QTURN WHITE PAPER #2

Measuring Socio-Emotional Skill, Impact, and Equity Outcomes

Updated 08/02/21

“Educational equity is when educational practices, policies, curricula, resources, and school cultures are representative of, constructed by, and responsive to all students, such that each student has access to, participates and makes progress in high-quality learning experiences, resulting in positive outcomes regardless of individual characteristics and cultural identities” (Michigan Department of Education, 2017, p. 9).

“[SEL] helps address the problem of inequity. Children from high-poverty, at-risk neighborhoods have less access to the kinds of enriching experiences that build social and emotional skills crucial for success in school and life. That’s an opportunity gap that we must close in the system, by ensuring that schools and afterschool programs are equipped to help children develop social and emotional skills as well as academic skills” (Loeb et al., 2016, p. 16).

“…the validity of performance assessment needs to be systematically addressed, as do other basic measurement issues such as reliability, comparability, and fairness. The latter reference to fairness broaches a broader set of equity issues in testing that includes fairness of test use, freedom from bias in scoring and interpretation, and the appropriateness of the test-based constructs or rules underlying decision making or resource allocation, that is, distributive justice” (Messick, 1995, p. 741-742).

Contents

Summary **3**

I. Introduction **4**

Realistic Approach 5

Note on Domains 6

II. Selecting SEL Measures **8**

Step 1. Integrated Model of Skill 8

Step 2. Objects and Measures 9

Schemas
Beliefs
Awareness
Behavior

Step 3. Indicator Purpose 11

Mental or Behavioral
Functional or Optional

Step 4. Profile Measures 13

III. Recommendations **15**

References **17**

Notes **22**

Citation: Smith, C., & Peck, S. C. (2020). *Measuring Socio-Emotional Skill, Impact, and Equity Outcomes* [White Paper #2]. QTurn.

This paper is part of a series: White Paper 1 – *Socio-Emotional Skills, Quality, and Equity* (Peck & Smith, 2020b) – provides a translational framework for understanding the key parts of an SEL skill set. White Paper 2 – *Measuring Socio-Emotional Skill, Impact, and Equity Outcomes* (Smith & Peck, 2020a) – provides guidance for selecting feasible and valid SEL skill measures. White Paper 3 – *Realist(ic) Evaluation Tools for OST Programs* – integrates the SEL framework and measures with a pattern-centered approach to both CQI and impact evaluation. White Paper 4 – *Citizen Science and Advocacy in OST* (Smith & Peck, 2020b) – presents an alternative evidence-based approach to improving both the impact and equity of OST investments.

Summary

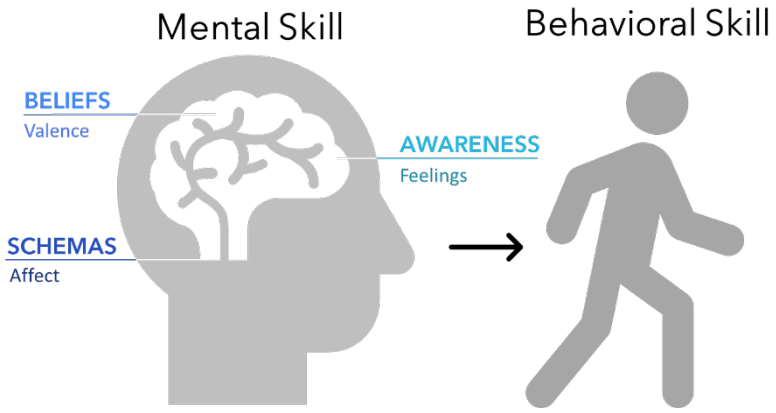
The *positivist* theory and methodology[1] used by most researchers and evaluators is poorly suited for addressing the formative explanations[2] that guide continuous quality improvement (CQI) processes and the nuanced impact models that pertain to questions about *how* and *how much*. QTurn's Quality-Outcomes Design and Methods (Q-ODM) toolbox (Peck & Smith, 2020b) was created to address fundamental problems in the evaluation of out-of-school time (OST) programs (e.g., afterschool, child care, drop-in, mentoring, tutoring, etc.). In this white paper, we extend from a framework for individual socio-emotional (SEL) skills (Peck & Smith, 2020a) to address several issues in the applied measurement of individual SEL skills.

We present steps to (a) identify the real objects we seek to represent with measurement and models (i.e., the parts of an individual's SEL skill set and the type and amount of skill change that is likely to occur during the program) and (b) produce SEL skill indicators and measures that are feasible and valid for both CQI and impact evaluation uses. With improved reasoning and evidence about the parts of SEL skill and individual skill change, we hope to help organizations produce local evidence and advocate both internally and externally for improved OST policies and increased investment.

I. Introduction

The best possible measurement choices follow from clear definitions of the objects of measurement (Cronbach & Meehl, 1955) – the "things" that assessment data reflect or indicate. The socio-emotional learning (SEL) literature provides a wide range of theories and definitions of SEL skills as objects of measurement (Berg et al., 2016; Olderbak & Wilhelm, 2020). Most of these theories and definitions have yet to be fully investigated, integrated, or presented in a way that simplifies measurement decisions. As a result, both researchers and practitioners are faced with the daunting task of making measurement decisions without clear information about (a) what SEL skills *are*, (b) which SEL skills are important in a specific type of setting, and (c) how to select measurement instruments that are most likely to produce feasible and valid information about SEL skill change, given the program design.

Figure 1. The Neuroperson Model: Four Parts of SEL Skill.



The neuroperson model (Peck & Smith, 2020a) helps us understand the necessity of addressing at least four parts of a person's integrated set of SEL skills: schemas, beliefs, awareness, and behavior (see Figure 1). Advanced forms of self-regulation and agency generally involve at least three out of these four parts, in any given instance. However, conceptualizing SEL skills mainly in terms of general domains (e.g., self-awareness, prosocial behavior), as is typical in the SEL literature, tends to obscure the unique roles played by the four key parts of the neuroperson model. In fact, because most SEL measures reflect only one or two of these key parts, some of the most important SEL skill information is neglected and may fail children on two counts: First, they may fail to identify mental parts of SEL skill that are directly associated with both high performance and misbehavior (i.e., schemas). Second, they may fail to identify mental parts of SEL skill that help children overcome disruptions and reactivity (i.e., shifting and sustaining the focus of awareness). If measures, and therefore subsequent models, fail to represent both the situation of emotional challenge and the mindful labors of children who do manage to overcome personal obstacles, then they are poorly aligned with the priorities of the OST field, which is focused on exactly these aspects of positive youth development (Eccles & Gootman, 2002; Napolatino et al., 2011).

If measures, and therefore subsequent models, fail to represent both the situation of emotional challenge and the mindful labors of children who do manage to overcome personal obstacles, then they are poorly aligned with the priorities of the OST field, which is focused on exactly these aspects of positive youth development.

In addition to theoretical confusion, we are in this situation due to the ubiquitous use of positivist evaluation methods in the OST field. The assumptions about positivist measurement are familiar: Behaviorist theory emphasizes verbally-organized concepts delivered by adults and peers as operant stimuli that influence children's behavioral responses. The same or similar behavioral responses are repeated by students when the operant stimulus appears (e.g., reading a

standardized test item). In order to measure the mental parts and processes that mediate between stimuli and responses, objective verbal facts and observable behaviors are indicated through either self-reports or adult ratings that are specified psychometrically (e.g., one indicator at a time, and normatively referenced to whoever happens to be in the study sample. The Schema and Awareness parts of SEL skill, along with their corresponding nonconscious and conscious emotional states, tend to remain unmeasured because behaviorist theory and psychometrics do not see them as *real* objects to be represented through measurement.

Finally, when it comes to making decisions with data, positivist traditions of modeling skill and skill change favor the group *average* of measured beliefs or behaviors as the best representation for each individual child in the group. Worse, where positivist evaluation models do attempt to differentiate between individual children, it is typically by reference to race/ethnicity, family socio-economic status, and other "social address" variables that have weak theoretical and empirical validity for SEL skill measurement. For example, these methods relegate exposure to stressors and cultural differences in SEL skills to the error term, effectively removing the most important information from the conversation.

Most people will recognize these assumptions as the No Child Left Behind (NCLB) paradigm of measurement and school evaluation. Within the context of NCLB, the overwhelming emphasis was on delivery of the stimulus – covering the curriculum content – regardless of the discomfort in feeling and thought that this caused for both students and staff. Unfortunately, these positivist assumptions are reflected by most SEL measures, including school climate and culture. This remains the case, even though most SEL measure developers and users want to use SEL data to create a more active and holistic subjective experience for individual children.

Realist Approach

In direct contrast to positivist methods, our approach draws upon *realist*[3] assumptions (Arocha, 2020), where the three parts of SEL mental skills have substance, knowable properties, and causal relations to behavior (e.g., SEL skills are *formal* and *final* causes of behavior and a wide variety of associated outcomes[4]). Given these assumptions, it is possible to measure and model a cascade of real causes and effects, from program quality to SEL skill growth, as in an explanation. Using measures to model an explanation about how and how much children's SEL skills grow requires empirical integration of (a) specific patterns of program quality that address (b) specific patterns of child mental skills that drive (c) specific patterns of child behavioral skills. By focusing on the pattern from a few parts of each person's mental and behavioral skill in a setting, we can generate holistic descriptions (e.g., profile measures) of each person's SEL skill set. These multi-part representations of individuals' SEL skills then, in turn, increase the validity and explanatory power of more aggregate models used to evaluate program quality and impacts.[5]

Multilevel thinking is a hallmark of realist science and has been an especially important part of our prior work on improving OST program quality at the point of service (POS) by targeting technical assistance at higher levels of the organization (e.g., organization and network levels; Smith et al., 2006, 2012, 2016a, b). Multilevel

Using measures to model an explanation about how and how much children's SEL skills grow requires empirical integration of (a) specific patterns of program quality that address (b) specific patterns of child mental skills that drive (c) specific patterns of child behavioral skills.

Unfortunately, SEL skill domains such as Emotion Management, Problem Solving, Self-Awareness, and Teamwork do not necessarily identify specific objects of measurement that are relevant, measurable, and malleable.

principles also apply to different levels of children's psychological system (aka, the "self-system") such that, for example, SEL skills operating at different levels of the self-system change on different timelines, change in response to different kinds of cues and messages, and produce qualitatively different states of emotion and agency when activated. Valid description of children's integrated SEL skill sets does not occur until these parts are modeled together, as in a holistic "profile measure" of an individual's integrated SEL skill set at a specific time point in a specific setting.

The reality is that many OST practitioners are already providing the solutions to social problems (e.g., impacting child outcomes by applying their own theories about how children learn). However, the evidence about how program staff are meeting children's needs and producing equity effects is typically unrecognized by the positivist gold standard for evaluation (e.g., behaviorist theory, psychometric precision, and randomized controlled trials). By improving the validity of theories, measures, and models, we can more easily see, understand, and document solutions that already exist.[6]

Note on Domains

Domains are category labels used to organize the many possible SEL skills into a few meaningful sets. Domain categories are immensely useful when thinking about how the guiding meanings, practices, and purposes in a program all fit together. Because SEL skills are understood most accurately in relation to specific program settings where specific kinds of meanings and behaviors occur by design, domains are a useful way to identify which practices and skills are critical to the program mission. For example, if Emotion Management is critical for younger students, then basic regulation skill-building games like "Red Light, Green Light" and "Simon Says" may be a regular part of the daily routine. Building emotion management skills with older youth may require explicit modeling of emotion management by the adult.

Unfortunately, SEL skill domains such as Emotion Management, Problem Solving, Self-Awareness, and Teamwork do not necessarily identify specific objects of measurement that are relevant, measurable, and malleable. As an example, the SEL domain of Self-Awareness might include indicators of teacher practices (e.g., all students complete self-reflection activity), child mental skills (e.g., understands emotion words, reflects on mistakes), and child behaviors (e.g., child accepts critical feedback). This combination of teacher practice, child mental skills, and child behavioral skill are often mixed into the same measurement construct (i.e., the items are averaged to indicate the child's Self-Awareness). This fusing together of information about different kinds of objects into a single score yields invalid data and is likely to obscure as much as it reveals about causes and effects, even in a simple analytic model such as a difference in means.

Table 1. SEL Domains by Context and Parts of SEL Skill.

SEL Domain	Teacher Practice	Parts of SEL Mental Skill			SEL Behavioral Skill
		Schemas	***Beliefs***	***Awareness***	
Emotion Management	Younger – Play "Red Light, Green Light" Older – Adult modeling during one on one check in	Secure attachment Fear of sharing, rejection	Self and social efficacy, emotion words	Shifting focus when excited	Moving to safe space in the room Sharing feelings when appropriate
Teamwork	Planning a group art project	Fear of social situations	Goals shared with the team	Redirect group when off task	Cooperate toward shared goals
Problem Solving	Break task down through backward mapping	Fear of failure	Self-efficacy, prerequisite knowledge for task	Mapping steps despite fear of failure, lack of efficacy, or task prerequisites	Execute the planned sequence of steps

The "situatedness" of SEL skill sets causes the different parts of SEL mental and behavioral skills to be activated by adults and children in domain-specific ways. For example, in a program where taking responsibility is a major objective of a life skills curriculum, SEL mental and behavioral skills will be expressed in ways that are related to the meaning of responsibility and responsible behavior in that setting. Table 1 provides an example of sorting the four parts of SEL skill in three common SEL skill domains with an extensive evidence-base in the OST field: Emotion Management, Teamwork, and Problem Solving. [7] Each of these domains reflects a broad set of thematically-related staff practices, youth mental skills, and youth behavioral skills, *but not a single specific object of measurement*. Rather, Table 1 suggests that there are at least five different types of skill indicator that should be sorted out in each domain. Each cell represents a different part of the integrated SEL skill in the given domain.

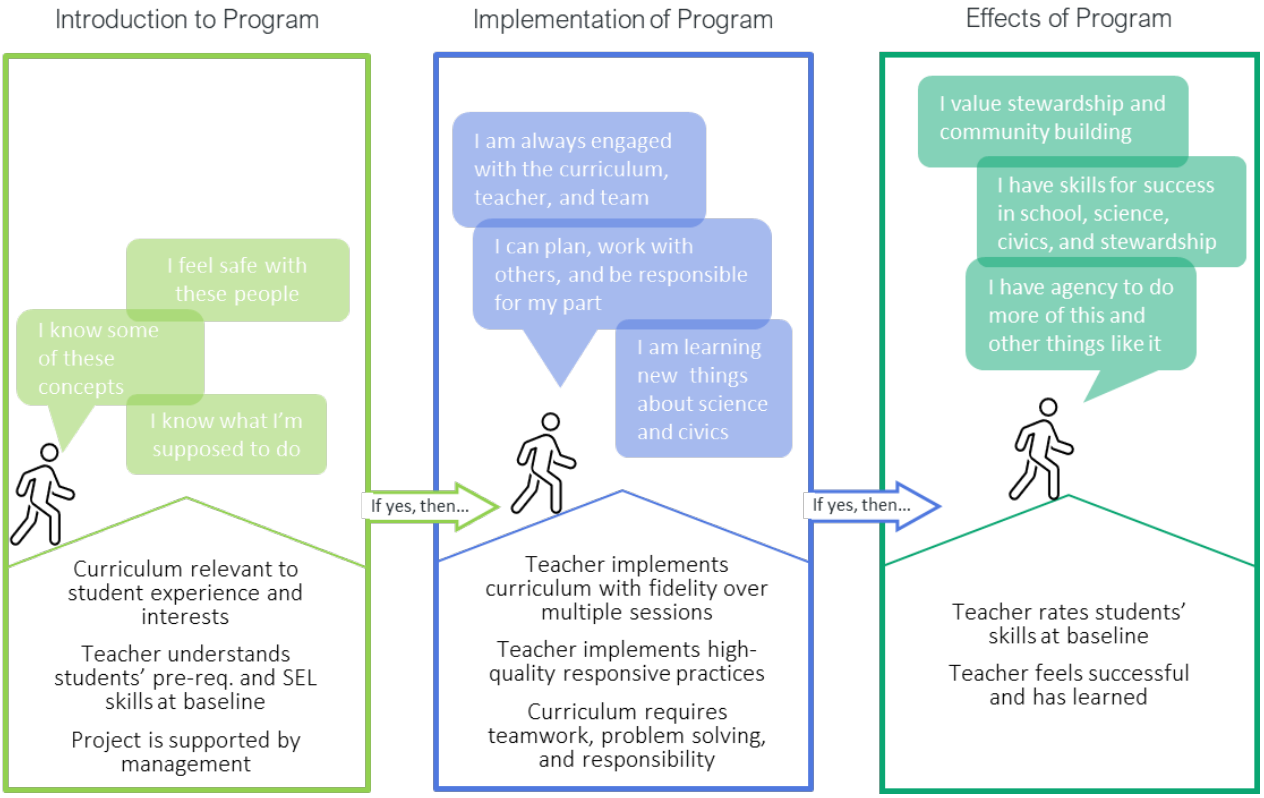
II. Selecting SEL Measures

Because most SEL measures are underdeveloped (e.g., lacking clear validation evidence, guidance for use, and supports), selecting SEL measures can involve a lot of guesswork. The following four steps reflect our best guidance on how to select SEL measures that will be characterized by both construct and consequential validity.

Step 1. Integrated Model of Skill

Integrated Models, as described by Grice (2015), focus on integrating the tangible causes and effects associated with mental and behavioral skills in specific settings. Figure 2 provides an integrated model for an OST program: the Southeastern Michigan Stewardship (SEMIS) Coalition, a regional school-based ecological stewardship project for K-12 students (Flanagan et al., 2019; G. Smith, 2016). Figure 2 shows a sequence of moments in the SEMIS theory of change: where the student and teacher first meet and introduce the skill-building opportunity, the sequence of SEMIS supports at high-fidelity implementation, and the impacts/outcomes (including equity effects) that occur during the process. In this model, only positive response states are indicated in each panel (listed in the dialogue balloons), and impacts are identifiable as mental skills (i.e., schemas, beliefs, and awareness) or behavioral skills.

Figure 2. Integrated Model of the SEMIS OST Program. Integrated Model of the SEMIS OST Program.



Prior to selecting measures, we recommend developing an integrated model so that the actual cascade of causes and effects becomes more apparent. Integrated models are a tool to improve reasoning about how to achieve desired child outcomes by accurately detecting the presence or

absence of the proximal causes of those outcomes. More precise causal theories support more valid measurement of objects in the actual cascade of causes and effects that result in SEL skill growth.

For example, in programs that focus on more therapeutic responses, or that serve only high-risk communities, the schema level of child skill (e.g., their pattern of reactivity) is likely a critical part of both intervention and assessment. Similarly, in programs focused on advanced self-regulation skills that are the developmental hallmarks of adolescence (e.g., creativity, critical thinking, planning, identity exploration), awareness (as *executive attention*) and its effects on thinking, feeling, and behavior should be an integral part of the adults mental model for what OST programs do. An integrated model requires attention to the parts, processes, and integrated SEL skill set for each individual in the sample.

Because most SEL measures do not achieve analogue-like correspondence to actual causes and effects, the explanatory power of their indicators (e.g., items and scales) is reduced for both research and CQI purposes. For example, because successively more-elaborated scales and models may compound misinformation, valid impact models are difficult to achieve. The purpose of an integrated model of staff practices, child mental skills, and child behavioral skills (along with measures aligned to the specific objects) is to create data that is a direct analogy to what is really happening for the children and adults in the OST setting.

The purpose of an integrated model of staff practices, child mental skills, and child behavioral skills (along with measures aligned to the specific objects) is to create data that is a direct analogy to what is really happening for the children and adults in the OST setting.

Step 2. Objects and Measures

The objects of interest are the parts of SEL skill, attributes of real people and settings. Each are summarily defined here, and we refer the reader to a more extensive discussion in WP1 (Peck & Smith, 2020a). Because each of the parts of SEL skill are expected to change at different rates and in response to different interventions, we also discuss a "natural" rate of "developmental" change for each part. Although these are very general guidelines, the general principle is that the longer a skill takes to change in its "natural state," the more intensive practice will be required from the intervention. Finally, we also include a note on existing measures used to indicate or represent these attributes of real individual people. This discussion of measures is offered mostly to familiarize the reader with the measurement situation for each part of skill and not to provide an exhaustive review or suggest any specific measures.

Schemas. We use the term s*chema* to describe how information (e.g., knowledge, memory) about the self and world is, in part, stored, organized, and processed in a particular area of the brain: the limbic system. More specifically, the term schema refers to non-verbal, non-symbolic, affectively-charged representations of the self and world (Peck, 2007; Peck et al., 2019),[8] as in *attachment schemas* (Bowlby, 1988). As relatively-enduring parts of the self-system, attachment schemas act like *set points* for the way children initially engage in and respond to program offerings; for this reason, we often refer to them as *basic regulation* skills. Schemas tend to change slowly, on the order of years, so a Time 2 (T2) schema measure (e.g., attachment, social phobia) sensitive to change would likely be on the order of months or years after T1. Schema-level SEL skill change is often a therapeutic goal for clinical settings. However, consistent with the social movement around trauma-informed care and learning, there is increasing pressure on OST programs to be responsive to children who have had adverse childhood experiences.

There are many kinds of measures used to indicate the status of children's attachment schemas – e.g., self-report, interview, and projective measures of child and adolescent attachment schemas (Dwyer, 2005; Jewell et al., 2019) – most of which may be too burdensome for OST programs. Relatively brief self-report measures of feelings of belonging, fear of abandonment, social phobia, or rejection sensitivity can be used to indicate the status of children's attachment schemas. Baseline assessments can also include information about young people's family and community situations, histories of education and enrichment experiences, and exposure to stressful life events (e.g., adverse childhood experiences) as proxies for the schema-level parts of SEL skill. On the behavioral side, baseline issues at the schema-level are tangibly observable by teachers in the process of coming to know children (e.g., shy, aggressive) and, for this reason, basic forms of self-regulation are often successfully rated by program staff and youth workers by simply observing children's behavior during program sessions.

Beliefs. We use the term *belief* to describe how information (e.g., knowledge, memory) is stored, organized, and processed in a second area of the brain: the neocortex. The term belief refers to verbal-symbolic representations of the self and world.[9] Basic beliefs (i.e., *beliefs in* and *beliefs about* a thing; Fishbein & Raven, 1962) are the most fundamental unit of information in the neocortex and are stored as long-term memory. Basic beliefs differentiate and integrate across time to form higher-order belief systems, such as *attitudes* (i.e., a belief in a thing integrated with a belief about the goodness or badness of that thing) that combine to form goals that combine to form plans.

Beliefs about the self (e.g., I can do this), beliefs about the world (e.g., school is important), and beliefs about the self in relation to the world (e.g., I could get better at math if my teacher was good) become more complex, and become embedded in belief systems (e.g., goals, plans, and ideologies), across developmental time. Beliefs are also relatively enduring, but on the order of months, so a T2 belief assessment sensitive to change could be on the order minutes, days, or months.

Basic beliefs differentiate and integrate across time to form higher-order belief systems, such as attitudes (i.e., a belief in a thing integrated with a belief about the goodness or badness of that thing) that combine to form goals that combine to form plans.

There are many youth self-report measures used to indicate SEL-related beliefs (e.g., SEL word definitions, SEL efficacy self-assessments) and behaviors (e.g., self-reports of prosocial behavior); indeed, these are the most readily available types of SEL skill measures. It is perhaps not surprising then that commercially available SEL curricula overwhelmingly focus on group-based discussions of SEL concepts (cf. Jones et al., 2019).

Awareness. The term *awareness* (or *executive functions*) refers specifically to consciously controlling the focus of awareness in relation to immediate thoughts and feelings.[10] Currently activated schemas and beliefs produce thoughts and feelings, which may or may not enter an individual's conscious awareness. Executive functions (e.g., shifting and sustaining the focus of awareness) operate directly on thoughts and feelings to provide the basis for all forms of self-reflection (e.g., secondary appraisal,[11] planning, and the effortful control of impulses). In this view, shifting and sustaining the focus of awareness are the mental skills that allow individuals to participate *intentionally* in their own learning and development.

Although the skills of consciously shifting and focusing awareness appear to be at least as, or even more, enduring than schemas,[12] evidence of successful interventions demonstrates that practicing these skills can yield tangible improvements in the abilities to shift and sustain the focus of awareness (Basso et al., 2019; Brefczynski-Lewis et al., 2007; Fortenbaugh et al., 2017; Lutz et al., 2009; MacLean et al., 2010; Roeser & Pinela, 2014; Tang & Leve, 2016; Zanesco et al., 2016). There are many measures used to

indicate executive functions, focused both on mental skills (e.g., Blair et al., 2005; Chan et al., 2008; Isquith et al., 2014; Mullane et al., 2016; Nyongesa et al., 2019; Raver et al., 2013; Suchy, 2009) and the

Even where no measure of awareness is used, the most important thing is to understand the critical role that awareness skills play in developing all other skills and outcomes.

behavioral manifestations of executive functions (e.g., Gioia et al., 2001; Rescorla et al., 2019; Van der Elst et al., 2012). Most of these instruments would probably also be considered too burdensome to be used within the context of OST programs.

As with schema assessment, we recommend using the best feasible measure to indicate children's ability to shift and sustain the focus of awareness. In practice, this strategy may reduce to using a few indicators of the extent to which children are able to pay continuous attention to, and avoid or recover from distractions during, program activities. Even where no measure of awareness is used, the most important thing is to understand the critical role that awareness skills play in developing all other skills and outcomes. For example, as has been amply demonstrated (e.g., Barker & Munakata, 2015; Diamond & Lee, 2011; Liew, 2012; Meltzer, 2018; Riggs et al., 2006), there are a wide range of curriculum strategies available for helping children develop awareness skills that do not necessarily require perfect, or even any, direct measures of executive functions.

Behavior. Only by defining and measuring child behavior as a distinct object can we model the causal relations between mental skills and behavior or between behavior and subsequent outcomes (e.g., grades). The term *behavior* refers to both brief bodily expressions (e.g., talking, smiling) and patterns of bodily expression that extend over longer time periods (e.g., cooperating, responsibility, studying). Behavioral rating scales should be focused on the specific behaviors enacted by children in a specific circumstance. Behavioral measures are widely available and, generally, both observational (e.g., Smith et al., 2016a) and direct (e.g., McKown et al., 2013, 2019) assessments of behavioral skills (a) are the most tangible outcomes desired by OST program managers and staff, (b) are generally more feasible to implement than most mental measures, and (c) have ethical advantages over other measurement strategies.[13]

Step 3. Indicator Purpose

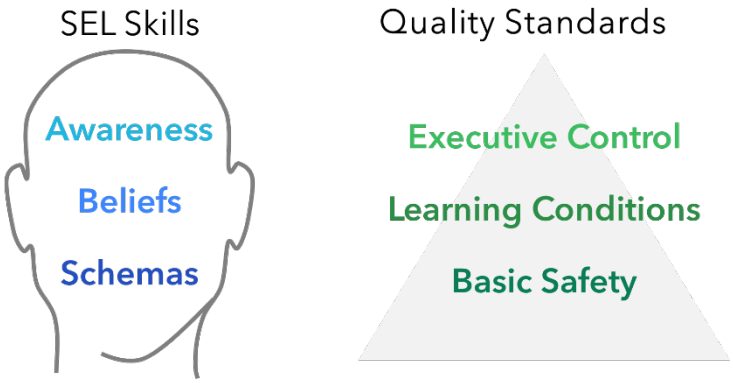
When reviewing SEL measures, it is critical to review all of the measurement indicators[14] to make sure that they are aligned with evaluation purposes (e.g., relevant characteristics of children are reflected by the indicators). In addition, scale scores and resulting statistical estimates should be interpreted with caution because, unfortunately, the positivist psychometric assumption of equivalent distance among scale points – that is, that psychological and behavioral constructs can be treated as interval or ratio – is rarely met (Michell, 2003). In White Paper 3 (Peck & Smith, 2020b), we discuss the critical role of *quasi-absolute scaling* in helping to make meaning from the ordinally- and categorically-structured objects that most social science data actually represent. Here, we consider distinctions that are simpler to address yet critical for measure and model validity; that is, indicator fit with evaluation purposes.

First, by distinguishing indicators that focus on mental versus behavioral skills, the validity of scales can be quickly evaluated, and decisions can be made about reconfiguring indicators to better reflect the relevant parts of SEL skill. Second, by distinguishing the two different ways a skill can be expressed, as

functional versus optimal skill levels (Fischer et al., 2006), greater confidence in the validity of skill growth models can be obtained.

Mental or Behavioral. A core finding in the science of child development is that targeting and promoting SEL mental skills will result in improvements to SEL behavioral skills where opportunities to enact those behaviors are also provided. Understanding the SEL mental skills that children bring to OST programs (e.g., an avoidant attachment schema coupled with a goal to make new friends) is key to promoting their most advanced regulation skills (e.g., understanding how to use their awareness to modify their avoidant feelings) and demonstrating new behavioral skills (e.g., studying with a partner). Getting to know children's mental skills in a nuanced way – through past experiences, current interests, academic beliefs, etc. – is a key aspect of OST best practice. Unfortunately, if measurement indicators for these different parts of SEL skill are fused together (e.g., as an average score), the validity of the measure and subsequent models is compromised, failing to represent the nuanced reality of each child's unique set of mental skills.

Figure 3. Parts of SEL Skill and Standards for Quality.



Where the theory linking mental and behavioral skills is strong, behavioral skill scores can sometimes be used as valid indicators of mental skills. For example, because children's behavior is a manifestation of mental skill engagement in setting-specific tasks and relationships, behavior can be a meaningful indicator of basic and advanced mental skills, even though it may not reveal specific information about those mental skills.[15] Further, given the lack of specificity about schemas and awareness in most SEL theoretical frameworks, clear and valid inferences about the schema and awareness parts of SEL skill may be particularly challenging.

Functional skill is the best an individual can do with no additional supports from the setting.

Optimal skill is the best someone can do while receiving the highest-quality supports.

Functional or Optimal. *Functional skill* is the best an individual can do with no additional supports from the setting. Supports, in this sense, might be an adult who demonstrates how to do something or provides encouragement at moments of struggle. *Optimal skill* is the best someone can do while receiving the highest-quality supports.[16] Children's functional SEL skills are relatively-enduring SEL skill *traits* that are transferred into and out of any setting that they may enter. Functional skills are not static, but changes require opportunity, practice, and relative mastery. In contrast, children's optimal SEL skills are *states* of active skill engagement that occur where staff are actively scaffolding children to their highest skill level.

The relationship between functional and optimal can be pictured in the sawtooth shape of skill growth, where peaks of well-supported optimal skill punctuate a more gradually-sloping and lower functional skill line. This relationship reflects a developmental dynamic in which pushing optimal skill levels higher and higher causes functional skill levels to rise.

Measures targeting optimal skill levels are better fit to the type of skill change that actually occurs during OST programs, where staff intentionally scaffold students.

Figure 3 reflects the alignment of OST quality standards with the specific parts of SEL skill, suggesting that over the course of OST program sessions, children's SEL skills should grow increasingly beyond their normal unsupported levels. This access to supports, where peak optimal skill levels are achieved, is the fundamental equity issue that most OST programs were designed to address: giving less advantaged children access to optimal experiences that drive functional skill levels upward. The functional SEL skill level is what transfers to other settings and influences other outcomes.

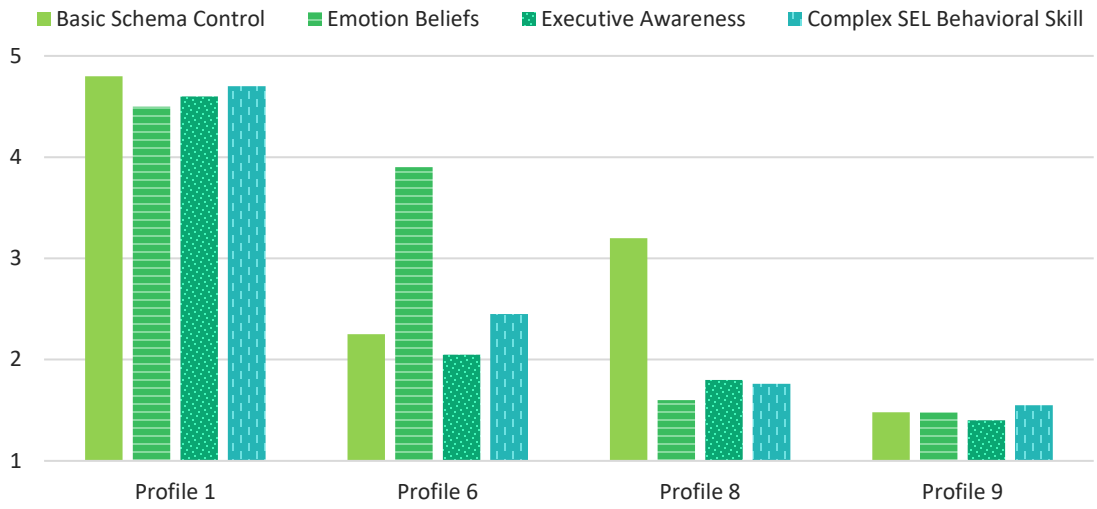
Unfortunately, while functional skill indicators are less sensitive to SEL skill growth, the majority of published SEL measures recommended to the OST field seem to be of this type.

The *referent* of an indicator is also important for understand which skill level the indicator targets. The general issue is that indicators that reference a child's skill state in a specific place and time in which optimal performance is expected – like an OST program offering – are likely to reflect an optimal skill level. Conversely, similar survey items that refence skill "in general" – at any time and in any space – reflect the more stable trait-like functional skill level. Unfortunately, while functional skill indicators are less sensitive to SEL skill growth, the majority of published SEL measures recommended to the OST field seem to be of this type (cf. Berg et al., 2017; Child Trends, 2014; Humphrey et al., 2011; Jones et al., 2019).

Step 4. Profile Measures

The final step uses pattern-centered thinking to create a "profile measure" of the integrated SEL skill set for each individual. Profile measures can include separate indicators for different parts of SEL skill sets, are valid measures of an individual's SEL skill set at a specific moment and context, and can be used to create valid skill change variables (e.g., growth, stability, & decline). Figure 4 shows partial baseline results for a group of over 1,100 children from 23 OST programs. The SEL skill measures map onto the four SEL skill parts – schemas, beliefs, awareness, behavior – and each grouping of bars represents a subgroup of individual children who share a specific pattern that represents their integrated SEL skill set. Profile 1 represents a group of students with very strong SEL skills and can be contrasted with the other three lower-skill profiles. Profile 6 represents elementary aged children who enter the OST setting with some emotion knowledge but who cannot manage their bodies or focus their attention very effectively. Profile 8 represents children who have basic self-control but appear to be withdrawn from engagement with the setting. The children in Profile 9 are struggling on all fronts. Why should this figure give us pause?

Figure 4. SEL Skill Profiles from Adult Ratings of Optimal Mental and Behavioral Skills.



Source: Lindeman et al. (2019). N = 1159 youth, ages 8 to 18. Selected profiles represent: 1 = 11%, 6 = 10%, 8 = 10%, and 9 = 12% of the sample. Original variable names: Behaviorally Manages Emotion, Expresses Emotion Knowledge, Displays Goal-Striving Mastery, and Displays Social-Role Mastery.

First, consider the average of a total score derived by combining these measures: It would accurately represent none of the children in the group. Second, consider the two moderate profiles (Profiles 6 and 8). Children from these two different profiles require different supports – the children in Profile 6 are externalizing and the children in profile 8 are internalizing – but their average total score would be nearly identical. Further, using either the Emotion Beliefs indicator in Profile 6 or the Basic Schema Control indicator in Profile 8, it is possible to rate the children in each profile as having moderately-high skill when, in fact, they scored low on all other parts of the integrated SEL skill set.

One of the critical purposes in this step is reflected by the fact that having more accurate information about individuals translates into having more accurate information about aggregates formed from these individuals.

One of the critical purposes in this step is reflected by the fact that having more accurate information about individuals translates into having more accurate information about aggregates formed from these individuals. For example, with more realistic and valid information about the SEL skills of individual children at baseline, it is then possible to see the *prevalence of children in the lower SEL skill profile, both within and across sites*. In a program setting where 70% of the children are not successfully self-regulating, there are very different and immediate needs in that setting (e.g., more staff) compared to a setting where only 10% of children are having a difficult time engaging with the program content.

III. Recommendations

These distinctions are critical to selecting measures that (a) target the right skill objects, (b) yield valid data and models, and (c) are sensitive to skill changes that actually occur.

Our first recommendation is to encourage managers and evaluators to lead their staff teams through the sequence of steps outlined above; for example, to think through the specific parts of SEL skill likely to grow and whether those skills are best represented as mental or behavioral, functional or optimal. These distinctions are critical to selecting measures that (a) target the right skill objects, (b) yield valid data and models, and (c) are *sensitive* to skill changes that actually occur. The selected measures must also be *feasible* to administer, meaning both providers and children must have the time, motivation, and infrastructure necessary to complete the measurement process. For example, despite having selected measures with evidence of reliability and validity, it may not be feasible to use an on-line version of those measures to assess mental skills if they require an hour to complete and/or if providers have limited access to computers. The objective is *validity and feasibility.*

Several additional recommendations for SEL skill measurement in OST settings follow. First, for the program to be fully responsive to each child and achieve optimal skill levels for all children, a valid baseline measure of SEL skills is practically useful. Although adults who work with children must make their own informal ratings in order to do the work every day, information from a valid measure at baseline adds value to adults' informal perceptions and speeds up the process and precision of responding to children, especially those with challenging SEL histories. Baseline data combined with data from a second time point makes it possible to ensure that children who enter programs with low skills experience skill growth and that children with high skills do not lose ground in chaotic or nonresponsive program settings.

Second, for evaluations that seek to understand SEL skill change, we recommend adult ratings of children's SEL behavioral skills at multiple timepoints, as indicated in the upper right hand cell of Figure 5. These ratings should reflect optimal skill levels and are more sensitive to variations in setting quality. The same measures should be used again at subsequent time points to assess change in SEL skills (e.g., behavioral skill ratings are repeated after at least three months of program participation). Deriving estimates of change from ratings of

Figure 5. Recommended Measures by Type and Purpose

	Mental	Behavior
Optimal	Requires a performance task	**Recommended for feasibility and sensitivity to program effects on SEL Skill**
Functional	**Youth Survey**	Requires a rating from home, school, etc.

children's optimal SEL behavioral skill at the point of service is ideal for generating impact estimates of the effects of OST program quality on children's SEL skill growth. These impact estimates are then powerful moderators in evaluations seeking to understand impact on transfer outcomes during the school day and other settings (see White Paper 3 [Peck & Smith, 2020b] for a more detailed discussion).

Finally, the traditional OST focus on basic safeguarding of physical and emotional safety has only become more important during the COVID-19 pandemic. Although we recommend a compassionate approach to evaluation that minimizes the burden on children, self-report measures of children's feelings can be a useful approach. For example, per the bottom left cell in Figure 5, we recommend using three subjective well-being (SWB) items during virtual sessions to get a quick idea about how children are doing. The items reflect functional schema-level emotion management skills and were selected to fulfill requests for a few questions that staff could ask children to check on their mental health and well-being in virtual learning environments, often the child's own home. For more information on these SWB items, and additional adult observational rating tools (upper right cell of Figure 5) for staff to use during virtual sessions, see http://www.qturngroup.com/ourtools/.

References

Arocha, J. F. (2020). Scientific realism and the issue of variability in behavior. Theory & Psychology, 1-20. doi.org/10.1177%2F0959354320935972

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology*, *20*(1), 385–406.

Bailey, J. R., & Eastman, W. N. (1994). Positivism and the promise of the social sciences. *Theory & Psychology*, *4*(4), 505–524.

Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin*, *112*(3), 461–484.

Barker, J. E., & Munakata, Y. (2015). Time isn't of the essence: Activating goals rather than imposing delays improves inhibitory control in children. *Psychological Science*, *26*(12), 1898–1908.

Barnard, P. J., Duke, D. J., Byrne, R. W., & Davidson, I. (2007). Differentiation in cognitive and emotional meanings: An evolutionary analysis. *Cognition and Emotion*, *21*, 1155-1183.

Basso, J. C., McHale, A., Ende, V., Oberlin, D. J., & Suzuki, W. A. (2019). Brief, daily meditation enhances attention, memory, mood, and emotional regulation in non-experienced meditators. *Behavioural Brain Research*, *356*, 208–220.

Berg, J., Osher, D., Same, M. R., Nolan, E., Benson, D., Jacobs, N. (2017). *Identifying, defining, and measuring social and emotional competencies*. Washington, DC: American Institutes for Research.

Bertalanffy, L. von (1968). *General system theory: Foundations, development, and applications* (revised edition). New York: Braziller.

Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The Measurement of Executive Function in Early Childhood. *Developmental Neuropsychology*, *28*(2), 561–571.

Bowlby, J. (1988). *A secure base: Parent-child attachment and healthy human development.* New York: Basic Books.

Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B., & Davidson, R. J. (2007). Neural correlates of attentional expertise in long-term meditation practitioners. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *104*(27), 11483–11488.

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist, 19*(1), 1-15.

Buchanan, D. R. (1998). Beyond positivism: Humanistic perspectives on theory and research in health education. *Health Education Research: Theory & Practice, 13*(3), 439-450.

Bunge, M. (2014). *Chasing reality: Strife over realism*. University of Toronto Press.

Burton-Jones, A., & Lee, A. S. (2017). Thinking about measures and measurement in positivist research: A proposal for refocusing on fundamentals. *Information Systems Research, 28*(3), 451-467.

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201–216.

Child Trends (2014). *Measuring Elementary School Students' Social and Emotional Skills: Providing Educators with Tools to Measure and Monitor Social and Emotional Skills that Lead to Academic Success*. Publication #2014-37. https://www.childtrends.org/

Chirkov, V., & Anderson, J. (2018a). Statistical positivism versus critical scientific realism. A comparison of two paradigms for motivation research: Part 1. A philosophical and empirical analysis of statistical positivism. *Theory & Psychology*, *28*(6), 712–736. doi.org/10.1177/0959354318804670

Chirkov, V., & Anderson, J. (2018b). Statistical positivism versus critical scientific realism. A comparison of two paradigms for motivation research: Part 2. A philosophical and empirical analysis of statistical positivism. *Theory & Psychology*, *28*(6), 737–756. doi.org/10.1177/0959354318816829

Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology*, *8*(4), 597–600.

Cronbach, L. J. (1949). *Essentials of psychological testing*. New York: Harper.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge University Press.

Derryberry, D., & Rothbart, M. K. (1997). Reactive and effortful processes in the organization of temperament. *Development and Psychopathology*, *9*(4), 633–652.

Deutsch, R., & Strack, F. (2006). Duality Models in Social Psychology: From Dual Processes to Interacting Systems. *Psychological Inquiry*, *17(3),* 166-172.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959–964.

Dwyer, K. M. (2005). The Meaning and Measurement of Attachment in Middle and Late Childhood. *Human Development*, *48*(3), 155–182.

Eccles, J. S., & Gootman, J. A. (2002). *Community programs to promote youth development*. Washington, DC: National Academy Press.

Epstein, S. (2003). Cognitive-experiential self-theory of personality. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology: Personality and social psychology, Vol. 5* (pp. 159-184). Hoboken, NJ: Wiley.

Federico, F., Marotta, A., Martella, D., & Casagrande, M. (2017). Development in attention functions and social processing: Evidence from the Attention Network Test. *British Journal of Developmental Psychology*, *35*(2), 169–185.

Fischer, K. W., Rose, L. T., & Rose, S. P. (2006). Growth cycles of mind and brain: Analyzing developmental pathways of learning disorders. In K. W. Fischer, J. H. Bernstein, & M. H. Immordino-Yang (Eds.), *Mind, brain, and education in reading disorders*. Cambridge, U.K.: Cambridge University Press.

Flanagan, C., Gallay, E., Pykett, A., & Smallwood, M. (2019). The environmental commons in urban communities: The potential of place-based education. *Frontiers in Psychology*, *10*. https://doi-org.proxy.lib.umich.edu/10.3389/fpsyg.2019.00226

Fortenbaugh, F. C., Degutis, J., & Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences, 1396,* 70–91.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *86*, 186–204.

Gioia, G. A., Isquith, P. K., & Guy, S. C. (2001). Assessment of executive functions in children with neurological impairment. In R. J. Simeonsson & S. L. Rosenthal (Eds.), *Psychological and developmental assessment: Children with disabilities and chronic conditions* (pp. 317–356). Guilford.

Grice, J. W. (2015). From means and variances to persons and patterns. *Frontiers in Psychology*, *6*. https://doi-org.proxy.lib.umich.edu/10.3389/fpsyg.2015.01007

Haig, B. D., & Evers, C. W. (2015). *Realist inquiry in social science*. Sage.

Humphrey, N., Kalambouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., & Wolpert, M. (2011). Measures of social and emotional skills for children and young people: A systematic review. *Educational and Psychological Measurement*, *71*(4), 617–637.

Isquith, P. K., Roth, R. M., & Gioia, G. A. (2014). Assessment of executive functioning using tasks of executive control. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functioning.* (pp. 333–357). Springer Science + Business Media.

Izard, C. E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology*, *60*, 1-25.

Jewell, T., Gardner, T., Susi, K., Watchorn, K., Coopey, E., Simic, M., Fonagy, P., & Eisler, I. (2019). Attachment measures in middle childhood and adolescence: A systematic review of measurement properties. *Clinical Psychology Review*, *68*, 71–82.

Jones, S., Baily, R., Brush, K., & Nelson, B. (2019). Introduction to the taxonomy project: Tools for selecting & aligning SEL frameworks. https://measuringsel.casel.org/wp-content/uploads/2019/02/Frameworks-C.1.pdf

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251.

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*(2), 220–232.

Kosslyn, S. M., & Pomerantz, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, *9*(1), 52–76.

Liew, J. (2012). Effortful control, executive functions, and education: Bringing self-regulatory and social-emotional competencies to the table. *Child Development Perspectives*, *6*(2), 105–111.

Lindeman, L. M., Smith, C., Peck S. C., & Harvey, S. L. (2019). *Evidence of positive impact on afterschool and summer programs in Palm Beach County: A 10-year validation study of Prime Time's integrated quality improvement system*. Boynton Beach, FL: Prime Time Palm Beach County. https://simplebooklet.com/10yearvalidationstudy

Mackenzie, B. (1977). *Behaviourism and the limits of scientific method.* Atlantic Highlands, NJ: Humanities.

MacLean, K. A., Ferrer, E., Aichele, S. R., Bridwell, D. A., Zanesco, A. P., Jacobs, T. L., King, B. G., Rosenberg, E. L., Sahdra, B. K., Shaver, P. R., Wallace, B. A., Mangun, G. R., & Saron, C. D. (2010). Intensive meditation training improves perceptual discrimination and sustained attention. *Psychological Science*, *21*(6), 829–839.

McKown, C., Allen, A. M., Russo-Ponsaran, N. M., & Johnson, J. K. (2013). Direct assessment of children's social-emotional comprehension. *Psychological Assessment*, *25*(4), 1154–1166.

McKown, C., et al. (2019). *Student social and emotional competence assessment: The current state of the field and a vision for its future.* Assessment Work Group: Collaborative for Academic, Social, and Emotional Learning.

Meltzer, L. (2018). Creating strategic classrooms and schools: Embedding executive function strategies in the curriculum. In L. Meltzer (Ed.), *Executive function in education: From theory to practice., 2nd ed.* (pp. 263-299). Guilford.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749.

Michell, J. (2003). Pragmatism, positivism and the quantitative imperative. *Theory & Psychology*, *13*(1), 45–52.

Michigan Department of Education (2017). Early Childhood to Grade 12 Social and Emotional Learning (SEL) Competencies and Indicators. https://www.michigan.gov/documents/mde/SEL_Competencies-_ADA_Compliant_FINAL_605109_7.pdf

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14.

Mullane, J. C., Lawrence, M. A., Corkum, P. V., Klein, R. M., & McLaughlin, E. N. (2016). The development of and interaction among alerting, orienting, and executive attention in children. *Child Neuropsychology*, *22*(2), 155–176.

Napolitano, C. M., Bowers, E. P., Gestsdóttir, S., & Chase, P. A. (2011). The development of intentional self-regulation in adolescence: Describing, explaining, and optimizing its link to positive youth development. In R. M. Lerner, J. V. Lerner, & J. B. Benson (Eds.), *Advances in child development and behavior, Vol 41: Positive youth development* (pp. 19-38). Elsevier.

Nyongesa, M. K., Ssewanyana, D., Mutua, A. M., Chongwo, E., Scerif, G., Newton, C. R. J. C., & Abubakar, A. (2019). Assessing executive function in adolescence: A scoping review of existing measures and their psychometric robustness. *Frontiers in Psychology*, *10*. https://doi-org.proxy.lib.umich.edu/10.3389/fpsyg.2019.00311

Olderbak, S., & Wilhelm, O. (2020). Overarching principles for the organization of socioemotional constructs. *Current Directions in Psychological Science*, *29*(1), 63–70.

Peck, S. C. (2007). TEMPEST in a gallimaufry: Applying multilevel systems theory to person-in-context research. *Journal of Personality, 75*, 1127-1156.

Peck, S. C. (2016). *The multilevel systems problem of gene-by-environment "interactions."* London, UK: The Early Intervention Foundation.

Peck, S. C. (2018). Revisiting Erikson's neglected concepts of ego-identity and self-identity: Hidden persuaders in the twilight zones of awareness. *Archives of Psychology, 2(1),* 1-25.

Peck, S. C., & Smith, C., (2020a). *Socio-Emotional Skills, Quality, and Equity: The Multilevel Person-in-Context~neuroperson (MPCn) Framework.* [White Paper #1]. QTurn.

Peck, S. C., & Smith, C. (2020b). *Realist(ic) Evaluation Tools for OST Programs: The quality-outcomes design and methods (Q-ODM) toolbox* [White Paper #3]. QTurn.

Peck, S. C., Smith, C., & Smith, L. (2019). *The Multilevel Person-in-Context~Neuroperson (MPCn) Model: Guidance for Quality Improvement Systems (QIS) Focused on Socio-Emotional Skill Growth and Transfer Outcomes* [Working Paper #1]. QTurn. https://qturngroup.com/MPC*n*

Raver, C. C., McCoy, D. C., Lowenstein, A. E., & Pess, R. (2013). Predicting individual differences in low-income children's executive control from early to middle childhood. *Developmental Science*, *16*(3), 394–408.

Rescorla, L. A., Blumenfeld, M. C., Ivanova, M. Y., & Achenbach, T. M. (2019). International comparisons of the dysregulation profile based on reports by parents, adolescents, and teachers. *Journal of Clinical Child and Adolescent Psychology*, *48*(6), 866–880.

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social-emotional competence. *Journal of Applied Developmental Psychology*, *27*(4), 300–309.

Roeser, R. W., & Peck, S. C. (2009). An education in awareness: Self, motivation, and self-regulated learning in contemplative perspective. *Educational Psychologist, 44*, 119-136.

Roeser, R. W., Peck, S. C., & Nasir, N. S. (2006). Self and identity processes in school motivation, learning, and achievement. In P. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd edition) (pp. 391-424). Mahwah, NJ: Erlbaum.

Roeser, R. W., & Pinela, C. (2014). Mindfulness and compassion training in adolescence: A developmental contemplative science perspective. In E. Oberle & K. Schonert-Reichl (Eds.), *Mindfulness in Adolescence.* (Vol. 142, pp. 9–30). Jossey-Bass.

Rothbart, M. K. (2007). Temperament, development, and personality. *Current Directions in Psychological Science*, *16*(4), 207–212.

Rueda, M. R. (2013). Development of attention. In K. N. Ochsner & S. Kosslyn (Eds.), *The Oxford Handbook of Cognitive Neuroscience* (Vol.1, pp. 296-316). DOI: 10.1093/oxfordhb/9780199988693.013.0015

Schultheiss, O. C. (2001). An information processing account of implicit motive arousal. In M. L. Maehr & P. Pintrich (Eds.), *Advances in motivation and achievement*: *Vol. 12. New directions in measures and methods* (pp. 1-41). Greenwich, CT: JAI Press.

Smith, C., Akiva, T., & Henry, B. (2006, March). *Quality in the out-of-school time sector: Insights from the Youth PQA Validation Study*. Paper presented at the biennial meeting of the Society for Research on Adolescence, San Francisco, CA.

Smith, C., Akiva, T., Sugar, S., Lo, Y. J., Frank, K. A., Peck, S. C., & Cortina, K. S. (2012). Continuous quality improvement in afterschool settings: Impact findings from the *Youth Program Quality Intervention study*. Ypsilanti, MI: David P. Weikart Center for Youth Program Quality.

Smith, C., McGovern, G., Larson, R. W., Hillaker, B., & Peck, S. C. (2016a). *Preparing youth to thrive: Promising practices for social emotional learning*. Ypsilanti, MI: David P. Weikart Center for Youth Program Quality.

Smith, C., McGovern, G., Peck, S. C., Larson, R. W., Hillaker, B., & Roy, L. (2016b). *Preparing youth to thrive: Methodology and findings from the social and emotional learning challenge*. Ypsilanti, MI: David P. Weikart Center for Youth Program Quality.

Smith, C., & Peck, S. C. (2020). *Citizen science for OST evidence and advocacy* [White Paper #4]. QTurn. https://qturngroup.com/knowledgecenter/white-paper-4/

Smith, G. (2016). Michigan's Great Lakes Stewardship Initiative: A Statewide Place-Based Education Effort to Connect Schools to Communities. *LEARNing Landscapes*, *10*(1), 289-306.

Staats, A. W. (1991). Unified positivism and unification psychology: Fad or new field? *American Psychologist*, *46*(9), 899–912.

Steele, A., Karmiloff, S. A., Cornish, K., & Scerif, G. (2012). The multiple subfunctions of attention: Differential developmental gateways to literacy and numeracy. *Child Development*, *83*(6), 2028–2041.

Suchy, Y. (2009). Executive functioning: Overview, assessment, and research issues for non-neuropsychologists. *Annals of Behavioral Medicine*, *37*(2), 106–116.

Tang, Y.-Y., & Leve, L. D. (2016). A translational neuroscience perspective on mindfulness meditation as a prevention strategy. *Translational Behavioral Medicine*, *6*(1), 63–72.

Thomas, A., & Chess, S. (1977). *Temperament and development*. Brunner/Mazel.

Van der Elst, W., Ouwehand, C., van der Werf, G., Kuyper, H., Lee, N., & Jolles, J. (2012). The Amsterdam executive function inventory (AEFI): Psychometric properties and demographically corrected normative data for adolescents aged between 15 and 18 years. *Journal of Clinical and Experimental Neuropsychology*, *34*(2), 160–171.

Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Weissberg, R. P., Kumpfer, K. L., & Seligman, M. E. P. (2003). Prevention that works for children and youth: An introduction. *American Psychologist*, *58*(6–7), 425–432.

Zanesco, A. P., King, B. G., MacLean, K. A., Jacobs, T. L., Aichele, S. R., Wallace, B. A., Smallwood, J., Schooler, J. W., & Saron, C. D. (2016). Meditation training influences mind wandering and mindless reading. *Psychology of Consciousness: Theory, Research, and Practice*, *3*(1), 12–33.

Notes

[1] We refer to variable-centered theory and methods, psychometrics, and generalized linear modeling strategies in particular (e.g., ANOVA, regression) as *positivist* theory and methodology. As the discipline of history and philosophy of science has described (Arocha, 2020; Baily & Eastman, 1994; Buchanan, 1998; Burton-Jones & Lee, 2017; Chirkov & Anderson, 2018a, 2018b; Michell, 2003; Staats, 1991), application of positivist methods results in pacification of the research participant by limiting the range of performance to behavior and simple beliefs (Danziger, Mackenzie, 1977), in part because of the tight linkage between positivist methodology and the psychological ideology of behaviorism that, in its modernized form, also emphasizes a reductionist perspective (in this case *operationism*) on mental skills.

[2] *Formative explanations* describe the specific causes and effects characterizing the dynamic interactions among the objects of study (e.g., how different kinds of staff behavior at the point of service are expected to promote or undermine different kinds of child mental and behavioral engagement at the point of service) as they apply to specific individuals (e.g., staff person A in relation to child B during program-offering activity C).

[3] According to Arocha (2020), "all scientific realists adopt at least two basic theses: one ontological and one epistemological (Bunge, 2014; Haig & Evers, 2015). The ontological thesis is that the world is real and it exists independently of our knowledge of it. The epistemological thesis is that reality is knowable, albeit often approximately and mostly indirectly" (p. 3).

[4] Realist approaches tend to reflect an Aristotelian fourfold understanding of causal explanation: material, efficient, formal and final. The positivist approach to measurement tends to struggle with formal and final causes because it ignores the interior structure of mental skills.

[5] See Peck & Smith (2020b) for a more detailed discussion. Our approach to measurement and impact modelling draws upon marginalized theory and methodology for dealing with (a) the ordinal and categorical nature of mental and behavioral skills, (b) the formative rather than reflective relation of SEL behavioral skill indicators to their constructs, (c) the use of integrated models to describe multilevel, person-in-context patterns of SEL skill growth, (d) the use of pattern-centered analytics, with replication, as a more informative "gold standard" for understanding the impact of OST program features (e.g., instructional quality) on SEL skill growth and a wide variety of subsequent outcomes (e.g., academic achievement).

[6] Unlike the positivist paradigm for precision and meaningfulness, we do not assume that all but the simplest mental skills are unknowable or that mental skills can be meaningfully represented in the aggregate for any but a strictly homogenous population (of which there are few in the education field). In our view, where seeking explanations about the causes of children's behavior and skill growth, mental SEL skills must be conceptualized explicitly in the causal model and assessed with measures that indicate the presence or absence of specific SEL skills for each individual.

[7] The work of Reed Larson and colleagues provided the primary evidence base for developing these SEL skill domain labels and definitions. A list of published work related to each domain can be found in Smith et al., 2016a; a complete list of Larson's work in this area can be found at http://youthdev.illinois.edu/).

[8] The schema system in the neuroperson model corresponds to what we have described elsewhere as the iconic representation system (Peck, 2007, 2016, 2018; Peck et al., 2019; Roeser & Peck, 2009; Roeser et al., 2006; Smith et al., 2016b) and is similar to Barnard et al.'s (2007) *implicational meaning* system,

Epstein's (2003) *experiential* system, Deutsch and Strack's (2006) *impulsive* system, Kahneman & Tversky's (1973) *intuitive* system, Bowlby's (1988) *working model* system, Baldwin's (1992) *relational schema* system, and Izard's (2009) *emotion schema* system.

[9] The belief system in the neuroperson model corresponds to what we have described elsewhere as the symbolic representation system (Peck, 2007, 2016, 2018; Peck et al., 2019; Roeser & Peck, 2009; Roeser et al., 2006; Smith et al., 2016b) and is similar to Barnard et al.'s (2007) *propositional meaning* system, Epstein's (2003) *rational* system, Deutsch and Strack's (2006) *reflective* system, Kahneman & Tversky's (1973) *reasoned* system, Kosslyn & Pomerantz's (1977) *conceptual* system, Bruner's (1964) *symbolic* system, and Schultheiss's (2001) *verbal-symbolic* system.

[10] In contrast to other approaches (e.g., Weissberg et al., 2003), and consistent with the idea that "EF [executive function] characterizes dimensions of cognition and social-emotional functioning that are not captured by previous constructs" (Blair et al., 2005, p. 565), we do not use the term awareness to mean "understanding" or "knowledge about" the self (e.g., self-awareness) or others (e.g., social awareness). According to the multilevel neuroperson model, understanding, or having knowledge about, the self or world refers specifically to, and requires, beliefs about the self and world but does not refer specifically to or require conscious awareness of those beliefs.

[11] We use the term *primary appraisal* where referring to the first 300-400 milliseconds of the brain's processing of incoming sensory stimuli (i.e., before the nature of the stimuli can become objects of conscious awareness), and we use the term *secondary appraisal* where referring to the conscious processing of incoming sensory stimuli (and associated information) that becomes available to conscious awareness approximately 500 milliseconds after the initiating stimuli.

[12] Depending on how they are defined, executive attention skills can be viewed as aspects of temperament (Rothbart, 2007; Thomas & Chess, 1977) and essentially stable or as peaking during infancy (Kochanska, Murray, & Harlan, 2000), early childhood (Rothbart, 2007), late childhood (Rueda, 2013), or early adolescence (Anderson et al., 2001). These different estimates likely reflect the *task impurity* problem (Miyake & Friedman, 2012; Friedman & Miyake, 2017; Steele et al., 2012), or the difficulty of measuring executive attention skills without confounding them with, for example, representational complexity and working memory capacity (Derryberry & Rothbart, 1997). In other words, measurement demands vary widely across specific aspects of the awareness system (Federico et al., 2017; Mullane et al., 2016; Rueda, 2013), so, as always, users should take care that the measures selected map closely onto the specific skills they intend to measure and promote.

[13] In particular, behavioral measures minimize the potentially anxiety-provoking exposure to the *positivist gaze* (e.g., the subjective experience of being asked questions about feelings by people who are not known and trusted).

[14] In our use, *indicator*, as a representation and reflection of an object, refers to the symbolic category that designates the presence, absence, or some middling condition (i.e., not fully present and not absent) of an object or attribute of an object. In these terms, an indicator is the lowest level of measurement that reflects an object of measurement. All items, scales, and other composites are constructed from indicators of the referenced object.

[15] Behavior provides loosely coupled information about specific mental skills (Bertalanffy, 1968; Cicchetti & Rogosch, 1996). This coupling becomes looser where the setting is poorly designed or the setting press is either unknown or ignored by the participants.

[16] The distinction between functional and optimal skills is analogous to the distinction between "typical" and "maximal" behavioral performance (Cronbach, 1949; Olderbak & Wilhelm, 2020). In addition, Vygotsky (1978) referred to the "distance" between functional and optimal skills as the *zone of proximal development* (ZPD).