

## **A Comparison of Methods for Measuring Fidelity of Implementation Outcomes**

**Pamela Drake, PhD;  
ETR Associates, Inc.**

**Seow Ling Ong, MPH; Jill Glassman, PhD; Lisa Unti, MPH; Regina Firpo-Triplett, MPH; Jennifer Larkin,  
MPH; Bayla Greenspoon; Kay Todd  
ETR Associates, Inc.**

Implementation fidelity is an important process variable to measure, and measure accurately, when evaluating evidence-based programs (EBPs). It becomes even more important to measure it accurately when implementation fidelity is the outcome of a study, such as when evaluating training programs for educators implementing EBPs. The simplest and most commonly used data sources for measuring implementation fidelity are educator self-report implementation logs. These typically ask educators to indicate which activities in a lesson they implemented and what type of changes they made. However, these self-report measures are fraught with error and often overestimate implementation fidelity. Other measures such as interviews and in-person observations are sometimes used, but are more burdensome and costly. Through a NIH-funded SBIR, we evaluated an online training program designed to improve the implementation fidelity of teachers implementing the Reducing the Risk program. Two-hundred nineteen educators across the United States were randomly assigned to either the intervention (training) or control (standard preparation using only the teaching guide) condition. Educators completed a pre-survey prior to training and implementation and a post-survey after completing implementation. They also completed implementation logs for all 16 lessons (162 completed all 16 logs). We also conducted 25 observations of educators implementing one of 4 key lessons. All educators, furthermore, were asked to audio-tape each of the 4 key lessons (88 turned in audiotapes). Our analyses compared fidelity scores and agreement across methods.

- Fidelity scores based on percent of activities covered based on teacher logs were very high
- Teachers and observers agreed on activity ratings only about half the time
- Teachers and coders agreed on activity ratings on average 61% of the time
- Fidelity scores based on observations and codings were consistently lower than the teacher self report
- Fidelity scores and agreement between teachers and observers/audio coders varied across types and focus of items
- Differences likely due to social desirability, recall, and interpretation of items
- Online tracking system, online logs, incentives increase completion of logs but still had a significant lag time
- In-person observation, are likely more accurate than self-report but cost-prohibitive and logistically difficult
- Audio observations may be an alternative, but only for behaviors that can be picked up via recorder
- Questions:
  - Does in-person observation really yield the most accurate fidelity score? Perhaps it's a combination of information that is most accurate.
  - How do we balance the need for quality results against available resources?
  - Can we provide any type of training on self-report logs that doesn't interfere with our outcomes?
  - Are errors random across TX and C groups?