



California's approach to supporting high quality data.
Session presented at the
Measuring Child and Family Outcomes meeting, Bethesda, MD.

Dimensions of Quality Assurance

Fidelity: The degree to which procedures are implemented as planned with intended recipients

Dunst, C.J., Trivette, C.M., McInerney, M., Holland-Coviello, R., Masiello, T., Helsel, F., & Robyak, A. (2008). Measuring training and practice fidelity in capacity-building scaling-up initiatives. *CELLpapers*, 3, 1 – 11.

Available for download at http://www.earlyliteracylearning.org/cellpapers/cellpapers_v3_n1.pdf

This article describes a framework for measuring fidelity in scaling-up initiatives. The purpose of the article is to provide guidance on the development of fidelity indicators and the collection of fidelity data. Three types of fidelity are defined including Implementation Fidelity, Diffusion Fidelity, and Intervention Fidelity.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Retrieved from the National Implementation Research Network: http://www.fpg.unc.edu/~nirn/resources/publications/Monograph/pdf/Monograph_full.pdf.

From the author: "This monograph summarizes findings from the review of the research literature on implementation and proposes frameworks for understanding effective implementation processes. The results of this literature review and synthesis confirm that systematic implementation practices are essential to any national attempt to use the products of science - such as evidence-based programs - to improve the lives of its citizens." In addition to an overview of implementation there is a section that addresses evaluation and fidelity specifically.

Agreement: Extent to which two independent assessors assign the same scores under similar conditions

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis (2nd ed.)*. New York: Cambridge University Press.

This book describes procedures for collecting and analyzing observational data. Techniques reviewed include: systematic observation, developing a coding scheme, recording behavioral sequences, assessing observer agreement, representing observational data, and analyzing sequential data. A detailed description of the difference between agreement and reliability and step-by-step instructions for selecting and computing agreement statistics is provided.

Completeness: The extent to which the appropriate respondents are identified and correctly sampled and sufficient response rates are obtained

Seastrom, M.M. (2002). *NCES statistical standards*. (Report No. NCES 2003601). Retrieved from National Center for Educational Statistics: <http://nces.ed.gov/statprog/2002/stdtoc.asp>

This technical report describes the procedures implemented by the National Center for Educational Statistics to maintain high quality data for national statistics. Several important topics are included: collection of data, processing and editing of data and analysis of data. The section on collection of data includes procedures for designing and collecting appropriate samples, including acceptable response rates.

Reliability: extent to which assessment scores are consistent across different aspects of the measurement setting (for example, time, raters, items)

Thompson, B. (2007). *Score reliability: Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage Publications.

This is a short edited book focused on the topic of score reliability. The author describes in detail the meaning of reliability and some common misconceptions. Importantly, the author describes the need to talk about the reliability of scores vs. the reliability of tests. The point is made that the reliability of scores changes with changes in the measurement context and the reliability of scores will often differ from those reported in the testing manual.

Validity: the degree to which scores from an assessment measure the intended construct

AERA, APA, NCME (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.

These standards were developed jointly by the American Educational Research Association, American Psychological Association and the National Council on Measurement in Education. Relevant topics include: test construction, evaluation, and documentation and testing applications. The section on validity provides a general overview of current thinking about validity and specific details for several different types of validity. In addition, several standards guiding selection of appropriate assessment instruments are described.

Utility: the extent to which stakeholders find the information generated from the assessment to be useful (accessible, relevant, timely, complete and accurate)

Lee, Y.W., Strong, D.M. (2003). Knowing-why about data processes and data quality. *Journal of Management Information Systems*, 20, 13 – 39.

This article describes the relationship between how useful data are and the quality of data. The article defines data of high quality at “data that are fit for use by the data consumers.” They describe five dimensions of data quality including: accessibility, relevancy, timeliness, completeness, and accuracy. The effect of knowledge about the data process on each of these dimensions is explored.