the K-12 student population becomes increasingly demographically, culturally, and linguistically diverse (National Center for Education Statistics, 2022), educational assessments must account for such variation, enabling test-takers to demonstrate their knowledge, skills, and abilities in ways that are most appropriate considering their cultural, linguistic, and social contexts (Mislevy, 2018; Sireci and Randall, 2021). Test items can include content reflective of situations, contexts, and practices students encounter in their lives (Randall, 2021), which can tap into students' home and community funds of knowledge (Moll et al., 1992; González et al., 2005) in ways that foster deeper student learning through meaningful connections to familiar, interesting contexts (Walkington and Bernacki,

2018). Math problems assessing knowledge of fractions within a recipe context could vary the context to align with students' cultural background (e.g., beans and cornbread vs. peanut butter sandwich). Positive effects have been shown for African American students interacting with pedagogical agents that employ dialects similar to their own in personalized learning systems (Finkelstein et al., 2013).

Emerging work is exploring cultural responsiveness in the context of scenario-based assessments (SBAs). SBAs are a useful context for exploring cultural factors in assessment performance and potential for implementing personalization within the CA framework (Sparks et al., 2023a,b). SBAs intentionally situate students in meaningful contexts for problem solving, providing a purpose and goal for responding to items (Sabatini et al., 2019). SBA developers have emphasized how scenarios can be made relevant to students from diverse racial, ethnic, and cultural backgrounds by intentionally incorporating contexts and content that celebrate students' cultural identities and integrate funds of knowledge from an asset-based perspective (O'Dwyer et al., 2023). Similar work has been conducted in designing robots for educational purposes in which students serve as co-creators to enable cultural relevance and responsiveness (Li et al., 2023). For example, SBA topics with greater cultural relevance to Black students (i.e., the Harlem Renaissance) show comparable reliability and validity but smaller group differences in performance versus more general topics (Ecosystems, Immigration), potentially due to Black students' greater engagement (Wang et al., 2023). Our current research (Sparks et al., 2023a,b) involves measuring students' self-identified cultural characteristics to examine relationships among their engagement and performance on SBAs, their racial, ethnic, and cultural identities, as well as their emotions, interests, motivations, prior knowledge, and experiences (i.e., home and community experiences, values, and practices related to assessment topics; Lave and Wenger, 1991; Gutiérrez and Rogoff, 2003; González et al., 2005). In future research, we aim to incorporate these characteristics into student profiles and evaluate how the profiles can be leveraged to provide a personalized assessment experience. This combination of cultural responsiveness and personalization has been explored in the learning context (Blanchard, 2010); however, additional research is needed to understand these dynamics to provide caring support within assessments.

6 Challenges and opportunities for caring assessment

Personalization within a CA framework introduces several challenges as well as opportunities when considering implementation of this framework within a digital learning system. The holistic view of students reflected in the ILM – going beyond measures of cognitive skill or performance to incorporate emotions, motivations, knowledge, interest, and other characteristics – requires access to data that is not typically collected during educational assessments (Zapata-Rivera, 2017). Contextual variables are often collected via survey methods (e.g., Braun et al., 2009; Abrahams et al., 2019) but could increasingly be collected by other means such as embedded assessment (Zapata-Rivera and Bauer, 2012; Zapata-Rivera, 2012; Rausch et al., 2019), and *stealth* assessment (Shute et al., 2009, 2015; Shute and Ventura, 2013) approaches which use logfile data from the assessment interaction and are less intrusive. For example, student

interest could be measured by utilizing time-on-task and clickstream behaviors, versus a survey. Such approaches may collect multimodal interaction data (e.g., audio, or visual data) and leverage this information in an ILM. Collection of such multimodal data introduces the potential for privacy concerns regarding what is being collected, where data is stored, and who has access, especially to the extent that Personally Identifiable Information (PII) may be collected. Policy prohibitions may prevent collection and storage of certain data types (Council of the European Union, 2023). The importance of ethical and secure data handling and transparency with users about what and how data will be collected, retained, and used, is paramount, especially for K-12 students. Thus, implementation of the CA framework will require innovative measurement and modeling methodologies as well as close collaboration with students and teachers to build trust. Much like the ILM, it will be critical to integrate these independent lines of work in new research efforts that apply the CA framework in practice. Such integrated research is being actively explored in the INVITE institute[1] toward development of "caring" STEM learning environments for K-12 students.

A further challenge relates to the inherent tradeoffs in selecting the key student characteristics and behaviors that should be used to implement personalization. Variable selection requires care to ensure that measures are reliable and appropriate, so that personalization can be implemented along the dimensions that are most pertinent to students' needs. However, this challenge also inspires new research opportunities – particularly ones that focus on students that have been historically underrepresented in both research and educational technology to determine what characteristics and behaviors are most relevant for different student groups. Research that is more inclusive and aware of the diverse experiences that students bring to personalized digital assessment and learning experiences can support effective variable selection. Open learner modeling approaches (Bull and Kay, 2016; Bull, 2020; Zapata-Rivera, 2020) introduce an opportunity to further refine CAs while building user trust by giving teachers and students the chance to inspect and reflect on the ILM, highlighting where the model and its interpretations should be revised or qualified. Development of the infrastructure needed to collect variables, classify behaviors, deploy adaptations, and continually update a caring system requires computational modeling, machine learning, and artificial intelligence expertise to help develop, test, and iterate on the learner models. ILMs can be leveraged toward effective decision cycles within the caring system that, for example, provide necessary supports, route students to appropriate versions of subsequent tasks, and provide tailored, asset-based feedback.

A related issue concerns teachers' perceptions of personalization and whether they prioritize mastery of content or embrace a more holistic view and a need to personalize based on a broader set of emotional, motivational, or cultural aspects. The effectiveness of CAs will rest on their ability to effectively integrate with teacher practice by supporting students with different constellations of strengths and challenges, detecting for teachers the students who are most in need of their additional attention and support. Again, this challenge offers an opportunity for new research that incorporates teachers into the research and development process to bring CAs into practice that are reflective of

---

1   https://invite.illinois.edu

current best practices and work with teachers in achieving the shared goal of student learning in a caring and supportive environment.

Integrating cultural responsiveness into the CA framework introduces additional challenges. While personalization implies treating students as individuals, culturally situated perspectives emphasize how individual students are positioned as members of socially-and historically-defined racial, ethnic, and cultural groups (Gutiérrez and Rogoff, 2003). Such views acknowledge that groups are not monolithic and that identification with the racial, ethnic, and cultural contexts individual students experience also varies (Tatum, 2017). Adapting at the group level necessitates acknowledgment of this individual variation as well as the potential for individuals to identify in ways that may (not) be congruent with demographic group membership. Demographics may intersect in meaningful ways that impact students' lived experiences (Crenshaw, 1989). However, culture is embodied in participation in practices with shared meaning and significance (Lave and Wenger, 1991; Gutiérrez and Rogoff, 2003; Nasir et al., 2014). This implies that CA should enable student self-identification of demographic characteristics, cultural group memberships, and engagement in home and community practices (i.e., in terms of their funds of knowledge). Further research is needed to best understand how the complexity of student identities interact and impact their learning experiences.

Intersections among students' cultural backgrounds, knowledge, and experiences might be leveraged to increase the relevance and responsiveness of assessments (Walkington and Bernacki, 2018). Meaningful co-design activities in which the knowledge, interests, values, and experiences of students and teachers from historically marginalized groups can be centered, celebrated, and prioritized has the potential to result in more engaging, relevant, and valid assessments and would support more responsive personalized designs (O'Dwyer et al., 2023; Ober et al., 2023). Open learner models that can be interrogated and critiqued by students and teachers will be essential for a culturally responsive CA framework, so that student profiles and ILMs do not reflect biases or stereotypes, that misclassifications are appropriately corrected, and that contextual factors are considered when interpreting students' performance. Continued partnerships with teachers and students are needed to maximize the benefits for learning through connections to students' funds of knowledge while also minimizing unintended consequences.

7 Discussion

The CA framework can be leveraged toward personalized and culturally responsive assessments designed to support K-12 teaching and learning. This article outlines the current state of CA research on student emotions, individual differences, and cultural contexts, and highlights key challenges and opportunities for future research. Critical issues for future research include collection and handling of student data (characteristics, behavioral, multimodal) and associated privacy and security concerns, selection of characteristics for learner modeling, teacher perceptions of personalization, individual variation and self-identification of students' cultural identities and contexts, and engaging students and teachers in co-design of personalized ILMs and responsive adaptations. Research that integrates these independent areas is needed to bring the CA conceptual framework into practice in personalized digital assessments.

Whether the primary aim is individual personalization or responsiveness to students' cultural contexts, it is imperative that researchers engage in deep, sustained co-design partnerships with teachers and students to ensure validity and utility for those most in need of support (Penuel, 2019). It is also important to consider the assessment context (e.g., formative vs. summative, group-vs. individual-level reporting) and implications for measurement (e.g., comparability, scoring, interpretation) when determining how best to apply CA in practice. CA introduces opportunities to enhance students' assessment experiences and to advance use of assessment outcomes to further individuals' educational opportunities and wellbeing (Bittencourt et al., 2023). However, effective design and implementation of personalized assessments is a complex endeavor, which may necessitate new processes for designing assessments (O'Dwyer et al., 2023). We invite other scholars to conduct research addressing these challenges, advancing the field's ability to provide personalized, culturally responsive assessments.

Author contributions

JRS, BL, and DZ-R contributed to manuscript conceptualization, writing, reviewing, and editing prior to submission. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

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

Publisher's note

Author disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the U.S. Department of Education.

References

Abrahams, L., Pancorbo, G., Santos, D., John, O. P., Primi, R., Kyllonen, P., et al. (2019). Social-emotional skill assessment in children and adolescents: advances and challenges in personality, clinical, and education contexts. *Psychol. Assess.* 31, 460–473. doi: 10.1037/pas0000591

Arslan, B., and Finn, B. (2023). The effects of nudges on cognitively disengaged student behavior in low-stakes assessments. *J. Intelligence* 11:204. doi: 10.3390/jintelligence11110204

Baker, R., D'Mello, S. K., Rodrigo, M. M. T., and Graesser, A. C. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *Int. J. Hum. Comput. Stud.* 68, 223–241. doi: 10.1016/j.ijhcs.2009.12.003

Bellarhmouch, Y., Jeghal, A., Tairi, H., and Benjelloun, N. (2023). A proposed architectural learner model for a personalized learning environment. *Educ. Inf. Technol.* 28, 4243–4263. doi: 10.1007/s10639-022-11392-y

Bennett, R. E. (2022). *The good side of COVID-19. Educational Measurement: Issues and Practice.* 41, 61–63. doi: 10.1111/emip.12496

Bennett, R. E. (2023). Toward a theory of socioculturally responsive assessment. *Educ. Assess.* 28, 83–104. doi: 10.1080/10627197.2023.2202312

Bittencourt, I. I., Chalco, G., Santos, J., Fernandes, S., Silva, J., Batista, N., et al. (2023). Positive artificial intelligence in education (P-AIED): a roadmap. *Int. J. Artif. Intell. Educ.* doi: 10.1007/s40593-023-00357-y

Blanchard, E. G. (2010). "Adaptation-oriented culturally-aware tutoring systems: when adaptive instructional technologies meet intercultural education" in *Handbook of research on human performance and instructional technology.* eds. H. Song and T. Kidd. IGI Global. 413–430. doi: 10.4018/978-1-60566-782-9.ch025

Blanchard, E. G. M., and Frasson, C. (2005). Making intelligent tutoring systems culturally aware: the use of Hofstede's cultural dimensions. In: *Proceedings of the 2005 International Conference on Artificial Intelligence (ICAI 2005).* eds. H. R. Arabnia and R. Joshua. Las Vegas, Nevada: CSREA Press. 2, 644–649.

Braun, H., Coley, R., Jia, Y., and Trapani, C. (2009). "Exploring what works in science instruction: a look at the eighth-grade science classroom" in *ETS policy information report* (Princeton, NJ: ETS). Available at: https://www.ets.org/Media/Research/pdf/PICSCIENCE.pdf

Bull, S. (2020). There are open learner models about! *IEEE Trans. Learn. Technol.* 13, 425–448. doi: 10.1109/TLT.2020.2978473

Bull, S., and Kay, J. (2016). SMILI☺: a framework for interfaces to learning data in open learner models learning analytics and related fields. *Int. J. Artif. Intell. Educ.* 26, 293–331. doi: 10.1007/s40593-015-0090-8

Council of the European Union. (2023). Artificial intelligence act: council and parliament strike a deal on the first rules for AI in the world. Press release. Available at: https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/

Craig, S., Graesser, A., Sullins, J., and Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning. *J. Educ. Media* 29, 241–250. doi: 10.1080/1358165042000283101

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *Univ. Chic. Leg. Forum* 1989, Available at: http://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8

D'Mello, S. K. (2013). A selective Meta-analysis on the relative incidence of discrete affective states during learning with technology. *J. Educ. Psychol.* 105, 1082–1099. doi: 10.1037/a0032674

D'Mello, S., and Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognit. Emot.* 25, 1299–1308. doi: 10.1080/02699931.2011.613668

D'Mello, S., and Graesser, A. (2012). Dynamics of affective states during complex learning. *Learn. Instr.* 22, 145–157. doi: 10.1016/j.learninstruc.2011.10.001

D'Mello, S., and Graesser, A. (2014). Confusion and its dynamics during device comprehension with breakdown scenarios. *Acta Psychol.* 151, 106–116. doi: 10.1016/j.actpsy.2014.06.005

D'Mello, S. K., Lehman, B. A., and Graesser, A. C. (2011). "A motivationally supportive affect-sensitive AutoTutor" in *New perspectives on affect and learning technologies.* eds. R. A. Calvo and S. K. D'Mello (New York: Springer), 113–126.

D'Mello, S., Lehman, B., Pekrun, R., and Graesser, A. (2014). Confusion can be beneficial for learning. *Learn. Instruct.* 29, 153–170. doi: 10.1016/j.learninstruc.2012.05.003

Du Boulay, B. (2018). Intelligent tutoring systems that adapt to learner motivation. In *Tutor. Intell. Tutor. Syst.*, Ed. S. D. Craig. New York: Nova Science Publishers, Inc. 103–128.

Du Boulay, B., Avramides, K., Luckin, R., Martinuz-Miron, E., Rebolledo Mendez, G., and Carr, A. (2010). Towards systems that care: a conceptual framework based on motivation, metacognition and affect. *Int. J. Artif. Intell. Educ.* 20, 197–229. doi: 10.5555/1971783.1971784

Duckworth, A. L., and Yaeger, D. S. (2015). Measurement matters: assessing personal qualities other than cognitive ability for educational purposes. *Educ. Res.* 44, 237–251. doi: 10.3102/0013189X15584327

Finkelstein, S., Yarzebinski, E., Vaughn, C., Ogan, A., and Cassell, J. (2013). "The effects of culturally congruent educational technologies on student achievement" in *Proceedings of 16th international conference on artificial intelligence in education (AIED2013).* eds. K. Yacef, C. Lane, J. Mostow and P. Pavlik (Berlin Heidelberg: Springer-Verlag), 493–502.

Forbes-Riley, K., and Litman, D. (2011). Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Comm.* 53, 1115–1136. doi: 10.1016/j.specom.2011.02.006

Gay, G. (2013). Teaching to and through cultural diversity. *Curric. Inq.* 43, 48–70. doi: 10.1111/curi.12002

González, N., Moll, L. C., and Amanti, C. (Eds.) (2005). *Funds of knowledge: theorizing practices in households, communities, and classrooms.* Lawrence Erlbaum associates publishers.

Grafsgaard, J. F., Boyer, K. E., Wiebe, E. N., and Lester, J. C. (2012). Analyzing posture and affect in task-oriented tutoring. In *Proceedings of the Intelligent Tutoring Systems Track of the Twenty-Fifth International Conference of the Florida Artificial Intelligence Research Society (FLAIRS).* Marco Island, FL. Washington, DC: AAAI. 438–443.

Guthrie, J. T., and Wigfield, A. (2000). "Engagement and motivation in reading" in *Handbook of reading research.* eds. M. L. Kamil, P. B. Mosenthal, P. D. Pearson and R. Barr (Mahwah, NJ: Erlbaum).

Gutiérrez, K. D., and Rogoff, B. (2003). Cultural ways of learning: individual traits or repertoires of practice. *Educ. Res.* 32, 19–25. doi: 10.3102/0013189X032005019

Hood, S. (1998). Culturally responsive performance-based assessment: conceptual and psychometric considerations. *J. Negro Educ.* 67, 187–196. doi: 10.2307/2668188

Kay, J., Bartimote, K., Kitto, K., Kummerfeld, B., Liu, D., and Reimann, P. (2022). Enhancing learning by open learner model (OLM) driven data design. *Comput. Educ. Artif. Intell.* 3:100069. doi: 10.1016/j.caeai.2022.100069

Kay, J., and McCalla, G. (2003). The careful double vision of self. *Int. J. Artif. Intell. Educ.* 13, 1–18,

Khayi, N. A., and Rus, V. (2019). Clustering students based on their prior knowledge. In M. Desmarais, C. F. Lynch, A. Merceron and R. Nkambou (Eds.), Proceedings of the 12th international conference on educational data mining.

Kizilcec, R. F., Pérez-Sanagustín, M., and Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Comput. Educ.* 104, 18–33. doi: 10.1016/j.compedu.2016.10.001

Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: a.k.a. the remix. *Harv. Educ. Rev.* 84, 74–84. doi: 10.17763/haer.84.1.p2rj131485484751

Lave, J., and Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press.

Lee, C. D. (1998). Culturally responsive pedagogy and performance-based assessment. *J. Negro Educ.* 67, 268–279. doi: 10.2307/2668195

Lee, D., Rodrigo, M., Baker, R., Sugay, J., and Coronel, A. (2011). "Exploring the relationship between novice programmer confusion and achievement" in *ACII 2011.* eds. S. D'Mello, A. Graesser, B. Schuller and J. Martin (Berlin/Heidelberg: Springer), 175–184.

Lehman, B., and Graesser, A. (2015). "To resolve or not to resolve? That is the big question about confusion" in *Proceedings of the international conference on artificial intelligence in education.* eds. C. Conati, N. Heffernan, A. Mitrovic and M. F. Verdejo (Springer: Verlag), 216–225.

Lehman, B., Sparks, J. R., and Zapata-Rivera, D. (2018). "When should an adaptive assessment care?" in *Proceedings of ITS 2018: Intelligent tutoring systems 14th international conference, workshop on exploring opportunities for caring assessments.* eds. N. Guin and A. Kumar (Montreal, Canada: ITS), 87–94.

Lehman, B., and Zapata-Rivera, D. (2018). "Frequency, intensity, and mixed emotions, oh my! Investigating control-value theory in conversation-based assessments" *Paper presented at the annual meeting of the American Educational Research Association.* New York, NY.

Li, Y., Nwogu, J., Buddemeyer, A., Soloist, J., Lee, J., Walker, E., et al. (2023). "I want to be unique from other robots": positioning girls as co-creators of social robots in culturally-responsive computing education. In *Proceedings of the 2023 CHI conference on human factors in computing systems (CHI '23), Article No. 441.* eds. A. Schmidt, K. Vaananen, T. Goyal, P. O. Kristensson, A. Peters, S. Mueller, J. R. Williamson, et al. New York, NY: ACM, 1–14.

Lipnevich, A. A., MacCann, C., and Roberts, R. D. (2013). Assessing non-cognitive constructs in education: a review of traditional and innovative approaches. In *Oxford Handbook Child Psychol. Assess.* eds. D. H. Saklofske, C. R. Reynolds and V. Schwean. New York: Oxford University Press. 750–772. doi: 10.1093/oxfordhb/9780199796304.013.0033

Liu, Z., Pataranutaporn, V., Ocumpaugh, J., and Baker, R. (2013). "Sequences of frustration and confusion, and learning" in *EDM 2013.* eds. S. D'Mello, R. Calvo and A. Olney. *Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013).* International Educational Data Mining Society. 114–120.

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. New York: Routledge. doi: 10.4324/9781315871691

Mohammed, P. S., and Watson, E. N. (2019). "Towards inclusive education in the age of artificial intelligence: perspectives, challenges, and opportunities" in *Artificial intelligence and inclusive education. Perspectives on rethinking and reforming education. Book series on perspectives on rethinking and reforming education.* eds. J. Knox, Y. Wang and M. Gallagher (Singapore: Springer) 17–37. doi: 10.1007/978-981-13-8161-4_2

Moll, L. C., Amanti, C., Neff, D., and Gonzalez, N. (1992). Funds of knowledge for teaching: using a qualitative approach to connect homes and classrooms. *Theory Pract.* 31, 132–141.

Nasir, N. S., Rosebery, A. S., Warren, B., and Lee, C. D. (2014). "Learning as a cultural process: achieving equity through diversity" in *The Cambridge handbook of the learning sciences*. 2 Edition ed. R. K. Sawyer (Cambridge, UK: Cambridge University Press), 686–706. doi: 10.1017/CBO9781139519526.041

National Center for Education Statistics (2022). *Racial/ethnic enrollment in public schools. Condition of education.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

O'Dwyer, E., Sparks, J. R., and Nabors Oláh, L. (2023). Enacting a process for developing culturally relevant classroom assessments. *Appl. Meas. Educ.* 36, 286–303. doi: 10.1080/08957347.2023.2214652

Ober, T., Lehman, B., Gooch, R., Oluwalana, O., Solyst, J., Garcia, A., et al. (2023). Development of a framework for culturally responsive personalized learning. Paper presented at the annual meeting of the International Society for Technology in Education (ISTE). Philadelphia, PA.

Paris, D. (2012). Culturally sustaining pedagogy: a needed change in stance, terminology, and practice. *Educ. Res.* 41, 93–97. doi: 10.3102/0013189X12441244

Pekrun, R. (2006). The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educ. Psychol. Rev.* 18, 315–341. doi: 10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Frenzel, A., Barchfield, P., and Perry, R. (2011). Measuring emotions in students' learning and performance: the achievement emotions questionnaire (AEQ). *Contemp. Educ. Psychol.* 36, 36–48. doi: 10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Perry, R., Kramer, K., Hochstadt, M., and Molfenter, S. (2004). Beyond test anxiety: development and validation of the test emotions questionnaire (TEQ). *Anxiety Stress Coping* 17, 287–316. doi: 10.1080/10615800412331303847

Penuel, W. R. (2019). "Co-design as infrastructuring with attention to power: building collective capacity for equitable teaching and learning through design-based implementation research" in *Collaborative curriculum Design for Sustainable Innovation and Teacher Learning*. eds. J. Pieters, J. Voogt and N. P. Roblin (Cham, Switzerland: SpringerOpen), 387–401.

Pintrich, P. R. (2000). "The role of goal orientation in self-regulated learning" in *Handbook of Self-regulation*. eds. M. Boekaerts, P. R. Pintrich and M. Zeidner (San Diego, CA: Academic Press), 452–502.

Qualls, A. L. (1998). Culturally responsive assessment: development strategies and validity issues. *J. Negro Educ.* 67, 296–301. doi: 10.2307/2668197

Ramasubramanian, S., Riewestahl, E., and Landmark, S. (2021). The trauma-informed equity-minded asset-based model (TEAM): the six R's for social justice-oriented educators. *J. Media Lit. Educ.* 13, 29–42. doi: 10.23860/JMLE-2021-13-2-3

Randall, J. (2021). Color-neutral is not a thing: redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educ. Meas. Issues Pract.* 40, 82–90. doi: 10.1111/emip.12429

Rausch, A., Kögler, K., and Siegfried, J. (2019). Validation of embedded experience sampling (EES) for measuring non-cognitive facets of problem-solving competence in scenario-based assessments. *Front. Psychol.* 10, 1–16. doi: 10.3389/fpsyg.2019.01200

Richardson, M., Abraham, C., and Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol. Bull.* 138, 353–387. doi: 10.1037/a0026838

Sabatini, J., O'Reilly, T., Weeks, J., and Wang, Z. (2019). Engineering a 21st century reading comprehension assessment system utilizing scenario-based assessment techniques. *Int. J. Test.* 20, 1–23. doi: 10.1080/15305058.2018.1551224

Schneider, M., and Preckel, F. (2017). Variables associated with achievement in higher education: a systematic review of meta-analyses. *Psychol. Bull.* 143, 565–600. doi: 10.1037/bul0000098

Self, J. A. (1999). The distinctive characteristics of intelligent tutoring systems research: ITSs care, precisely. *Int. J. Artif. Intell. Educ.* 10, 350–364,

Shemshack, A., Kinshuk, , and Spector, J. M. (2021). A comprehensive analysis of personalized learning components. *J. Comput. Educ.* 8, 485–503. doi: 10.1007/s40692-021-00188-7

Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., et al. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Comput. Educ.* 86, 224–235. doi: 10.1016/j.compedu.2015.08.001

Shute, V. J., and Ventura, M. (2013). *Measuring and supporting learning in games: Stealth assessment*. Cambridge, MA: The MIT Press.

Shute, V. J., Ventura, M., Bauer, M. I., and Zapata-Rivera, D. (2009). "Melding the power of serious games and embedded assessment to monitor and foster learning: flow and grow" in *Serious games: mechanisms and effects*. eds. U. Ritterfeld, M. Cody and P. Vorderer (Mahwah, NJ: Routledge, Taylor and Francis), 295–321.

Sireci, S. G. (2020). Standardization and UNDERSTANDardization in educational assessment. *Educ. Meas. Issues Pract.* 39, 100–105. doi: 10.1111/emip.12377

Sireci, S. G., and Randall, J. (2021). "Evolving notions of fairness in testing in the United States" in *The history of educational measurement: Key advancements in theory, policy, & practice*. eds. B. E. Clauser and M. B. Bunch (New York: Routledge), 111–135.

Spangler, G., Pekrun, R., Kramer, K., and Hofmann, H. (2002). Students' emotions, physiological reactions, and coping in academic exams. *Anxiety Stress Coping* 15, 413–432. doi: 10.1080/1061580021000056555

Sparks, J. R., Lehman, B., Ober, T., Zapata-Rivera, D., Steinberg, J., and McCulla, L. (2023b). *Linking students' funds of knowledge to task engagement and performance: toward culturally responsive "caring assessments"*. Paper presented in coordinated symposium session at the 7th Culturally Responsive Evaluation and Assessment Conference, Chicago, IL.

Sparks, J.R., Ober, T., Steinberg, J., Lehman, B., McCulla, L., and Zapata-Rivera, D. (2023a). *PK and IDK are OK, IMO: correlating students' knowledge, metacognition, and perceptions of scenario-based assessments*. Poster presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Sparks, J. R., Peters, S., Steinberg, J., James, K., Lehman, B. A., and Zapata-Rivera, D. (2019). *Individual difference measures that predict performance on conversation-based assessments of science inquiry skills*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada.

Sparks, J. R., Steinberg, J., Castellano, K., Lehman, B., and Zapata-Rivera, D. (2020). *Generating individual difference profiles via cluster analysis: toward caring assessments for science*. Poster accepted at the annual meeting of the National Council for Measurement in Education, San Francisco, CA. (Conference canceled).

Sparks, J. R., Steinberg, J., Lehman, B., and Zapata-Rivera, D. (2022). *Leveraging students' background characteristics to predict performance on conversation-based assessments of mathematics*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Talandron-Felipe, M. M. P. (2021). "Considerations towards culturally-adaptive instructional systems" in *Adaptive instructional systems. Design and evaluation. HCII 2021*. eds. R. A. Sottilare and J. Schwarz, Lecture Notes in Computer Science (Cham: Springer), 12792.

Tatum, B. D. (2017). *Why are all the black kids sitting together in the cafeteria? And other conversations about race. Revised* Edn. New York: Basic Books.

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

Walkington, C., and Bernacki, M. L. (2018). Personalization of instruction: design dimensions and implications for cognition. *J. Exp. Educ.* 86, 50–68. doi: 10.1080/00220973.2017.1380590

Wang, Z., Bruce, K. M., O'Reilly, T. P., Sparks, J. R., and Walker, M. (2023). *Group differences across scenario-based reading assessments: examining the effects of culturally relevant test content*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Weitekamp, D., and Koedinger, K. (2023). Computational models of learning: deepening care and carefulness in AI in education. In: N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda and O.C Santos. (Eds.), Artificial intelligence in education. Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky. AIED 2023. Communications in Computer and Information Science, vol 1831. Cham, Switzerland: Springer. doi: 10.1007/978-3-031-36336-8_2

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., et al. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educ. Eval. Policy Analysis* 38, 148–170,

Winne, P. H., and Hadwin, A. F. (1998). "Studying as self-regulated engagement in learning" in *Metacognition in educational theory and practice*. eds. D. Hacker, J. Dunlosky and A. Graesser (Hillsdale, NJ: Erlbaum), 277–304.

Wise, S. L., Bhola, D., and Yang, S. (2006). Taking the time to improve the validity of low-stakes tests: the effort-monitoring CBT. *Educ. Meas. Issues Pract.* 25, 21–30. doi: 10.1111/j.1745-3992.2006.00054.x

Wise, S. L., and DeMars, C. E. (2006). An application of item response time: the effort-moderated IRT model. *J. Educ. Meas.* 43, 19–38. doi: 10.1111/j.1745-3984.2006.00002.x

Wise, S. L., Kuhfeld, M. R., and Soland, J. (2019). The effects of effort monitoring with proctor notification on test-taking engagement, test performance, and validity. *Appl. Meas. Educ.* 32, 183–192. doi: 10.1080/08957347.2019.1577248

Yadegaridehkordi, E., Noor, N. F. B. M., Ayub, M. N. B., Affal, H. B., and Hussin, N. B. (2019). Affective computing in education: a systematic review and future research. *Comput. Educ.* 142:103649. doi: 10.1016/j.compedu.2019.103649

Zapata-Rivera, D., and Bauer, M. (2012). "Exploring the role of games in educational assessment" in *Technology-based assessments for twenty-first-century skills: theoretical and practical implications from modern research*. eds. J. Clarke-Midura, M. Mayrath, D. Robinson and G. Schraw. Charlotte, NC: Information Age. 147–169.

Zapata-Rivera, D., Forsyth, C., Sparks, J. R., and Lehman, B. (2023). "Conversation-based assessment: current findings and future work" in *International encyclopedia of education*. eds. R. Tierney, F. Rizvi and K. Ercikan. *4th* ed (Amsterdam, Netherlands: Elsevier Science) 504–518.

Zapata-Rivera, D. (2012) Embedded assessment of informal and afterschool science learning. Summit on Assessment of Informal and After-School Science Learning Available at: https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_072564.pdf

Zapata-Rivera, D. (2017). Toward caring assessment systems. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization (UMAP '17)*. eds. M. Tkalcic, D. Thakker, P. Germanakos, K. Yacef, C. Paris and O. Santos. New York: ACM. 97–100.

Zapata-Rivera, D. (2020). Open student modeling research and its connections to educational assessment. *Int. J. Artif. Intell. Educ.* 31, 380–396. doi: 10.1007/s40593-020-00206-2

Zapata-Rivera, D., and Arslan, B. (2021). "Enhancing personalization by integrating top-down and bottom-up approaches to learner modeling" in *Proceedings of HCII 2021*. eds. R. A. Sottilare and J. Schwarz (Switzerland: Springer Nature), 234–246.

Zapata-Rivera, D., Lehman, B., and Sparks, J. R. (2020). "Learner modeling in the context of caring assessments" in *Proceedings of the second international conference on adaptive instructional systems, held as part of HCI international conference 2020, LNCS 12214*. eds. R. A. Sottilare and J. Schwarz (Cham, Switzerland: Springer Nature), 422–431.