

Diversity, Equity, and Inclusion Validity Arguments

And the *Matrix of Evidence for Validity Argumentation* for the design or evaluation of research instruments

by **Dakota W. Cintron, PhD, EdM, MS** and **Erin Hagan, PhD, MBA**

Introduction

Assessment¹ (or the act of assessing something) and **evaluation** (or the act of determining merit, worth, or significance) are vital to society's understanding of science, the development of knowledge, and innovation. Diversity, equity, and inclusion (**DEI**) has not always been prioritized in assessment and evaluation practices, which form the basis of scientific evidence; especially in the context of quantitative data analysis. As an institution, the Academy has not historically required or incentivized DEI, and the socioeconomic and cultural backgrounds of researchers influence assessment and evaluation choices. As a result, decisions about assessment and evaluation risk overlooking or disregarding the perspectives of individuals from historically marginalized communities with limited social or political power.

At the inaugural RWJF RELEvent conference (a convening of the foundation's Research Evaluation and Learning grantees), the use, value, and validity of quantitative data, for both research and advancing health equity, were juxtaposed through a **DEI lens**. On one hand, quantitative data was seen as a potential roadblock to research advancing

health equity (e.g., quantitative data may not be valid or reliable for certain individuals and communities and doesn't provide nuanced insights explaining *why* an outcome is occurring). On the other hand, quantitative data was seen as essential to research and advancing health equity (e.g., by elucidating health inequities themselves).

In light of these challenges, the E4A Methods Lab chose to reflect on the ways in which the **instrument design process** could be more fully inclusive of diverse audiences, especially as it relates to validly measuring or collecting quantitative data on outcomes that help support a culture of health (e.g., well-being) and health equity. Research instruments (or simply **instruments**) include, e.g., survey questionnaires, cognitive or psychosocial assessments, interviews, tests, or checklists.

What is validity? What is validity through a DEI lens?

Before diving into what it means to use a DEI lens throughout the instrument design process, we provide an overview of our definition of **validity** and **DEI validity arguments**. Traditionally, at the core of all validity studies, is the simple notion of whether we are measuring what we claim to measure. The 2014 Standards for Educational and Psychological Testing provides a recent definition of validity that is common parlance in the fields of mea-

¹ We use boldface to denote terms we define in a glossary.

surement and assessment [1]:

"Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests. The process of validation involves accumulating relevant evidence to provide a sound scientific basis for the proposed score interpretations. It is the interpretations of test scores for proposed uses that are evaluated, not the test itself." (p. 11).

As the quote demonstrates, validity is an important concept for justifying the interpretations and use of a test (a specific research instrument).² Further, the definition emphasizes that validity is not a property of the instrument, but rather a property regarding the interpretations or use of scores derived from an instrument. For instance, the validity of an instrument has been framed as a form of an argument.

“The description of validity as ‘argument’ emphasizes the need for various kinds of evidence arranged so that the ‘argument’ as a whole is coherent and convincing. It draws attention to the importance of plausible rival hypotheses. And, it indicates the openness of the enterprise; real arguments about important issues are hardly ever resolved by a simple ‘yes’ or ‘no’ answer. Arguments are plausible or credible, rather than certain” (Kane, 1992, p. 20).

Consistent with Kane’s definition of validity argumentation, we define DEI validity arguments as the arguments or evidence necessary to ensure validity through a DEI lens. For an instrument to be useful among diverse

² In this definition of validity, the focus is on tests (a specific research instrument). In this method note, we consider this definition of validity as appropriate for all research instruments.

individuals, that instrument needs to have evidence arranged in a way that coherently and convincingly demonstrates DEI throughout the instrument design process.

How do we construct DEI validity arguments for the instrument design process?

We focus on the recently developed Matrix of Evidence for Validity Argumentation (MEVA; Solano-Flores, 2019) because it offers a tool or schema (i.e., the MEVA itself) for constructing, organizing, and arranging validity arguments throughout the instrument design process. The MEVA was designed to facilitate the construction of validity arguments for cultural validity in large-scale tests. We modify the MEVA from its original presentation to narrowly focus on the instrument design process. A simplified MEVA for the instrument design process is in **Table 1**. Along the columns of the matrix are the components of the instrument design process (see **Table 2** for

definitions) and along the rows of the matrix are a sample of possible procedural assumptions³ that are critical to valid and fair measurement of diverse populations (see

³ Solano-Flores (2019) defines procedural assumptions as “criteria of conceptual and methodological rigor that need to be met throughout the entire assessment process in order to validly, fairly test culturally diverse populations. Altogether, these procedural assumptions formalize the kinds of actions that need to be taken to ensure the defensibility of the interpretation and use of test scores across cultural groups.” For instance, a construct validity study using multiple-group confirmatory factor analysis would be an example of a probabilistic reasoning procedural assumption common in the instrument evaluation stage of the instrument design process.

Table 3 for definitions). Solano-Flores (2019) notes that the intersection of the rows and columns in the MEVA are “key to operationalizing (and ultimately elucidating what it takes to attain) cultural responsiveness in large-scale assessment.”

For example, consider the cells along the first row of the example MEVA in **Table 1** (i.e., cells [1,1], [1,2], and [1,3]; note we hold the row constant at one but vary the columns from one to three). Row one in the MEVA focuses on the procedural assumption of inclusion, representation, and sampling.

- In cell [1,1], we are concerned with inclusion, representation, and sampling in the theory generation and scoping stage of the instrument design process. As a DEI validity argument, cell [1,1] is concerned with the question of whether or not diverse and representative groups are included in defining the constructs or dimensions an instrument intends to measure. For example, were diverse individuals included in qualitative interviews to hone the definition of the construct being measured by the instrument? Were diverse individuals consulted to assess whether the dimensions correspond with their understanding of the construct or what it means in their community?
- In cell [1,2], we are concerned with inclusion, representation, and sampling in the content generation stage of the instrument design process. As a DEI validity argument, cell [1,2] is concerned with the question of whether or not diverse individuals were included, for example, in the pilot testing of an instrument. That is, when initially piloting an instrument, was sampling done in a way that was inclusive and representative of diverse perspectives?
- In cell [1,3], we are concerned with inclusion, representation, and sampling in the instrument evaluation phase. Similarly, cell [1,3] is concerned with the question of whether or not diverse and representative groups were included, for example, in a second pilot study that is often used for construct validation purposes (i.e., a probabilistic reasoning approach for instrument evaluation or cell [2,3]). We leave the remaining interpretations to the reader (please see Solano-Flores, 2019 for further details).

Table 1: Simplified Matrix of Evidence for Validity Argumentation for instrument design or development

Examples of Procedural Assumptions	Simplified steps in the instrument design process		
	Theory Generation and Scoping	Content Generation	Instrument Evaluation
Inclusion, representation, and sampling	[1,1]	[1,2]	[1,3]
Probabilistic reasoning	[2,1]	[2,2]	[2,3]
Implementation	[3,1]	[3,2]	[3,3]

Note. We adapt the MEVA from Solano-Flores (2019) to the instrument design process (see, for example, McCoach et al. [2013] for a more detailed description of instrument design or development). We only include a subset of the procedural assumptions from Solano-Flores (2019). Solano-Flores (2019) notes that each cell of the MEVA may contain multiple pieces of evidence and may contain conflicting pieces of evidence (i.e., confirming or disconfirming evidence). Furthermore, there may be blank cells because relevant information is unavailable. However, absence of evidence is not evidence of absence. Blank cells in the MEVA may be an indication of poor implementation or poor documentation (Solano-Flores, 2019).

Table 2: Simplified stages in the instrument design process**1. Theory generation and scoping**

- Specify the purpose of an instrument. Preliminary definitions and dimensions of a construct are specified.

2. Content generation

- Hone in on the items and content that will be used to measure a construct of interest. At this stage, operational definitions are created, a scaling technique is selected, and a draft of the final version of an instrument is generated. Pre-piloting of the instrument takes place within small groups from a target population.

3. Instrument Evaluation

- Conduct a second pilot study. Factor analysis is done, reliability analyses are conducted, and item and scale properties are evaluated. A final version of the instrument items and a manual for the instrument use are generated. Steps for instrument administration and suggestions for decision-making are formulated.

Table 3: Examples of procedural assumptions**1. Inclusion, representation, and sampling**

- Sampling includes and is representative of individuals from diverse groups and their social contexts.

2. Probabilistic reasoning

- Error due to uncertainty (e.g., from the characteristics of socioeconomic or cultural groups) is recognized and estimated using quantitative data or methods.

3. Implementation

- Resources and efforts are allocated to ensure that methods and procedures are applied with fidelity and consistently across individuals and socio-cultural contexts.

Overall, the MEVA approach offers a systematic strategy for applying DEI principles at each step in designing or evaluating an instrument. To comprehensively evaluate validity from a DEI perspective, each cell of Table 1 should be considered through a DEI lens, and confirming or disconfirming evidence of whether certain principles are followed must be provided (e.g., if a study did/did not consult diverse individuals in the theory generation and scoping stage, this would be confirming/disconfirming evidence of DEI for an instrument). Importantly, Solano-Flores (2019) notes how the evidence in a cell of the MEVA enables the construction of narratives. These narratives provide an illustrative record of the disconfirming/confirming validity evidence in a cell and form the basis of a coherent DEI validity argument. The MEVA should not be treated simply as a checklist to demonstrate sensitivity to DEI. The narratives are important because they describe what was done throughout the instrument design process and how the evidence in a cell is relevant to DEI.¹ The process of ensuring that an instrument adequately addresses or captures issues of DEI is not an easy task. As this simplified example of the MEVA demonstrates, many components can influence the validity of an instrument. Safeguarding against these components requires rigorously constructing validity arguments that help ensure that the design of an instrument considers issues of DEI.

¹ In a future blog post, we provide an example of constructing these narratives when evaluating the evidence of an existing instrument.

Is there a threshold for deciding whether there is sufficient DEI validity evidence to use an instrument?

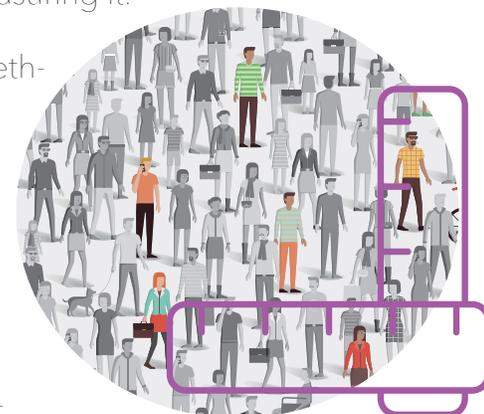
As the definition of validity given by the Standards indicates, “The process of validation involves accumulating relevant evidence to provide a sound scientific basis for the proposed score interpretations” (p. 11). In the process of evaluating a set of instruments (e.g., to measure well-being), the MEVA is consulted and DEI validity evidence is arranged. The decision to use a particular instrument (or to seek out an alternate instrument or design one’s own instrument in the belief that DEI validity evidence is insufficient) is a decision that needs to take additional considerations into account. The MEVA provides us with a tool to reason about the extent of validity evidence that supports claims that an instrument is valid for use among relevant diverse individuals or communities. However, other considerations (e.g., cost-benefits or negative consequences of use for particular communities or individuals) should factor into the decision to use an existing instrument or design a new one. In all cases, the use of an instrument should be prefaced by an explicit statement outlining what interpretations of the instrument scores can be made based on the DEI validity evidence for an instrument.

Why does applying a DEI lens to the instrument design process matter?

A DEI lens helps build **validity evidence** that aligns with values that promote social justice, health equity, and community empowerment. Given the critical role that instruments play in providing quantitative data to evaluate or test theories in research, assessment, and evaluation, we need to be sure that instruments are valid and reliable, especially

among marginalized, minoritized, or otherwise vulnerable groups. Otherwise, it may not be clear what the instruments are measuring or for whom they are measuring it.

In this [blog](#) and methods note, we aim to provide awareness about DEI in the instrument design process. We hope this blog highlights best practices and how we might construct



validity evidence to ensure that the instruments guiding research and knowledge generation are equitable and inclusive of all people. The MEVA provides a schema for developing DEI validity arguments to ensure that what we measure with an instrument is ultimately reflective of diverse individuals’ input and perspectives. In practice, the procedural assumptions that are included in the MEVA (the rows) can be more comprehensive than what we detail here. Furthermore, we might use the MEVA to consider the entire research and evaluation process (e.g., we might reformulate the MEVA to assess to what degree an evaluation used a DEI lens).²

Instruments developed using a DEI lens can offer insight into how accurate the instrument may be in rendering an inclusive and representative picture of society as a whole. There is also evidence that involvement in the instrument design process may help promote cultural validity and increase the participation and engagement of diverse individuals who may be underrepresented in research studies (see, for example, Sheldon et al., 2007). Simply, instruments designed to promote DEI can help establish trusts, norms, and practices that may help improve

² For example, one might formulate an MEVA where the steps in the CDC framework for program evaluation are along the columns of the matrix and the equitable evaluation framework principles are the procedural assumptions along the rows of the MEVA.

participation research and response behavior (e.g., the decision to omit a response or not).

References

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain*. New York, NY: Springer.

Scriven, M. (2007). The Logic of Evaluation. OSSA Conference Archive. Retrieved from <https://scholar.uwindsor.ca/ossaarchive/OSSA7/papersand-commentaries/138>

Sheldon, H., Graham, C., Potheary, N., & Rasul, F. (2007). *Increasing response rates amongst black and minority ethnic and seldom heard groups*. Europe: Picker Institute.

Solano-Flores, G. (2019, June). Examining Cultural Responsiveness in Large-Scale Assessment: The Matrix of Evidence for Validity Argumentation. In *Frontiers in Education* (Vol. 4, p. 43). Frontiers.

Glossary

Assessment. The action or an instance of making a judgement about something: the act of assessing something. (Merriam-Webster).

DEI lens. A diversity, equity, and inclusion (DEI) lens is the deliberate commitment to incorporate DEI principles and practices into the environments that we work and the knowledge that we produce.

DEI validity arguments. The arguments necessary to ensure that an instrument is valid through a DEI lens. That is, for an instrument to be valid and useful from a DEI perspective, that instrument needs to have evidence arranged in a way that coherently and convincingly demonstrates DEI throughout the instrument design process.

Evaluation. "Evaluation is the process determining merit, worth, or significance; an evaluation is a product of that process" (Scriven, 2007, p. 1). In this regard, the use of instruments for assessing or evaluating some outcome construct can be framed as an evaluation. For more information on evaluation, see the [CDC's Practical Strategies for Culturally Competent Evaluation](#).

Instruments. Instruments or research instruments are devices used to collect quantitative data in research. Instruments might include, for example, survey questionnaires, cognitive or psychosocial assessments, tests, or checklists

Instrument design. The process by which an instrument is designed.

Validity. See definition of validity from the 2014 [Standards for Educational and Psychological Testing](#) provided at the top of page 2 in the "What is Validity? What is Validity Through a DEI lens?" section.

Validity evidence. Various sources of evidence that are used to evaluate the validity of a proposed interpretation of instrument scores for a particular use. Classical forms of validity evidence include, for example, content, criterion, or construct validity.

The E4A Methods Lab was developed to address common methods questions or challenges in Culture of Health research. Our goals are to strengthen the research of E4A grantees and the larger community of population health researchers, to help prospective grantees recognize compelling research opportunities, and to stimulate cross-disciplinary conversation and appreciation across the community of population health researchers. We welcome suggestions for new topics for briefs or training areas.

Support for this note was provided by the Robert Wood Johnson Foundation. The views expressed here do not necessarily reflect the views of the Foundation.