SmarterAI Think Tank:

Student-Centric Design Principles for Responsible Use of AI

May 2025

Table of Contents

SmarterAI Think Tank: Student-Centric Design Principles for Responsible Use Of AI

Smarter Balanced collaborated with IBM Consulting in early 2024 to draft Student-Centric Design Principles for Responsible Use of AI when considering the process for designing and integrating AI in large-scale assessment. These principles serve as a resource to the educational measurement community. The Student-Centric Design Principles offer a foundation for considering ongoing conversations and decisions related to how AI is best leveraged in assessment system designs. Consistent with the foundational principles of the Smarter Balanced Assessment Consortium, Smarter Balanced engaged a variety of interest holders in this process to minimize partiality and promote accessibility.

Smarter Balanced convened several multi-disciplinary work groups that included experts in educational assessment and measurement; those serving in state, district, and local educational contexts; higher education experts; and policymakers. In collaboration with IBM Consulting, we used a design-thinking methodology. The methodology and specific outcomes associated with a selected use case are explored in Appendix A. The process was iterative over several months and included various interest holder groups.

Eight Student-Centric Design Principles for Responsible Use of AI are presented as outcomes of this process. The initial draft was developed iteratively over several months and with varying groups to manage a proposed use case that serves to demonstrate how the principles can be implemented within a large-scale educational assessment context. Through the development process, it became clear that these principles also apply to broader aspects of assessment development, not only to the use of AI. Similarly, while the initial purpose of this process was focused on assessment development, the conversations also addressed other aspects of assessment system design such as test delivery, student engagement, scoring, and reporting.

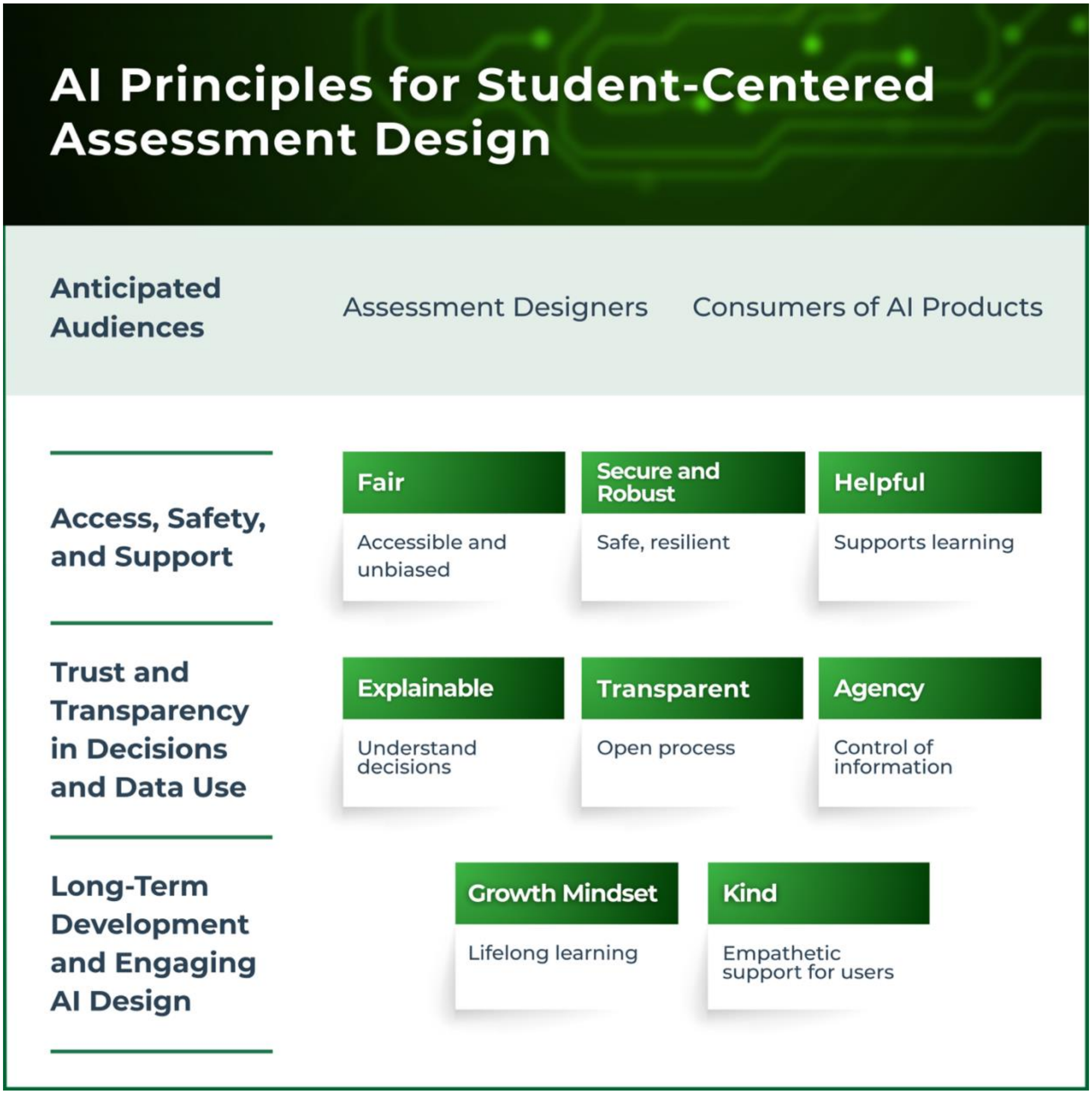
To solicit additional feedback on the Student-Centric Design Principles, presentations were conducted with external interest holders at conferences and other invited sessions. These sessions included the Council of Chief State School Officers (CCSSO) National Conference on Student Assessment (NCSA), the National Assessment Governing Board, and IBM Think 2024.

The principles described below represent not only the contributions of the SmarterAI Think Tank members, but also a refinement based on conversations and presentations with a larger body of interest holders to help make the principles useful for a broad set of large-scale educational assessments, so that the principles might better serve a wider audience.

Smarter Balanced Student-Centric Design Principles for Responsible Use Of AI

The eight Student-Centric Design Principles for Responsible Use of AI were adapted from IBM's Artificial Intelligence Pillars (2023) and applied in the context of AI use cases for student or educator engagement within the context of educational assessments:

Figure 1. Principles



AI Principles for Student-Centered Assessment Design

Anticipated Audiences

Assessment Designers

Consumers of AI Products

Access, Safety, and Support

Fair
Accessible and unbiased

Secure and Robust
Safe, resilient

Helpful
Supports learning

Trust and Transparency in Decisions and Data Use

Explainable
Understand decisions

Transparent
Open process

Agency
Control of information

Long-Term Development and Engaging AI Design

Growth Mindset
Lifelong learning

Kind
Empathetic support for users

Principle 1: Fair

Fair refers to the promotion of accessibility for all student populations. Adapted from IBM (2023), properly calibrated, AI can assist humans in making fairer choices, promoting accessibility and openness. Certain tendencies can be present in the algorithm of the AI system and in the data used to train and test it. Openness means creating a development team and seeking out the perspectives of disadvantaged populations and communities. The premise of including "fair" is to offer fair treatment of individuals, or groups of individuals, when developing and implementing an AI system. An example of fairness includes the building of an AI model through data that represents the communities the model serves, and incorporates representatives of those varied communities in the validation and subsequent improvement of the AI systems. The *Standards for Educational and Psychological Testing* (2014) Standard 3: Fairness in Testing note, "Test tasks and items should be designed to maximize access and be free of construct-irrelevant barriers as far as possible for all individuals and relevant subgroups in the intended test-taker population." Fairness promotes the engagement of all the users in the assessment system.

Principle 2: Secure and Robust

Secure and Robust describes how developed AI-powered systems actively defend against adversarial attacks, minimize security risks, and enable confidence in system outcomes. Robust AI effectively handles exceptional conditions, such as abnormalities in input or malicious attacks, without causing unintentional harm to any user group. For example, external compromises such as viruses or bots must be anticipated and protected against by specific strategies. According to The *Standards for Educational and Psychological Testing* (2014) Standard 6: Test Administration, Scoring, Reporting, and Interpretation, "Organizations and individuals who maintain or use confidential information about test takers or their scores should have and implement an appropriate policy for maintaining security and integrity of the data, including protecting from accidental or deliberate modification as well as preventing loss or unauthorized destruction." Technology inherently is vulnerable to outside attacks and requires specific action to protect, store, and manage the test events and subsequent reporting steps.

Principle 3: Helpful

Helpful refers to the models where there is a measurable improvement in the assessment system. The effects of utilizing a model should create a measurable improvement to the student's experience and engagement; create efficiencies in managing the content presented to the student; and support educators, students, and families interpreting assessment results. As an example, the AI may suggest or offer insights not currently available during traditional educational assessment to offer suggestions to the students about the content, such as scaffolding, or provide additional perspectives during the activity. The *Standards for Educational and Psychological Testing* (2014) Standard 3.11: Fairness in Testing offer, "When a test is changed to remove barriers to the accessibility of the construct being measured, test developers and/or users are responsible for obtaining and documenting evidence of the validity of score interpretations for intended uses of the changed test." In other words, it is helpful to provide appropriate access to the content for the student as part of the test design and delivery process to reflect outcomes or scores that align to a student’s true understanding of content being assessed.

Principle 4: Explainable

Explainable refers to the degree to which the actions of the AI are consistent with available documentation such that users can anticipate how the AI will act or react as well as how it arrives at particular determinations or conclusions about individuals’ knowledge and skills. AI engagement must be understood by the user and their families (parent/guardian) as appropriate, those responsible for the deployment of the system, and any other individuals who have an interest in the system. The owners and operators should make available — as appropriate and in a context that the relevant end-user can understand — documentation that details essential information for consumers to access. The *Standards for Educational and Psychological Testing* (2014) Standard 13: Uses of Tests for Program Evaluation, Policy Studies, and Accountability describe, “Those responsible for the release or reporting of test results should provide and explain any supplemental information that will minimize possible misinterpretations or misuse of the data.” Additionally, “Score information should be communicated in a way that is accessible to persons receiving the score report.”

Principle 5: Transparent

Transparent refers to the communication and disclosure of information about how the AI was designed and its appropriate uses. Transparent AI systems share information on what data are used, collected, accessed and stored, and reported. For example, information is presented and available about data, engagement, and reporting to users and relevant interest holders. According to The *Standards for Educational and Psychological Testing* (2014) Standard 7: Supporting Documentation for Tests, “Test documentation is effective if it communicates information to user groups in a manner that is appropriate for the particular audience.” Clear information and protocols are essential to ensure user groups have insights into the deployment of the test utilizing AI, use of student and system data, and in the deployment and interpretation of results.

Principle 6: Agency

Agency refers to the ownership of the data within the system and communication about how data may be stored and used. As an example, the AI model should be designed to ensure that the student has confirmed their response and is aware of the implications of the submission as part of the test event. Agency allows more deliberate ownership by students in the process. For example, the test administrator or the system may be designed to explain how the student will engage in the assessment, ways to modify or revise answers, and submit a final test event. The *Standards for Educational and Psychological Testing* (2014) Standard 8: Rights and Responsibilities of Test Takers state, “Test takers have the right to adequate information to help them properly prepare for a test so that the test results accurately reflect their standing on the construct being assessed and lead to fair and accurate score interpretations. They also have the right to protection of their personally identifiable score results from unauthorized access, use, or disclosure.”

Students participating in the test events should understand how their data and information submitted will be used. This insight can allow students to have agency as they engage in submission of individual items and the overall test event considering the overall purpose and use of the assessment.

Principle 7: Growth Mindset

Growth mindset describes a disposition to consider students and adults as lifelong learners who will continue to explore and develop over time. This concept in AI fosters the idea that students may continue to extend their learning during a test event, through AI. Examples may include scaffolding throughout the test event to manage engagement and promote learning beyond the current level of understanding by the student. Specific examples noted in The *Standards for Educational and Psychological Testing* (2014) Standard 1: Validity include, “Questioning test takers from various groups making up the intended test-taking population about their performance strategies or responses to particular items can yield evidence that enriches the definition of a construct. Maintaining records that monitor the development of a response to a writing task, through successive written drafts or electronically monitored revisions, for instance, also provides evidence of process.” Growth mindset in this context goes beyond simply eliciting a response but provides an opportunity for engagement by the student to promote additional learning and exploration of the content.

Principle 8: Kind

Kind describes an aspiration to develop models that establish trust and engagement with the user by incorporating directions and feedback that prioritize humanity, caring, motivation, and the pursuit of additional knowledge and skills. In the context of AI, an example of “kind” may be a system that shares results directly with the student and provides feedback on how to improve, redirecting when a student makes a mistake so that they learn and improve. As noted within the *Standards for Educational and Psychological Testing* (2014), Standard 3 (3.1) includes, “Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population.” Kind signals a system that adapts to the needs of the user and promotes engagement in a way that fosters participation through increased accessibility.

Contrasting Principles from Other Organizations

Alongside the work of Smarter Balanced, other organizations engaged in processes to create and adopt frameworks or principles to guide work with AI. Smarter Balanced reviewed recent releases from organizations to compare and contrast the Student-Centric Design Principles for Responsible Use of AI with these other frameworks. Frameworks from IBM, ETS, Duolingo, TeachAI, and the Office of Educational Technology (OET) of the U.S. Department of Education were used for this comparison. Table 1 below highlights areas of commonality and distinction across these frameworks.

Smarter Balanced Student-Centric Design Principles for Responsible Use of AI	Similarities and Distinctions					
	IBM	ETS	Duolingo	TeachAI	OET USED-Teaching and Learning	OET USED-Developers
Fair	Fairness	Fairness and Bias	Fairness		Center People	Advancing Equity and Protecting Civil Rights
Secure and Robust	Robustness	Privacy and Security	Privacy and Security			Ensuring Safety and Security
Helpful (Impact)		Educational impact and integrity			Advance Equity	Providing Evidence for Rationale and Impact
Explainable	Explainability	Transparency, explainability, and accountability		Knowledge		
Transparent	Transparency	Transparency, Explainability, and Accountability	Accountability and Transparency		Promote Transparency	Promoting Transparency and Earning Trust
Agency	Privacy			Integrity/ Compliance/ Agency	Center People/Ensure Safety, Ethics, and Effectiveness	
Growth Mindset						Designing for Teaching and Learning
Kind						

Table 1. Synthesis of AI Principles

Smarter Balanced's approach was unique in two principles — "Kind" and "Growth Mindset" were not explicitly represented by the other organizations' frameworks.

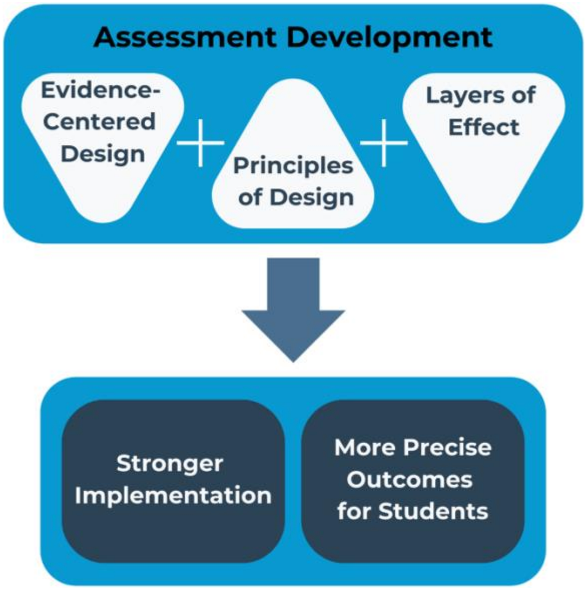
Smarter Balanced principles represented several commonalities across organizational guidelines. These included fairness, transparency, explainability, privacy, and security. Some nuances existed within the frameworks, including the notion of centering on people from the U.S. Department of Education (2023). Another framework from the U.S. Department of Education also noted trust as a key consideration for users engaged in AI (2024). The other foundations or principles reviewed ranged from four to seven total principles.

A small number of additional features were noted by other organizations, including: validity and reliability, continuous improvement, balance, and evaluation. Smarter Balanced used these features to revise the Student-Centric Design Principles incorporated into this document.

Design Approach in Leveraging Principles

Smarter Balanced views the principles as an aspect of Evidence-Centered Design (ECD) methodology. These principles may complement ECD to provide additional facets of accessibility while ensuring alignment of content and rigor through ECD. A graphic highlighting this complement is noted below.

Figure 2. Evidence-Centered Design and Principles Interaction



Application to State Policies

Many states also released AI guidance in 2024. Across states, several utilized the frameworks noted above as foundational components as they considered how to apply AI to teaching and learning. Most specifically, the U.S. Department of Education served as a springboard in several documents to highlight the human element of AI to ensure integrity of the solutions brought forward to engage with educators and students.

Common themes that were addressed were partiality, ethical and responsible use, privacy, and transparency as key considerations for various applications by interest holders. The states' guidance also addressed implementation so that educators may have access to guidance about how they might consider the integration of AI into learning environments.

Consistent with the principle of transparency identified above, many states noted the need to disclose and cite when and how AI is integrated into learning environments, and the thoughtfulness needed for appropriate citation and awareness. This aspect reflects the value of transparency related to the system and deployment. As part of their guidance regarding AI, several states also addressed the larger issues of using technology effectively and responsibly in the context of digital citizenship, computer science, and research practices.

Final Review and Recommendations

The primary motivation for the development of the Student-Centric Design Principles was to develop guidance that could foster innovation in the educational measurement community and support educators in being critical consumers of AI-related services. We welcome the opportunity to continue the conversation related to the Student-Centric Design Principles for Responsible Use of AI and how these may contribute to assessment development efforts.

Appendix A. SmarterAI Think Tank Use Case Methodology

Problem Statement

Historically, some educational assessments for K-12 students, including standardized tests and structured quizzes, have been criticized for various reasons. Research highlights several ways in which these assessments may be unfair:

Socioeconomic Considerations

Students from wealthier backgrounds often have access to more resources, such as tutoring, test preparation services, and supportive learning environments, which can lead to higher scores compared to their peers from lower socioeconomic backgrounds.

Cultural Considerations

Standardized tests can favor certain cultural and socioeconomic groups over others. Questions may be written in a way that assumes specific cultural or linguistic knowledge or experiences that are more familiar to certain groups, disadvantaging students from different backgrounds.

Accessibility Issues

Traditional assessments assume all students can be assessed in the same manner without sufficiently addressing their individual needs, including students with disabilities and students who are English language learners.

Impact on Teaching and Learning

Standardized testing can lead to a narrowing of the curriculum, where teachers focus primarily on rote memorization and test preparation rather than a broader educational experience. This can limit students' educational opportunities and engagement, particularly in subjects not included in the testing.

Organizational Structure

Smarter Balanced recruited and engaged multi-disciplinary panels utilizing three different structures:

The Quorum

The Quorum served as a small subset of thought leaders carefully chosen to give overall direction for this initiative. Their efforts included choosing the initial use case to focus on and helping determine the key Student-Centric Design Principles for the Responsible Use of AI and definitions. All Collective and Forum activities rolled up to the Quorum for review and further discussion.

The Collective

The Collective is the next largest group also made up of trusted professionals who engaged in the process to guide forum activities and speak to the effort *publicly.*

The Forum

The Forum is the largest group in which we engaged, and it consisted of teachers, students, IBM Trustworthy AI Center of Excellence practitioners, and others who expressed interest in participating in these thought-provoking activities.

The intent of the three working groups was to serve as the overarching SmarterAI Think Tank. The Quorum, Collective and Forum integrated together and discretely over several months to inform not only the use case, but also the principles as the use case was evaluated. The collaboration across the groups allowed for varied perspectives to be considered throughout the process.

Student-Centric Design Principles for Responsible Use Of AI

The Student-Centric Design Principles for Responsible Use of AI that our Quorum determined were necessary to have reflected in AI use cases that would be used in educational assessments include:

Fair

Secure and Robust

Helpful

Explainable

Transparent

Agency

Growth Mindset

Kind

Collaboration

Collaborating with a variety of stakeholders, both building AI models and governing these systems, is crucial for creating fair and effective assessments in grade school settings. Teams representing a variety of perspectives bring a wide array of experiences and cultural insights, which are essential to developing AI models that are impartial and representative of all students. This completeness helps to ensure that the AI systems do not inadvertently perpetuate existing variabilities or overlook the unique needs of different demographic groups. Governance frameworks that reflect this variety are better equipped to address ethical concerns, protect student privacy, and maintain transparency. By fostering an environment that is open and available to all, we can build AI models that truly enhance educational outcomes, providing every child with the opportunity to succeed and thrive in a way that respects their individuality and background.

Smarter Balanced was deliberative about those invited to contribute to this effort. Participants included subject matter experts with deep backgrounds in AI quality, EdTech, legal and compliance, universal design, community leaders from neurodivergent communities, and leaders representing various backgrounds. The forum included teachers, students, and staff from nonprofits. This engagement elicited these perspectives to support AI models that reflect the needs of various communities. Smarter Balanced recognizes that representation on teams that are designing these models and the systems of governance around these models offers outcomes aligned to accessibility and mitigating partiality.

There is a substantial body of information that documents the partiality that can be embedded in well-designed models based on insufficient or inadequate data used to train AI systems. Therefore, ensuring that the representation users is well characterized within the training data used in the model is another essential requirement to increase to address variety and completeness.

Design Thinking, Mural and Collaborative Effort

Design Thinking frameworks were offered by IBM Consulting to craft a human-centric approach to this investment in AI. The goal of this methodology is to support strategically aligned outcomes, responsive to both functional and non-functional requirements detailed by those establishing governance for the organization. Design Thinking enables developers and interest holders to deeply understand user needs, ideate innovative solutions, and prototype iteratively. This methodology is invaluable in identifying and assessing risks early in the development process, facilitating the creation of AI models that are ethical, open and available to all, and effective. By continuously engaging with varied communities of domain experts and other interest holders and incorporating their feedback, Design Thinking offers AI solutions that are not only technologically sound but also socially responsible. This approach aligns with IBM's commitment to designing AI systems that enhance human capabilities, foster trust, and promote fair outcomes. The Design Thinking frameworks used were consistent with IBM's own AI design guild.

The conversations with multiple interest holder groups involved several intentional steps that were designed to be iterative, with adjustments over time as more considerations became evident. First, groups addressed the question, "What is the relationship that we ultimately want to have with AI?" Following this conversation and refinement of the Student-Centric Design Principles for Responsible Use of AI, participants considered what they would want to see reflected in these AI models, along with their definitions. Once these aspects were defined, conversations focused on the functional and non-functional requirements that we would expect to see to operationalize those principles. Finally, the focus turned to the use case in which unintended effects of the model, mitigating risks, and preserving rights were contemplated. The final exercises utilized a voting method to refine the minimum requirements for the use case.

For our initial discussions, Mural was leveraged, allowing for a highly interactive tool with dozens of people collaborating at a time. We subsequently met across Quorum, Collective and Forum teams to crystallize our approach, story, and context around the requirements over several months.

Use Case Detail

Stakeholders engaged in the use case exploration refined the characteristics for consideration based on the following features noted in Table 2.

Table 2. Use Case Features and Considerations

Feature	Geography	Setting	Content Area	Context	Interaction	Grade Level
Use Case Consideration	United States	Public Schools	English/Language Arts	Formal Evaluation	Audio and Text (No Visual Representation	Sixth Grade

For the first use case, the SmarterAI Think Tank group members considered an AI-supported assessment of ELA/literacy competencies for sixth graders in the United States within a formal assessment context. To simplify and focus our use case, we chose to discuss a model that used audio, verbal, and text-based inputs but did not include an avatar, and with which a student could actively engage, asking and responding to questions. Considerations included that a conversational AI could be used to evaluate students' understanding of media (listening, multimedia understanding + reading), communication (speaking, writing) which would include research, critical analysis, and creativity as well as other skills — turn-taking, listening and reflecting, other healthy behaviors for conversation, and written communication (social-emotional competencies built in).

Layers of Effect Exercise

One sample exercise considered the primary intent of the AI model, then participants noted the secondary intended and known effects of that use case. The concluding aspect ascertained what are unintended and possibly known negative effects of that AI model use case.

Participants also worked to categorize the unintended harms. The categories that were refined through discussion included: potential harmful partiality, mental health, socioeconomic considerations, multilingual learners, cybersecurity, and communication challenges. For each of these categories of potential areas of harm, conversations collectively described mechanisms to mitigate these risks. This exercise better reflected the functional and non-functional requirements that we described in the prior exercise so that we could be better prepared for voting on the minimum requirements that we feel would be necessary for consideration.

Voting

For the voting exercise, participants reflected on the Student-Centric Design Principles for Responsible Use of AI, the unintended harms, and the ways that we described mitigating the risks before participants voted. Votes reflected those areas that should be considered and emphasized in subsequent efforts refining and implementing AI systems aligned to the use case.

Conclusion | Call to Action

The process in which we engaged provides a model for contemplating and evaluating levels of effect and risks. These efforts are published with the intent to share and highlight the types of conversations necessary before deploying AI models in educational assessments. Smarter Balanced and IBM Consulting recognize significant experimentation happening right now with AI without consideration for the unintended effects of those AI models. This is our opportunity to create responsible thinking and standards, through Student-Centric Design Principles for Responsible Use of AI, to support the educational measurement industry's move forward with the consideration, development, and adoption of AI in assessment.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing* (2014 ed.). American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf.

Arizona State University. (2024). *Generative Artificial Intelligence in K-12 Education*. Squarespace. https://static1.squarespace.com/static/64398599b0c21f1705fb8fb3/t/66b66995e1b1fc7f24f9d780/1723230615694/AZ+NAU.GAIGuide.pdf.

Colorado Department of Education. (2024). *Colorado roadmap for AI in K-12 education*. Squarespace. https://static1.squarespace.com/static/64398599b0c21f1705fb8fb3/t/66b669c8da2f3c4be35cf7a3/1723230664968/Colorado-Roadmap-for-AI-in-K-12-Education_August-2024.pdf.

Connecticut Department of Education. (2024). *Responsible AI policy framework*. Squarespace. https://static1.squarespace.com/static/64398599b0c21f1705fb8fb3/t/66b669d9616a3725434a699a/1723230681965/ct-responsible-ai-policy-framework-final-02012024.pdf.

Delaware Department of Education. (2024). *Generative AI guidance*. Squarespace. https://static1.squarespace.com/static/64398599b0c21f1705fb8fb3/t/66b669f7b7aafc138976e83e/1723230713888/delaware_generative_ai_guidance.pdf.

Duolingo. (2024). *DET+ Responsible AI Standards*. Duolingo. https://duolingo-papers.s3.amazonaws.com/other/DET%2BResponsible%2BAI%2BStandards%2B-%2B040824.pdf.

ETS. (2024). *ETS convening executive summary for the AI guidelines*. ETS. https://www.ets.org/Rebrand/pdf/ETS_Convening_executive_summary_for_the_AI_Guidelines.pdf.

IBM. (2024). *IBM artificial intelligence pillars*. IBM. https://www.ibm.com/policy/ibm-artificial-intelligence-pillars/.

TeachAI. (2024). *AI guidance for schools toolkit: Principles*. TeachAI. https://www.teachai.org/toolkit-principles

U.S. Department of Education, Office of Educational Technology. (2024). *Designing for education with artificial intelligence: An essential guide for developers*. U.S. Department of Education.

U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. U.S. Department of Education.

Washington State Department of Education. (2024). *Comprehensive AI guidance*. Squarespace. https://static1.squarespace.com/static/64398599b0c21f1705fb8fb3/t/66b66c3665fce31854dae63c/1723231288038/WA+comprehensive-ai-guidance.pdf