



Introduction to White Paper 2

December 30, 2021 | QTurn Team

Welcome back! In this second white paper, [Measuring Socio-Emotional Skill, Impact, and Equity Outcomes](#), we extend from the White Paper 1 skill framework to discuss implications for accurate measurement.

We are pleased to share these hard-won lessons from two decades of trying to describe the actual outcomes of “broad developmentally-focused programs” – which means trying to figure out how to measure socio-emotional skill changes of both adults (i.e., quality practices) and children over short time periods. In the paper, we work through the logic of measurement in a way that we hope non-technical readers can follow with minimal suffering. There are no Greek symbols!

We’re passionate about this subject because the potential is real. Getting measurement right will make a big difference for the oft-ignored questions about how and how much skills change during relatively short periods, such as a semester or school year. To put the conclusion up front: Maximize measurement sensitivity in applied settings by using (a) adult ratings of child behavior that (b) reference periods of not more than two weeks past, and (c) using a scale anchored by frequency of behavior – what we call *optimal skill measures*.

Another message is that regardless of measure choice, items should analogize to actual mental and behavioral skill states that occur in real time, using words that all raters understand in the same way. Without this power of analogy *from* the raters’ concept of the verb/predicate in the written item *to* an observed quality in the room, external-raters can’t make clear comparisons before checking the box. The same is true for self-raters observing thoughts and feelings happening inside their own mind/body.

The kicker is that as inaccurate data are aggregated, the extent of invalidity is compounded. What if the ambivalent impact findings repeatedly demonstrated by gold-standard evaluations in publicly-funded afterschool programs were caused by leaving out accurate information about socio-emotional skills? (This is, in fact, a key argument elaborated in White Papers 3 and 4.) Thanks for checking out our work!

P.S. for the psychometrically minded. **Why are many SEL skill measurement constructs likely to be inaccurate, despite psychometric evidence of reliability and validity?** First, many measures of SEL skill lump things together that they shouldn’t. For example, mixing self-report items about (beliefs about emotional control in general (efficacy), the felt level of charged energy in the body (motivation), and specific behaviors (taking initiative) that follow – creates scale scores that obscure distinct parts of skill that change on different timelines and with different causes.

Second, it turns out that most measures young people encounter in school day and OST settings are self-reports of beliefs about skills. Students are rarely trained in the meaning of the words in the items that they are responding to while coming from different histories and different “untrained” perspectives on

emotion-related words. We just don't know what the words mean to the self-reporter, particularly the relative intensities offered in multi-point Likert-type response scales.

Third, items that refer to the use of skills *in general* (i.e., a verb without clear predicating context or time period) are much less sensitive to specific skill changes that actually occur over short periods of time. We refer to these as measures of *functional skill* levels that change more slowly over time.

In the new year, we're highlighting the third white paper, *Realist(ic) Evaluation Tools for OST Programs: The Quality-Outcomes Design and Methods (Q-ODM) Toolbox*. In this paper, Charles Smith and Steve Peck extend the ideas introduced in White Paper 1 (socio-emotional framework) and White Paper 2 (socio-emotional measures) to program evaluation and impact evidence.

www.qturngroup.com/2021/12/30/introduction-to-white-paper-2