

## **Preliminary Recommendations to CORE On Measuring Students' Social-Emotional Skills**

### **Introduction**

On August 6<sup>th</sup>, 2013, the US Department of Education (USED) approved an ESEA waiver application from the California Office to Reform Education (CORE), authorizing the eight participating districts to roll out a new accountability and continuous improvement system called the School Quality Improvement System. CORE's revolutionary new approach to accountability breaks away from the trend of a singular focus on standardized test scores, focusing instead on a multifaceted vision for student success and a shared responsibility to prepare all students for college, career, and life.

The new School Quality Improvement Index (SQII), which will impact one million students, assesses school performance as a function of academic outcomes (60%), school climate and culture measures (20%), and social-emotional measures (20%). The social-emotional domain of the Index will include both administrative data, such as attendance and suspensions, and measures of students' social-emotional skills. This memo describes Transforming Education's initial recommendations regarding which measures of students' social-emotional skills should be considered for inclusion in CORE's SQII.<sup>1</sup>

These recommendations represent a set of preliminary hypotheses informed by the work of our Scientific Advisory Board and by leading SEL researchers from CASEL, ETS, and other national organizations. We hope this document will foster initial discussion and debate amongst CORE and district leaders. We welcome your feedback and hope to have the opportunity to engage collectively in a deeper Design Phase for this work, during which we would gather context-specific input from CORE practitioners, vet our initial recommendations with the teachers and leaders who will use these measures on the ground, and work with leading psychometricians to adapt the most promising social-emotional measures to the CORE context.

---

<sup>1</sup> Transforming Education (TransformEd) is a non-profit organization dedicated to helping educators and school systems equip students with the mindsets, skills and habits they need to succeed in college, career, and life. TransformEd arose from a set of research collaborations, funded by the Bill and Melinda Gates Foundation, between the National Center on Time & Learning (NCTL) and MIT Professor of Brain & Cognitive Science John Gabrieli. TransformEd's founding network includes content experts in the psychology and cognitive science fields, innovative education reform practitioners, and entrepreneurial policymakers with experience successfully building an education reform movement into a robust field. Our Scientific Advisory Board is led by cognitive neuroscientist John Gabrieli (MIT) and includes the following leaders of the psychology, education, and education policy fields: Angela Duckworth (UPenn), Carol Dweck (Stanford), Clancy Blair (NYU), David Yeager (UT Austin), Marty West (Harvard), Daniel Willingham (UVA), and Matt Kraft (Brown). TransformEd operates as an independent program of the National Center on Time & Learning.

## A Few Design Principles and Considerations

In formulating our recommendations, Transforming Education has adhered to a few key principles that we believe will support effective design and implementation of the School Quality Improvement Index. This section describes each of those key principles in greater detail.

### *Traits Included in the Index Must be Meaningful, Measurable, and Malleable*

Transforming Education (TransformEd) refers to “social and emotional skills” or “non-cognitive skills” as Mindsets and Essential Skills & Habits (MESH). We believe that any MESH trait that is to be incorporated into education policy and practice should ideally pass three tests we call the “Three M’s” – to wit they must be meaningful, measurable and malleable.

To be considered meaningful, a trait must matter for students’ success in college, career and life. More specifically, a “meaningful” MESH trait must have significant evidence of predicting students’ academic and life success over the middle to long term, above and beyond test scores and conventional academic measures. For example, there is strong evidence demonstrating the importance of self-control on life outcomes: Moffitt et al. have demonstrated that childhood self-control, independent of intelligence and socio-economic status, strongly predict such long-term outcomes as health, wealth, well-being and criminal activity in adulthood.<sup>2</sup> There is also compelling middle-term (i.e. more than a school year) evidence demonstrating the impact of growth mindset and social belonging on academic and attainment outcomes.<sup>3</sup>

By measurable, we mean that there must be practical measures available for a trait that meet reasonable tests of validity and reliability. This is a growing concern for the MESH field given the multiple sources of evidence for “anomalous” or “paradoxical” results that invalidate cross-school or cross-cultural comparisons. (Please see the *Reference Bias* section below for more details.) These issues may not loom as large when measures are used in a single school or are used only for formative purposes, but it is critical to address such biases when implementing MESH measures on a large scale or for accountability purposes, as CORE has proposed to do.

By malleable, we mean that there should be reasonable evidence that the MESH trait in question can be strengthened within a school setting. MESH traits are only relevant to CORE’s SQII insofar as they can be changed in the context of individual schools. Currently, there is a growing research base that demonstrates that several MESH traits are, in fact, malleable through school-based interventions: for example, Carol Dweck and her colleagues have demonstrated that both online program and in-person interventions can boost students’ growth mindset (i.e. their belief that academic ability increases with effort). More broadly, psychologists have found that certain MESH traits may be malleable later into life than

---

<sup>2</sup> Moffitt et al. (2011) A gradient of self-control predicts health, wealth, and public safety; Also see Mischel, Shoda & Rodriguez (1989) Delay of Gratification in Children and Segal (2012) Misbehavior, education, and labor market outcomes

<sup>3</sup> Examples include Blackwell, Trzesniewski, & Dweck (2007) Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention

fundamental cognitive skills, which provides a promising rationale for schools to focus on building these traits.<sup>4</sup>

### *The Measures Must be Parsimonious and Practical*

Given CORE's intent to begin piloting its new School Quality Improvement Index (SQII) this year, the MESH measures selected must be ready for immediate use and must be practical to administer to large groups of students. We assume that CORE will be able to combine the administration of student self-report measures with other existing instruments, such as standardized academic tests and school climate surveys. We anticipate that these measures will be collected in machine-scorable ways or, in some cases, online.

Furthermore, we value parsimony of measures because we know that instructional time is limited and that students, teachers, and leaders experience significant testing fatigue when faced with too many assessments. The MESH field has many different, overlapping measures arising from multiple frameworks (e.g. social-emotional learning, personality psychology, 21<sup>st</sup> century skills, character education, etc.) and many different researchers. CORE must pick a small number of the best validated and most practical of these in order to efficiently gather useful data on students' MESH traits.

While we value parsimony, we also understand the importance of collecting and combining multiple measures of a given trait to increase the validity of the resulting data. We believe that in addition to student self-report, it will also be important to collect teacher reports on some key measures, such as self-regulation. The process of combining teacher reports with self-reports for certain traits (e.g. self-control/self-regulation) has been shown to significantly increase the validity of the resulting measure. We understand that gathering teacher ratings represents another layer of complexity in the implementation process, but the incremental value of doing so makes this effort worthwhile. Similarly, while performance tasks and situational judgment tests may not yet be suitable for full-scale use, their advantages in objectivity should compel immediate piloting where they are ready (e.g. for self-regulation in younger children).

### *The Set of Measures Should Have Face Validity and Cover a Range of Important Traits*

Beyond selecting the few most validated and practical measures, we believe that CORE must also supplement its set of MESH measures to cover a reasonably complete range of characteristics that educators believe are important for student success. This means, to begin with, covering both the intrapersonal and interpersonal domains. Also, it likely means covering concepts that are familiar to educators interested in the MESH field, such as self-regulation/self-control, grit, growth mindsets, and teamwork. CORE's own team of educators possesses a wealth of knowledge on which skills and traits are important for student success. Our goal is to create an initial recommendation that will resonate with educators and then to

---

<sup>4</sup> Almlund, Duckworth, Heckman & Kautz (2011) *Personal Psychology and Economics*

get broad input from CORE's own teachers and school leaders during a Design Phase that will precede the actual piloting of measures during the 2013-14 school year.

### *We Must Acknowledge and Address Reference Bias in Selecting Measures for the SQII*

It has become increasingly clear that many MESH measures based on student self-report are susceptible to various biases, including a reference bias that renders them of little or no use in comparing heterogeneous student populations across schools. It is crucial that CORE understand this problem and take steps to reduce or eliminate the distortion of data that reference bias causes, as this issue is of the utmost concern in accountability settings.

To illustrate one example of reference bias, the 2003 PISA test showed evidence of paradoxical relationships between students' math achievement and their math self-concept (measured by students' agreement with such statements as "I get good grades in math" and "I learn mathematics quickly"). Within individual countries, these two variables were positively correlated ( $r = .40$ ). However, across countries, the same variables were negatively correlated ( $r = -.20$ ). In other words, within a given country, students who performed better on the PISA math test were more likely to have a positive math self-concept; however, with respect to country-level averages, higher-achieving countries tended to have lower self-concepts.<sup>5</sup>

Our own research and many other examples from the literature show similar paradoxes when comparing countries, schools, or student subgroups.<sup>6</sup> Researchers' prevailing hypothesis is that the anomalous results are caused by differences in the frame of reference that students use in responding to questions that require them to compare their own performance to a particular standard of excellence. For example, to answer the sample math self-concept questions described above, a student must conjure up a mental representation of what it means to "learn math quickly" and then compare him or herself to that mental representation. The representation itself can vary by ethnicity, school, country, or other frame of reference. As such, students' responses to such questions may be systematically biased in a way that invalidates comparisons across schools and subgroups. Put simply, when individuals are asked to rate themselves, they use the people around them as a frame of reference. If people in one sample have much higher standards for hard work than people in another sample do, individuals from each group who work equally hard may rate themselves differently in response to the question "Are you a hard worker?"

When considering self-reported measures for use in an accountability system, reference bias represents a serious concern. In constructing the SQII, CORE must acknowledge and address this concern to ensure that the Index itself is a valid and reliable measure of school performance. There are at least four ways that reference bias can be addressed. Firstly, we believe that placing a premium on measures that do not appear to suffer from such reference bias can mitigate the issue. For example, student self-reports on internal beliefs (e.g. student

---

<sup>5</sup> Kyllonen & Bertling (in press) Innovative questionnaire assessment methods to increase cross-country comparability

<sup>6</sup> Heine, Buchtel & Norenzayan (2007) What do cross-national comparisons of personality traits tell us?; Tuttle et al. (2013) KIPP Middle Schools: Impact on Achievement and Other Outcomes; Dobbie & Fryer (unpublished data)

mindsets, described below) seem to be less prone to reference bias because such questions ask students to describe their inner thoughts rather than compare their behavior to a particular standard of excellence.

Secondly, we believe that multiple reporters can help offset individual reference bias, increasing the validity of the resulting composite measures. For that reason, we think it is especially important that self-regulation/self-control be measured by both teacher report and student self-report. Teacher reports may also be susceptible to reference bias, but multiple reports should mitigate this risk. For example, Moffitt et al. created a composite measure of childhood self-control using reports from multiple stakeholders and subsequently found this measure to be significantly predictive of adult outcomes. In addition to using student and teacher reports, CORE may also be able to collect peer reports of MESH traits, not just individually, but at the class or grade level.

Thirdly, we should begin to pilot the most promising assessment techniques developed to offset reference bias. For example, researchers from ETS have been able to reverse the paradoxical trends of the 2003 PISA results by using anchoring vignettes and forced-choice surveys.<sup>7</sup> Through the anchoring vignettes technique, students are asked to rate several hypothetical individuals or situations on the scale of interest before rating themselves. Each student's self-rating is then adjusted based on his/her rating of the three hypothetical scenarios. In forced-choice questions, students are asked to rate or select from a limited number of options, each referring to different traits. For example, students may be asked to choose which of the following statements is more true of them: "I work well in groups" or "I am a hard worker." The statements intentionally describe two unrelated and equally desirable traits (e.g. teamwork and self-control) so that there is no obvious "correct" answer. Each student responds to a series of similar questions, and his or responses are then converted into a rating on each of the scales used (e.g. teamwork and self-control). Both of these techniques have been shown to increase the validity and reliability of social-emotional measures that are otherwise subject to reference bias.

Finally, we should explore the use of task-based performance measures, which may be fundamentally more objective than self-reported measures. Task-based measures may also be more useful in the youngest respondents, where self-reported measures are not practical. For example, the Flanker Task is recommended by the National Institute of Health to test attention and inhibitory control (forms of self-control) in individuals from age 3-85. The test requires the participant to focus on a particular stimulus while inhibiting attention to the other stimuli flanking it. The test takes approximately 3 minutes to complete, and participants are assessed based on a combination of accuracy and reaction time. Because such task-based measures may be significantly more objective than self-report and teacher reports, it will be crucial to begin piloting these approaches as soon as possible.<sup>8</sup>

---

<sup>7</sup> Kyllonen & Bertling (in press)

<sup>8</sup> See NIH website: <http://www.nihtoolbox.org/WhatAndWhy/Cognition/Attention/>

## The Perils of Using MESH for Accountability

We are delighted that CORE is pioneering the adoption of MESH measures as peers to academic measures in setting high expectations for schools. We believe that MESH skills are as important to students' life success as academic skills are, and therefore that helping students' build MESH skills should be a high priority for schools. That said, we also are quite concerned about the perils of using these measures for accountability purposes. At a practical level, MESH measures – including both the ones discussed in this memo and the additional measures of attendance and suspension – seem especially susceptible to inflation once they carry high stakes for accountability. Suspensions, for example, are issued at the discretion of the principal, and student attendance is a measure that could easily be tampered with.

With respect to self-report measures, schools could coach teachers and students to parrot back shallow, positive answers to survey questions, driving up results with little or no effect on the actual student traits underlying the measures. When teachers fill out similar questionnaires, which require subjective judgment of students' capacity for self-regulation, one could imagine that high stakes situations may lead to grade inflation. Such gaming of the measures would be counterproductive to the ultimate goal of raising students' MESH traits to improve their chances for success in college, career, and life.

Given these concerns, we are heartened by the fact that SEL measures account for a relatively small part of the total SQII score and by the fact that CORE has developed an accountability system in which a struggling school receives more support, rather than severe punishments. Further, the CORE philosophy is aimed at prompting intrinsic motivation for change in schools and districts and providing many avenues for continuous improvement. **Nonetheless, we urge CORE to be cognizant of these concerns and to consider ways to establish and maintain a high priority on MESH measures while minimizing the risk that educators will engage in unproductive gamesmanship to raise students' MESH scores without truly investing in these crucial assets.**

## Preliminary Recommendations

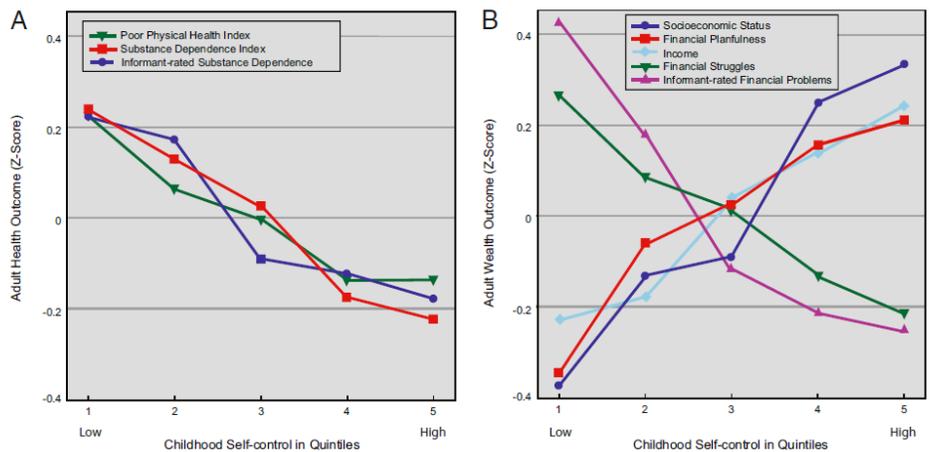
We recommend that CORE gather data on the following key intrapersonal and interpersonal traits to contribute to the social-emotional domain of the SQII. The outline below identifies specific mindsets, skills and habits to measure and summarizes what types of measures are likely to be most appropriate (e.g. self-report, teacher report, task-based measures). Specific scales and measures will need to be developed in a subsequent Design Phase to be carried out before the 2013-14 pilot begins.

### Intrapersonal Mindsets, Skills and Habits

- 1. Intrapersonal Self-Control:** Of the various human traits that psychologists and economists study, self-control is the trait most reliably related to school success.<sup>9</sup> This trait actually bridges the *intrapersonal* and *interpersonal* domains, as it refers to self-regulation of attention (*intrapersonal*) as well as self-regulation of social behavior (*interpersonal*). There is clear evidence that self-regulation of attention and of interpersonal behavior, even when measured at a young age, positively predict high school and college completion.<sup>10</sup> One notable recent report demonstrates that, controlling for school achievement, children who were rated one standard deviation above the mean on attention span (a measure of intrapersonal self-control) at age 4 had 39% greater odds of completing college by age 25.<sup>11</sup>

A major study published in 2011 by Duke psychologist Terrie Moffitt summarizes some of the most compelling evidence that self-control is linked to key life outcomes beyond school and well into adulthood. Following a cohort of 1,000 children from birth to age 32, Moffitt and her colleagues demonstrate that childhood self-control (measured from ages 3-11) predicts physical health, substance dependence, personal finances, and criminal offending outcomes in adulthood. (See Figure 1) The researchers controlled for students' intelligence and socio-economic status, isolating self-control as a key predictor of these adult outcomes.<sup>12</sup>

**Figure 1. Moffitt et al. show that self-control predicts health and wealth outcomes**



<sup>9</sup> Duckworth & Carlson (in press) Self-regulation and school success. In Sokol, Grouzet & Müller (Eds.) *Self-regulation and autonomy: Social and developmental dimensions of human conduct*

<sup>10</sup> Duckworth & Carlson (in press)

<sup>11</sup> McClelland, Piccinin, Acock & Stallings (2011) Relations between preschool attention and later school achievement and educational outcomes

<sup>12</sup> Moffitt et al. (2011); Knudsen, Heckman, Cameron & Shonkoff (2006) Economic, neurobiological, and behavioral perspectives on building America's future workforce

**Relevant measures:** TransformEd recommends measuring self-control in all students, using multiple measures to create a more valid and reliable composite rating. For elementary students, CORE could focus primarily on teacher ratings, gathering ratings from multiple teachers on each individual student where possible. For middle and high school students, CORE could combine multiple teacher ratings with student self-report measures. (See “Interpersonal Self-Control” below for sample survey questions.) While there is strong evidence that self-reported measures of self-control are subject to a reference bias, ETS has demonstrated that the effect of this bias can be minimized by combining self-reports with other measures and by using anchoring vignettes or forced-choice questions. Finally, we recommend that CORE explore, and ideally begin to pilot, performance tasks to assess self-control as well.<sup>13</sup>

- 2. Grit:** Grit is a trait closely related to self-control that Angela Duckworth and her colleagues define as “perseverance and passion for long-term goals.”<sup>14</sup> While self-control refers to the ability to control impulses in the short-term, grit refers to sustained effort and engagement in the pursuit of longer-term goals.<sup>15</sup> Duckworth and colleagues have demonstrated that, while controlling for intelligence, grit predicts educational attainment, adolescents’ and college students’ GPA as well as retention in demanding learning environments.<sup>16</sup>

**Relevant Measures:** TransformEd recommends measuring grit in high school and potentially middle school students through a combination of self-report and teacher-report instruments. Grit measures ask individuals to rate their level of agreement with statements such as “I have achieved a goal that took years of work” (high grit) and “I become interested in new pursuits every few months” (low grit). These measures are less relevant to elementary school students, as young children have typically had fewer opportunities to demonstrate perseverance and passion for long-term goals.

- 3. Student Mindsets:** “Student Mindsets” refer to students’ perceptions of themselves and their learning environments, which can affect how they perform academically and whether they exert or withdraw effort in the face of adversity. Past research has shown that addressing the following mindsets can improve students’ performance in enduring ways.<sup>17</sup>

---

<sup>13</sup> Sample performance tasks for self-control (inhibitory control and attention) are available on the NIH website: <http://www.nihtoolbox.org/WhatAndWhy/Cognition/Attention/>

<sup>14</sup> Duckworth et al. (2007) Grit: Perseverance and passion for long-term goals

<sup>15</sup> Dweck, Walton & Cohen (2011) Academic Tenacity: Mindsets and Skills that Promote Long-Term Learning

<sup>16</sup> Duckworth et al. (2007), Duckworth & Quinn (2009) Development and validation of the short grit scale

<sup>17</sup> Yeager, Paunesku, Walton & Dweck (2013) How Can We Instill Productive Mindsets at Scale? -- A Review of the Evidence and an Initial R&D Agenda; Garcia & Cohen (2012) A social-psychological approach to educational intervention; Yeager & Walton (2011) Social-psychological interventions in education: They're not magic

- a. **Growth mindset:** Growth mindset describes the belief that one’s academic ability can improve with effort, rather than being fixed at a particular level that is outside of one’s control. Multiple longitudinal studies show that growth mindset predicts students’ performance in school.<sup>18</sup> Research suggests that having a growth mindset is particularly important during transitions (e.g. from elementary to middle or middle to high school), since students often meet with new challenges during such transitions.<sup>19</sup>
- b. **Self-efficacy:** Decades of research show that self-efficacy, or students’ belief in their ability to perform a specific task or attain a certain goal, is a valid predictor of students’ motivation and learning.<sup>20</sup> There is evidence that students with high levels of self-efficacy participate more in class, work harder, persist longer, and have fewer adverse emotional reactions when encountering difficulties than their peers with lower self-efficacy.<sup>21</sup> High self-efficacy can also motivate students to use specific learning strategies and to engage in self-directed learning.<sup>22</sup>
- c. **Social belonging:** Social belonging describes the belief that one *belongs* and will be valued in a particular setting, such as school. Longitudinal research links social belonging in a school setting to long-term student motivation and academic success.<sup>23</sup> Students who have better relationships with teachers and peers experience a greater sense of social belonging. As a result, these students are more engaged in class and earn better grades, even when controlling for prior performance and motivation.<sup>24</sup>
- d. **Relevance/Motivation:** This fourth mindset refers to a student’s sense that the subject matter he or she is studying is interesting, relevant, or valuable. Value can be derived from the importance of doing well on a task (attainment value), from enjoyment of the task itself (intrinsic value), or from seeing the task as useful in fulfilling certain goals (utility value).<sup>25</sup> The degree to which a student values a particular academic task (i.e. “task value”) is a strong determinant of the student’s motivation to complete that task, influencing the student’s persistence

---

<sup>18</sup>Blackwell, Trzesniewski, & Dweck (2007) Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention

<sup>19</sup>Blackwell, Trzesniewski, & Dweck (2007); Dweck, Walton & Cohen (2011) Academic Tenacity: Mindsets and Skills that Promote Long-Term Learning

<sup>20</sup>Bandura (1997) Self-efficacy: The exercise of control; Zimmerman (2000) Self-Efficacy: An Essential Motive to Learn

<sup>21</sup>Bandura (1997); Zimmerman (2000)

<sup>22</sup>Zimmerman & Martinez-Pons (1990) Student differences in self-regulated learning; Zimmerman, Bandura, & Martinez-Pons (1992) Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting

<sup>23</sup>Dweck, Walton & Cohen (2011); Yeager & Walton (2011)

<sup>24</sup>Dweck, Walton & Cohen (2011); Furrer & Skinner (2003) Sense of relatedness as a factor in children’s academic engagement and performance; Roeser, Midgley & Urdan (1996) Perceptions of the school psychological environment and early adolescents psychological and behavioral functioning in school: The mediating role of goals and belonging

<sup>25</sup>Eccles et al. (1983) Expectancies, values, and academic behaviors; Wigfield & Eccles (Eds.) (2002) Development of achievement motivation

and performance.<sup>26</sup> Task value also predicts course choice and career choice in students and adults.<sup>27</sup>

**Relevant measures:** We recommend that these mindsets be measured in students of all ages, especially around common transition points (e.g. elementary to middle school, middle to high school, high school to college), as evidence suggests that the mindsets are particularly important mediators of successful transitions. Because mindsets are internal beliefs and perceptions, they can only be measured through self-report. Various measures of student mindsets have been shown to be more valid and reliable when they are administered in reference to a specific subject, task or skill. Thus, the student questionnaires used will likely need to refer to a specific academic subject, such as math or English language arts.

### *Interpersonal Mindsets, Skills and Habits*

- 4. Interpersonal Self-Control:** As described above, self-control is both an *intrapersonal* and an *interpersonal* trait. Measures can combine these two domains of self-control or can address each one separately. While intrapersonal self-control measures for students typically focus on habits like coming to class prepared, paying attention, and resisting distractions; interpersonal self-control measures focus on students' ability to remain calm when criticized and allow others to speak without interruption.<sup>28</sup>

**Relevant Measures:** As with *intrapersonal* self-control, we recommend measuring *interpersonal* self-control in all students and combining student reports and teacher reports to create a more valid and reliable composite rating. We should also explore the possibility of piloting performance tasks to assess interpersonal self-control.

- 5. Collaborative Problem Solving:** Collaborative problem solving is widely established as an important factor in workforce success. One recent employer survey conducted by the Partnership for 21<sup>st</sup> Century Skills and its supporting organizations found that collaboration was amongst the two most important applied skills for job entrants of all education levels.<sup>29</sup> The OECD has stated that its 2015 PISA exam will measure collaborative problem solving competency, which it defines as an individual's ability to engage effectively in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution.<sup>30</sup>

---

<sup>26</sup> Farrington et al. (2012) Teaching Adolescents to Become Learners – The Role of Noncognitive Factors in Sharing School Performance: A Critical Literature Review; Eccles et al. (1983); Wigfield (1994) Expectancy-value theory of achievement motivation: A developmental perspective

<sup>27</sup> Eccles (1994). Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices

<sup>28</sup> KIPP Character Growth Card

<sup>29</sup> Casner-Lotto, Barrington & Wright (2006) Are they really ready to work?

<sup>30</sup> OECD (2013) PISA 2015 Draft Collaborative Problem Solving Framework

**Relevant Measures:** Many public and private institutions (e.g. National Research Council, ETS, OECD, etc.) are in the process of developing scalable measures of collaborative problem solving, and the 2015 PISA measure is expected to be completed by December 2013. Existing approaches that may be appropriate for CORE’s purposes include situational judgment tasks (SJT), in which students select from a set of potential responses to a situation that is presented in writing or through video. In Belgium, SJTs assessing interpersonal skills are being used as one component of the medical school admissions process. These assessments have been shown to be predictive of students’ internship and job performance over approximately ten years.<sup>31</sup>

In addition to confirming that feasible measures of collaborative problem-solving will exist in time for CORE’s 2013-14 pilot, we recommend exploring one or two more interpersonal measures (e.g. communication skills, social efficacy, social intelligence, etc.) that could be included in the SQII over the course of the design phase of this work.

## Selecting Traits for Inclusion in the SQII is Just the Beginning

Reaching preliminary consensus on which MESH traits to focus on is just one step in a much larger process that will be required to ensure the success of CORE’s new SQII. This section outlines two phases of work that we believe will be integral to the successful design and implementation of the SQII’s social-emotional domain.

### *Design Phase*

While the recommendations above are based on the most promising scientific research in the MESH field, the Design Phase will focus on refining these recommendations and rendering them practical. The goals of the Design Phase include the following:

1. Gather context-specific input from CORE practitioners about the substance and form of social-emotional measures to be included in the SQII
2. Ensure that social-emotional measures are thoughtfully integrated into other data collection efforts already underway in schools (e.g. academic testing, other surveys, etc.)
3. Use this input to partner with leading psychometricians in adapting the most promising measures to the CORE context
4. Design the 2013-14 pilot to reflect the notion that the social-emotional measures and the data collection effort itself will require continuous improvement.

During the design phase, we would focus on sharing our initial hypotheses with a diverse range of practitioners, gathering their input, and reaching a firm conclusion about which traits can and should be measured in the social-emotional component of the SQII. This phase of work is

---

<sup>31</sup> Lievens & Sackett (2012) The Validity of Interpersonal Skills Assessment Via Situational Judgment Tests for Predicting Academic Success and Job Performance

not only crucial for helping CORE make more informed decisions, it also serves to lay the groundwork for broad-based buy-in from educators, which is a key part of the intrinsic motivation and continuous improvement approach that CORE aims to foster.

During this phase of work, we would also gather information about what surveys and measures are already being administered in schools that may overlap with those selected for inclusion in the Index. Collecting this information will enable CORE to reduce the testing burden on students and teachers wherever possible. It will also help identify which vehicles for data collection already exist in schools so that new measures could be integrated into these channels rather than designing separate data collection processes for social-emotional measures.

Using our initial hypotheses and the valuable feedback collected from CORE practitioners, we would partner with leading psychometricians to identify specific scales and surveys that are ready for use in the 2013-14 pilot and to adapt these measures as needed to mitigate the relevant biases (e.g. by adding anchoring vignettes or converting to forced-choice scales). Potential partner organizations for this effort include ETS and ACT, both of which have already invested significantly in developing social and emotional measures, some of which are already in use in postsecondary education and employment testing settings.

Once the specific measures that will be used in the 2013-14 pilot are identified, we would need to map out the remaining details of the pilot itself: how many schools will participate, which students and grades will be tested, how and when will tests be administered, and how will results be evaluated? We would also proactively identify which analyses need to be conducted with the pilot data to test the feasibility, reliability, and validity of the selected social-emotional measures. For example, intrapersonal self-control measures should correlate with administrative measures such as attendance, tardiness, and homework completion. Accordingly, CORE may want to capture this administrative data systematically during the pilot in order to validate the selected social-emotional measures at the end of the 2013-14 school year.

### *Outreach and Training Phase*

While a diverse subset of CORE educators will participate in the Design Phase, it will be crucial to communicate with *all* educators about the final design and rollout of the pilot measures during a subsequent Outreach and Training Phase. The goals of the Outreach and Training Phase include the following:

1. In concert with CORE's broader outreach and training efforts, share information about the SQII social-emotional domain with all teachers
2. Prepare teachers for the specific role they will be asked to play in assessing students' social-emotional skills during the 2013-14 pilot
3. Foster rich conversations about the results of the pilot in order to refine selected measures before broader rollout in 2014-15
4. Begin to identify promising practices that teachers may already be using to help build students social and emotional skills.

Specific outreach and training for the social-emotional domain must, of course, be tightly integrated with the communications and training efforts CORE designs to support the overall School Quality Improvement Index. If done well, these outreach and training efforts will help establish a common understanding of, language for, and commitment to CORE's holistic definition of college and career readiness.

Clear, consistent, and responsive communication with educators will be a crucial part of successful SQII implementation. In addition to learning more about the goals and plan for the SQII overall, teachers will need to receive training on how to use the teacher-report scales and how to administer student-report measures for the social-emotional domain. To provide training on a large scale while minimizing costs, CORE may be able to leverage MOOC-like technologies, creating an online course that explains the rationale for including social and emotional factors in the SQII, describes how these factors will be measured, and prepares teachers for the specific role they will be asked to play.

Once the pilot has been completed, the resulting data can be used to inform rich discussions with teachers and school leaders about what we are learning. For example, were there any notable barriers to successful data collection during the pilot? Why might multiple teachers have scored an individual student differently on a particular trait? Are there observable patterns in the differences between students' self-ratings and teacher ratings? By discussing these and related questions with educators, CORE will gather valuable feedback to inform the revision of measures before the broader rollout scheduled for 2014-15. These discussions also represent an opportunity to strengthen educators' level of comfort with and buy-in for the SQII overall.

Finally, data from the broader 2014-15 rollout of social-emotional measures can be used to highlight the instructional techniques and interventions that are most effective in helping build students' MESH skills. Using school- and classroom-level MESH results, CORE will be able to identify schools and teachers whose students outperform expectations and then study and scale their best practices. CORE will also be able to use these social-emotional measures to evaluate formal SEL programs and interventions that schools and districts are already using. The focus on MESH measures should ignite a high quality search for the most promising programs and approaches that schools can use to address students' most pressing MESH needs.

All of this can and should be done in the spirit of continuous improvement that CORE emphasizes in its waiver application. If CORE's new School Quality Improvement Index is planned and implemented effectively, it will have the power to dramatically change the understanding, awareness, and effort that schools and teachers place on helping students build the skills they need to succeed in college, career, and life.