January 2025

EXAMINING THE CONSEQUENCES OF ASSESSMENT:

An Integrated Framework for Developing Measurement Tools

This brief shares findings from Project ESSY, an IES-funded research grant (R305A220249) with aims to develop and evaluate the ESSY Whole Child Screener, a school-based screening instrument and associated data reporting structures for assessing a range of child and environmental indicators.

Consequential validity refers to the social consequences of test use. These can be positive or negative – and intended or unintended.

Tests, assessments, and screeners are used across the fields of education, psychology, and healthcare to diagnose, evaluate, and determine next steps. Use of these assessments can have both positive and negative consequences. An educational assessment, for example, might identify needed support that improves a student's success in school, generating positive consequences. Yet, the same assessment may lead to labeling or tracking practices that limit student opportunities, yielding negative consequences. These positive and negative effects of test use have been referred to as **consequential validity**, or the social consequences of measure use in the real world.[1]

Why is consequential validity important?

In addition to positive or negative consequences, a consequential validity lens further specifies that the consequences of test use can either be intended or unintended, resulting in four possibilities:

- **Positive intended consequences**: goal of measure use
- **Negative intended consequences**: should be avoided from the outset
- **Unintended positive consequences**: unexpected but favorable
- **Unintended negative consequences**: unexpected and unfavorable

The consequential validity of assessments has not been widely studied. There are multiple reasons for this. To evaluate consequential validity, the measure needs to be used for some time to allow for consequences of its use to unfold. Therefore, there is a time lag between when test developers do standard psychometric tests (e.g., reliability, validity) and when they could evaluate consequential validity. There is also disagreement about whether evaluating consequential validity is the responsibility of test developers or test users. Although recognized as an important aspect of validity, consequential validity has gone largely unstudied to date.

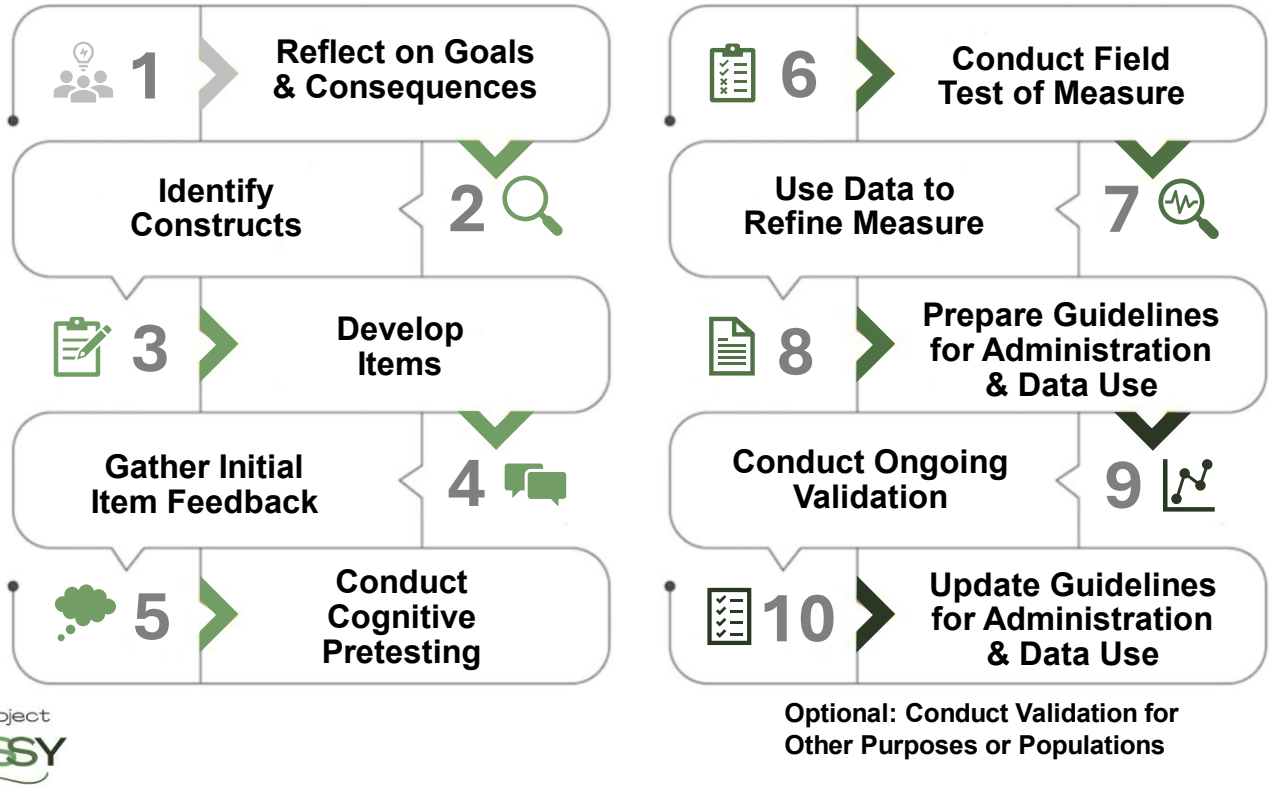
How could consequential validity be incorporated in measure development?

The Consequential-Validity Centered Measure Development Framework provides guidance for attending to potential and actual social consequences throughout the test development process.

In conjunction with developing the ESSY Whole Child Screener, the Project ESSY research team created an integrated measure development approach that centers consequential validity. Building from both traditional measure development[2,3] and recent mixed methods approaches,[4,5] this framework plans for consequential validity throughout the development and evaluation process. It involves multiple rounds of mixed methods (i.e., quantitative and qualitative data) input from key groups to identify constructs, develop items, and refine instrumentation. At each stage, key groups are also engaged in identifying potential positive and negative consequences of measure use. By doing so, test developers gain critical insight intended to promote positive consequences of their measure's use and mitigate any potential unintended consequences of its use, such as negatively affecting certain groups more than others (e.g., disproportionality or unfair outcomes for some groups).

The **Consequential-Validity Centered Measure Development Framework** includes ten steps, as shown in the figure.

Consequential Validity-Centered Measure Development Framework



Project ESSY

The 10 steps can be roughly organized into four phases of work:

Phase 1: Pre-work

The pre-work begins with test developers reflecting on goals and intended consequences of developing a new measure. (***Step 1***)

Phase 2: Initial measure development

Researchers developing measurement tools should carefully define the concepts or constructs they are trying to measure and begin developing measurement items by (1) reviewing the literature and existing measures and (2) gathering input from diverse advisory boards that include intended users of the measure. (***Steps 2-3***) After developing initial items for measurement tools, researchers should evaluate (1) content validity (appropriateness, relevance, and representativeness of the items) from experts, and (2) usability and potential consequences with intended users. (***Step 4***) After incorporating that initial feedback, researchers should again gather feedback from intended users regarding instructions and item clarity through cognitive interviews. This cognitive pretesting should be completed prior to any large field testing to determine the need for refinement and understandability. (***Step 5***)

Phase 3: Initial validation

After completing the previous steps, researchers should conduct a large field test, inclusive of quantitative and open-ended items. Open-ended items can facilitate collection of general feedback about items or the measure more broadly as well as potential consequences of measure use. (***Step 6***) After field testing, researchers should analyze qualitative and quantitative data separately, and then together, to refine the measurement tool. Researchers should also evaluate the potential or actual consequences of scores. (***Step 7***) After finalizing revisions, researchers should develop instructions for implementation (e.g., time period, any training requirements), and data/score interpretation with attention towards reducing bias and facilitating positive consequences. (***Step 8***)

Phase 4: Ongoing validation

Once the measurement tool has been published and is in use, researchers must continue to evaluate evidence of reliability and validity with new samples. This ongoing process should include looking at intended and unintended consequences in the real world. Again, an integrated mixed methods approach is recommended to more deeply explore how scores are being used and what consequences they produce, both short-term and long-term, in applied practice. (***Step 9***) After the measurement tool has been both in use and has been studied, the researchers and their wider network of practitioners should use those findings to revise implementation procedures as necessary in order to facilitate positive consequences (e.g. training manuals, score reports, how results feedback is given). (***Step 10***) Finally, an optional step is to conduct validation for other purposes and populations, using a mixed methods approach that involves both quantitative and qualitative data.

In summary, the Consequential-Validity Centered Measure Development Framework integrates traditional measure development and emerging mixed methods approaches into a framework that centers consequential validity in decisions. In doing so, it aims to maximize positive consequences of measure use by proactively planning for consequences from the outset of measure development and evaluating consequences at multiple points throughout the development and measure process.

To learn more, visit the Project ESSY website at https://expanding-school-screening.education.uconn.edu/.

To Cite this Brief:

> Caemmerer, J. M., Koslouski, J.B., Chafouleas, S.M., Briesch, A.M., Melo, B., & Marcy, H. M. (2025, January). *Examining the Consequences of Assessment: An Integrated Framework for Developing Measurement Tools*. Brief available from https://expanding-school-screening.education.uconn.edu/

Project ESSY represents a collaboration between researchers at the University of Connecticut and Northeastern University. The research project is supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A220249 to the University of Connecticut (PI: Chafouleas). The content in this brief does not represent the views of the Institute or the U.S. Department of Education.

---

[1] Messick, S. (1998). Consequences of test interpretation and use: The fusion of validity and values in psychological assessment. *ETS Research Report Series*, *1998*(2). https://doi.org/10.1002/j.2333-8504.1998.tb01797.x

[2] Bandalos, D.L. (2018). *Measurement theory and applications for the social sciences*. The Guilford Press.

[3] Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6,* 149. doi: 10.3389/fpubh.2018.00149

[4] Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research*, *4*(1), 56–78. https://doi.org/10.1177/1558689809355805

[5] Sankofa, N. L. (2022). Transformativist measurement development methodology: A mixed methods approach to scale construction. *Journal of Mixed Methods Research*, *16*(3), 307–327. https://doi.org/10.1177/15586898211033698