

Call for Proposals
Equity in Assessment and Measurement Webinar 2022

Goals for Webinar

1. To put a spotlight on the ways that the NCME community is taking up (or could be taking up) issues of equity and social justice in the context of assessment and measurement.
2. To engage members of the NCME community in discussion and debate.
3. To provide a forum for learning.

Proposals will be evaluated by the NCME Board of Governors. The winning proposal will be given a fund of \$3,000 to be allocated to session participants as honoraria, with a cap of \$2,000 as an honorarium for a keynote presentation. The webinar will be scheduled during the 3rd week of January 2022.

Deadline for proposal: October 1, 2021

Submit to NCME@talley.com

Criteria for Evaluating Proposals

A key criterion for any proposed session is that the featured presentation or discussion must be cast as a response to one or more focal questions related to research and/or practice for which there are *no agreed upon answers in the NCME community*. It is also encouraged (though not required) to reach outside of NCME for presenters/discussants that may be especially knowledgeable about a given topic.

The proposed formats for the session should be designed to best support the webinar goals (above).

Examples:

- Keynote presentation/demonstration followed by one or more discussants
- Facilitated panel discussion/debate

Session proposals should be no longer than 1500 words and address the following questions

1. What is/are the equity question(s) that are focus of the session?
2. Why is the proposed session significant? Why is it timely?
3. What is the interpretation-use-consequence context for the assessment(s) that will be the focus of the session (see Background section below)?
4. Who will be participating in the session and what are their qualifications?
5. What are at least two different positions that could be taken in answering these questions and who will be representing them?
6. To what extent will empirical data be brought to bear on the questions that are the focus of the session?
7. Who will be facilitating the session and ensuring that diverse viewpoints are being elicited?

Background

Equity represents the idea that the distribution of educational resources and services should be purposefully unequal so that the neediest of students are provided more support to compensate for their different starting points. With this definition in mind, it is not always clear what we mean when we speak of equity as a focus in the context of the activities undertaken for educational assessment and measurement. Historically, one purported equity value of large-scale assessments (e.g., standardized tests) is often attributed to their role as “flashlights” to reveal individual and group differences in performance and to shed light on *inequities* with respect to opportunities to learn. In fact, this is the value proposition underlying the two most recent instantiations of the Elementary and Secondary Education Act (ESEA), NCLB and ESSA. On the other hand, many have raised concerns about the unintended negative consequences associated with such uses. Similar concerns are often applicable when tests are used for selection, admission, and certification.

As large-scale assessments become more dynamic, and as classroom assessments become more thoughtful, there is some sense that it may be possible to incorporate principles of equity through purposeful design (e.g., Universal Design for Learning). In this sense what is meant by equity may be similar to the way *fairness* is currently used in the Standards, and perhaps even with newly emerging and

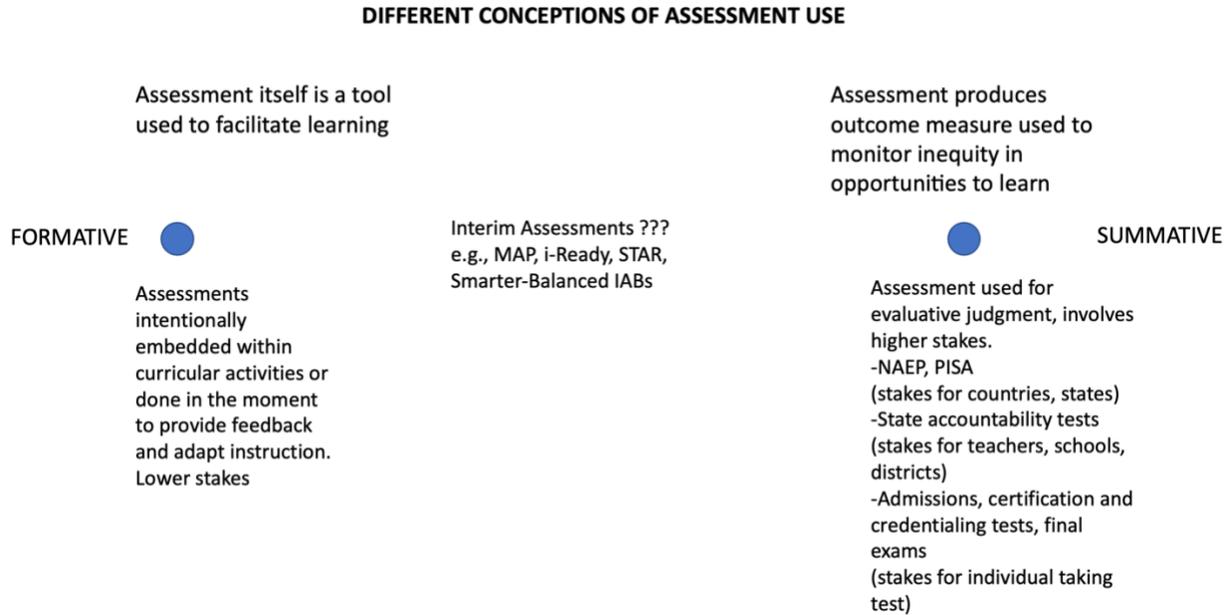
evolving conceptions of *anti-racism*. But the lines between these conceptions can be blurry, and commitments to equity, fairness and social justice may mean different things to different people.

The extent to which equity can be said to represent a focus of an educational assessment may depend upon interactions between construct conceptualization, assessment design, administration, analysis, interpretation, uses and consequences. One framework for conceptualizing the nature of these interactions (not intended to be definitive, but illustrative) is included below.

(1) From Construct to Interpretation: Distinguishable Stages of Assessment Life Cycle

Stage	Conceptualization of Construct(s)	Assessment Design	Assessment Administration	Analysis of Assessment Responses	Communication & Interpretation of Results
Some Conventional activities	<ul style="list-style-type: none"> *Creating content standards *Developing a model of student cognition *Establishing intended interpretative claims *Anticipating threats to validity *Role of culture and context 	<ul style="list-style-type: none"> *Choice of prompts to elicit evidence of relevant to construct *Item\task formats & design templates 	<ul style="list-style-type: none"> *Fixed, paper & pencil test *Dynamic (adaptive) digital environment *Embedded within a curriculum 	<ul style="list-style-type: none"> *Scoring item responses (informal vs. formal, human vs automated) *CTT *IRT *Linking & Equating 	<ul style="list-style-type: none"> *Choice of scale *Item Mapping *Normative information (percentiles) *Criterion-referenced information (performance levels) *Group Comparisons
Questions within a stage that bear on equity	<ul style="list-style-type: none"> *Who has been invited to the table? *To what extent are sources of cultural differences in representation and expression in the target population incorporated into meaning of the construct? 	<ul style="list-style-type: none"> *What are universal design elements? *What process is used to review items for bias? 	<ul style="list-style-type: none"> *What accommodations are available for students with construct-irrelevant barriers to participation? *What scaffolds can be introduced to maximize student motivation and engagement? 	<ul style="list-style-type: none"> *Is there empirical evidence of possible item and/or test bias (DIF, DTF)? *Are students responding to prompts/tasks as expected? *In what sense are different assessment instances comparable? 	<ul style="list-style-type: none"> *How are results of group comparisons being framed and/or contextualized relative to OTL? *What attention is being given to alternative interpretations?

(2) Different Uses of Assessment



Note: An oft forgotten assessment use that sits outside this continuum: research. (e.g., when an assessment serves as the outcome variable in evaluating the efficacy of some program or intervention, or to test theories about cognitive development and individual differences)

(3) Evaluation of Intended and Unintended Consequences of Assessment Use

The evaluation of consequences always requires evidence that derives from counterfactual reasoning, as any perceived consequence (i.e., effect) of assessment use can only be properly evaluated when compared to outcomes that would have been observed in the absence of the assessment.

Example:

- An intended consequence of using the SAT/ACT to inform college admission is to be more likely to identify students who are prepared to be successful in college.
- An unintended consequence is that use of the SAT/ACT may discourage students who would have been successful from applying to the more selective institutions.

In both cases, these consequences are difficult to evaluate in the absence of institutions of higher education that perform college admissions and do not use a standardized assessments, or use some alternative in their place. In place of this we have predictive studies relating SAT/ACT scores to college outcomes and mostly qualitative evidence about the effect of standardized testing on self-efficacy and motivation. Or we have quasi-experimental comparisons between schools that require admissions tests and those for whom it is optional. How well do these allow for a valid evaluation of consequences? What are some alternatives?

In summary, any proposed webinar session should be careful to situate the context for the focal question (or questions). For example:

- (1) Where does the question reside within the five stages of an assessment life cycle that support interpretations?
- (2) For what assessment use is it most relevant? (Formative? Summative? Both?)
- (3) Does the session take up the evaluation of consequences? (Anticipated and unanticipated? What is the counterfactual to a given use of an assessment?)