

# Should non-cognitive skills be included in school accountability systems? Preliminary evidence from California's CORE districts

Martin R. West

## Executive Summary

Evidence confirms that student skills other than academic achievement and ability predict a broad range of academic and life outcomes. This evidence, along with a new federal requirement that state accountability systems include an indicator of school quality or student success not based on test scores, has sparked interest in incorporating such “non-cognitive” or “social-emotional” skills into school accountability systems.

Yet important questions have been raised about the suitability of extant measures of non-cognitive skills, most of which rely on asking students to assess their own abilities, for accountability purposes. Key concerns include the possibility of misleading information due to reference bias in students' self-reports and that students may simply inflate their self-ratings to improve their school's standing once stakes have been attached.

The most ambitious effort to deploy common measures of non-cognitive skills as part of a performance management system is unfolding in California's CORE Districts, a consortium of nine school districts that collectively serve over one million students. In the 2014-15 school year, CORE conducted a field test of measures of four social-emotional skills involving more than 450,000 students in grades 3-12. Starting this year, information from these measures will be publicly reported and is expected to play a modest role in schools' performance ratings, comprising eight percent of overall scores.

Analysis of data from the CORE field test indicates that the scales used to measure student skills demonstrate strong reliability and are positively correlated with key indicators of academic performance and behavior, both across and within schools. These findings provide a broadly encouraging view of the potential for self-reports of social-emotional skills as an input into its system for evaluating school performance. However, they do not address how self-report measures of social-emotional skills would perform in a high-stakes setting – or even with the modest weight that will be attached to them within CORE. The data currently being gathered by CORE provide a unique opportunity for researchers to study this question and others related to the role of schools in developing student skills and the design of educational accountability systems.

A growing body of evidence confirms that student skills not directly captured by tests of academic achievement and ability predict a broad range of academic and life outcomes, even when taking into account differences in cognitive skills.<sup>i</sup> Both intra-personal skills (such as the ability to regulate one’s behavior and persevere toward goals) and inter-personal skills (such as the ability to collaborate with others) are key complements to academic achievement in determining students’ success. This evidence, in combination with a new federal requirement that state accountability systems include an additional indicator of school quality or student success not based on test scores, has sparked [widespread interest](#) in the possibility of incorporating such “non-cognitive” or “social-emotional” skills into school accountability systems.

At the same time, important questions have been raised about the suitability of extant measures of non-cognitive skills, most of which rely on asking students to assess their own abilities, for accountability purposes. In a 2015 paper in *Educational Researcher*, leading psychologists Angela Duckworth and David Yeager offer what they describe as a “simple scientific recommendation regarding the use of currently available personal quality measures for most forms of accountability: not yet.”<sup>iii</sup>

Duckworth and Yeager identify three key concerns with the use of student self-reports of non-cognitive skills into accountability systems. The first stems from the fact that students evaluating their own skills must employ an external frame of reference in order to reach a judgment about their relative standing. As a result, differences in self-reports may reflect variation in normative expectations rather than true differences in skills, a phenomenon known as “reference bias.”<sup>iii</sup> To the extent that students attending schools with more demanding expectations for student behavior hold themselves to a higher standard when completing questionnaires, reference bias could make comparisons of their responses across schools misleading. If schools with high expectations are actually more effective in improving students’ non-cognitive skills (something not yet known but often assumed), conclusions about school performance based on self-reports could even be precisely backward.

Duckworth and Yeager’s second concern is more obvious: that students may simply inflate – or be coached to inflate – their self-ratings to improve their school’s standing once stakes have been attached. Finally, they note that we have little evidence on the

ability of these measures when aggregated to the school level to distinguish statistically between schools with high and low levels of performance – something that depends on both the reliability of the measures and the extent to which students in the same school tend to respond in similar ways.

These concerns are worth taking seriously, especially when voiced by scholars who have done so much to enrich our understanding of the skills students need to succeed in the classroom and beyond. My own research has suggested the potential importance of reference bias due to differences in school climate, leading me to [caution](#) in this series against proposals to incorporate survey-based measures of non-cognitive skills into high-stakes accountability systems.<sup>iv</sup>

In addition to the concerns emphasized by Duckworth and Yeager, I would note the risk that deploying superficial measures of non-cognitive skills might lead to superficial instructional responses. Setting aside intentional faking, there’s clearly a difference between thinking of oneself as having strong self-management skills or a high level of social awareness and actually being able to demonstrate those capacities in one’s daily life, including in novel situations and environments.

Yet a few school systems are moving forward with using student self-reports to systematically track the development of non-cognitive skills and even with including them as a component of school accountability systems; others may well follow. This is understandable, given the ways in which the importance of these skills has been promoted. One of Duckworth’s seminal papers on self-control, for example, is entitled “What No Child Left Behind Leaves Behind: The Roles of IQ and Self-Control in Predicting Standardized Achievement Test Scores and Report Card Grades.”<sup>v</sup>

It is also, in my view, a positive development. Above all, it presents an enormous learning opportunity for the field – a chance to study not only the properties of the measures when administered at scale and how, if at all, they change once stakes are attached, but also schools’ role in developing non-cognitive skills and effective strategies to improve them. To the extent that there is skepticism about the value of student self-reports for school accountability, it presents an opportunity to subject that skepticism to an empirical test. Educational accountability systems serve many purposes, one of which is to signal to educators what is important in a way that will lead to desired changes

in instructional practice. Are we really so sure that the inclusion of measures of non-cognitive skills in such a system can't play a constructive role? And might not the use of current measures, despite their potential flaws, help drive the development of new and better ones?

Easily the most ambitious effort to deploy common measures of non-cognitive skills as part of a performance management system is unfolding in California's CORE Districts, a consortium of nine school districts that collectively serve over one million students in more than 1,500 schools.<sup>vi</sup> Six of these districts have been operating since 2013 under a waiver from the U.S. Department of Education to implement an accountability system that aims to be both more holistic and more useful for improving practice than they believe is possible based on test scores alone. In addition to student proficiency and growth as measured by state tests, the inputs into CORE's School Quality Improvement Index (SQII) include such indicators as suspension and expulsion rates, chronic absenteeism, and school culture and climate surveys administered to students, teachers, and parents. The most distinctive feature of the SQII, however, is the plan eventually to incorporate self-report measures of what CORE refers to as students' social-emotional skills directly into school performance ratings.

CORE has approached the development of this component of its accountability framework in a cautious, thoughtful manner. Working with a partner organization known as Transforming Education, they selected the specific social-emotional skills on which to focus based on a review of evidence on the extent to which those skills are measurable, meaningfully predictive of important academic and life outcomes, and likely to be malleable through school-based interventions. This process was constrained by a commitment to limit the total assessment burden on students to less than 20 minutes each spring, and ultimately led them to settle on four skills: self-management, social awareness, self-efficacy, and growth mindset (see Appendix Table 1). After piloting the collection of measures of those skills in a small number of schools during the 2013-14 school year, including conducting multiple experiments to compare the performance of alternative survey items, CORE conducted a broader field test involving more than 450,000 students in grades 3-12 the following spring. Starting with the 2015-16 school year, information from these measures will be publicly reported and is expected to factor into school performance ratings –

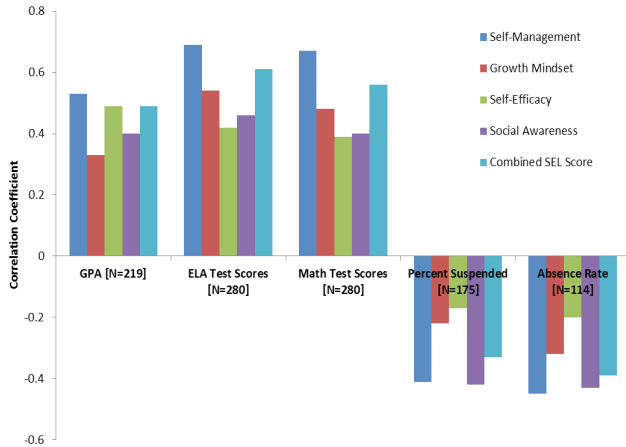
but in a very modest way, comprising just eight percent of the scores schools receive on the SQII. Perhaps most important, CORE has made both the student survey data and district administrative data available to independent researchers at the John W. Gardner Center and Policy Analysis for California Education at Stanford and Harvard's Center for Education Policy Research (CEPR).

My CEPR colleagues and I have used data from the 2014-15 field test to perform preliminary analyses of the reliability of students' survey responses and their validity, when aggregated to the school level, as an indicator of school performance.<sup>vii</sup> With respect to reliability, we first examined the extent to which students' responses to specific items used to measure the same skill were correlated, as would be expected to be the case if they captured a common underlying construct. Across all students in grades 3-12, we found that three of the four of the scales demonstrated strong internal reliability. The exception was the scale used to measure growth mindset, which had an internal reliability coefficient of 0.7, somewhat below the commonly used benchmark for acceptable reliability of 0.8. A closer inspection of the data suggested that the reliability of each scale, and in particular the scale measuring growth mindset, was pulled down by lower inter-item correlations among the youngest students completing the survey – those in the third and fourth grades. This may indicate that students below grade five struggled to understand some survey items or are less well-positioned to assess their own skills, and CORE is currently in the process of deciding which grades it will ultimately include. Overall, however, the scales performed well along this dimension, both overall and for important student subgroups such as English language learners and students with disabilities.

CORE selected its measures of social-emotional learning based on evidence from other settings that they were valid predictors of academic success. Do those same relationships hold when administered at scale in its districts? Figure 2 shows the correlations between school-average social-emotional skills and key indicators of academic performance (GPA and state test scores) and student behavior (the percentage of students receiving suspensions and average absence rates) across CORE district middle schools.<sup>viii</sup> As expected, social-emotional skills are positively related with the academic indicators and negative correlated with the two indicators of student (mis-) behavior, with the correlations for academic indicators ranging from 0.33 to 0.69. The strongest relationships

with academic indicators are observed for self-management, a pattern consistent with other research, while self-management and social awareness are equally important predictors of behavior.

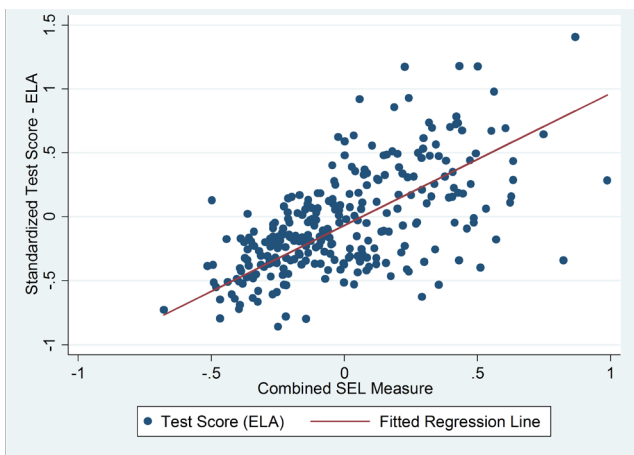
**Figure 1. School-level correlations of average student social-emotional skills and indicators of academic performance and behavior for CORE district middle schools**



Note: All correlations are statistically significant at the 95 percent confidence level or higher. ELA and math test scores are standardized by grade and subject level. GPAs are standardized within district due to variation in scales. Combined SEL Score is an equally weighted average of the four other scales. Schools with fewer than 25 students with valid survey responses excluded.

Figure 2 illustrates the strong correlation between CORE’s summary social-emotional learning measure (the average of the four scales) and English language arts (ELA) achievement, but also reveals ample dispersion of schools around the regression line.<sup>ix</sup>

**Figure 2: School-level relationship between combined social-emotional learning (SEL) measure and English language arts (ELA) test scores for CORE district middle schools**



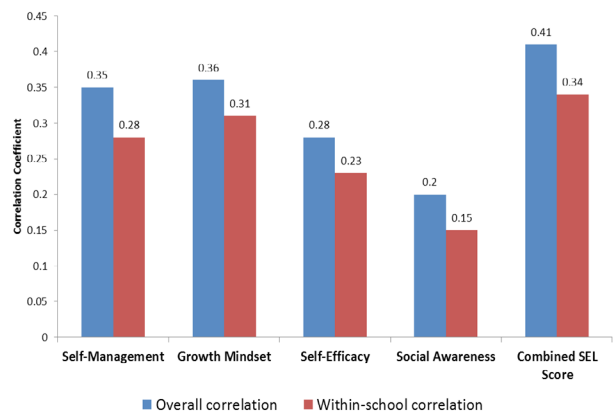
Note: ELA test scores and SEL skills are standardized by grade. Schools with fewer than 25 students with valid survey responses excluded.

In other words, students in some middle schools in which academic performance (as measured by ELA test scores) is high report relatively low social-emotional skills, and vice versa. On one hand, this could reflect authentic variation in performance across academic and social-emotional domains – and therefore the value of a more holistic indicator. On the other, it could be that students in some schools rate their social-emotional skills more critically than in others, perhaps due to variation in norms across schools that leads to reference bias.

To probe for evidence of reference bias, we compared the strength of the student-level correlations between social-emotional skills and academic indicators overall (i.e., across all students attending CORE middle schools) with those obtained when we limit the analysis to comparisons of students attending the same school. The logic of this exercise is straightforward: If students in higher-performing schools rate themselves more critically, then average self-ratings in those schools will be artificially low. This would cause the overall correlation to be biased downward, and lower than that observed among students responding to surveys within the same school environment.

Figure 3 shows the results of this comparison for ELA test scores.<sup>x</sup> It shows that the overall and within-school correlations do differ modestly but that the former are stronger than the latter – precisely the opposite pattern that would result from systematic reference bias due to varying expectations.

**Figure 3. Student-level correlations between social-emotional skills and English language arts (ELA) test scores in CORE district middle schools, overall and within schools**



Note: N=110,293. All correlations and all differences between overall and within-school correlations are statistically significant at the 99 percent confidence level or higher. ELA test scores are standardized by grade and subject level.

To be sure, this analysis does not rule out the possibility that reference bias may lead to misleading inferences about specific schools with particularly distinctive environments. It does, however, provide some preliminary evidence that the form of reference bias that would be most problematic in the context of a school accountability system may not be an important phenomenon in the CORE districts as a whole.

In sum, our preliminary analysis of the data from CORE's field test provides a broadly encouraging view of the potential for self-reports of social-emotional skills as an input into its system for evaluating school performance. That said, the view it provides is also quite limited. It says nothing about how self-report measures of social-emotional skills would perform in a high-stakes setting – or even with the very modest weight that will be attached to them this year within CORE.<sup>xi</sup> Nor can we say anything about how CORE's focus on social-emotional learning will alter teacher practice and, ultimately, student achievement. The results presented above are best thought of as a baseline for future analysis of these issues – and many more.

One reason researchers don't have much to say about these questions currently is that the No Child Left Behind Act effectively required all fifty states to adopt a common approach to the design of school accountability systems. Fifteen years later, we know a lot about the strengths of this approach and even more about its weaknesses – but next to nothing about those of potential alternatives. The recently enacted Every Student Succeeds Act provides both opportunity and incentive for experimentation. What is important is that we learn from what happens next. We need to let evidence speak.

**Appendix Table 1. Social-emotional skills assessed by the CORE Districts**

<b>Skill</b>	<b>Definition</b>	<b>Sample item</b>
Self-management	The ability to regulate one’s emotions, thoughts, and behaviors effectively in different situations. This includes managing stress, delaying gratification, motivating oneself, and selecting and working toward personal and academic goals.	Please answer how often you did the following during the last 30 days. During the past 30 days...  I remained calm even when criticized or otherwise provoked.
Growth mindset	The belief that one’s abilities grow with effort. Students with a growth mindset see effort as necessary for success, embrace challenges, learn from criticism, and persist in the face of setbacks.	Please indicate how true each of the following statements is for you:  My intelligence is something that I can’t change very much.
Self-efficacy	The belief in one’s ability to succeed in achieving an outcome or reaching a goal. Self-efficacy reflects confidence in the ability to exert control over one’s motivation, behavior, and environment.	How confident are you about the following at school?  I can do well on my tests even when they are difficult.
Social awareness	The ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources and supports.	During the past 30 days...  How carefully did you listen to other people’s points of view?

Note: Definitions and items are drawn from CORE Districts documents available at <http://coredistricts.org/school-quality-improvement-system-waiver/>.

- 
- <sup>i</sup> Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. D. (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education*. (vol. 4) (pp. 1-181). Amsterdam: Elsevier, North-Holland; Duckworth, A. L., Tsukayama, E., & May, H. (2010). Establishing causality using hierarchical linear modeling: An illustration predicting achievement from self-control. *Social Psychological and Personality Science* 1(4), 311-317.
- <sup>ii</sup> Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251. In a recent New York Times interview, Duckworth was even more emphatic, telling the reporter: "I do not think we should be doing this; it is a bad idea."
- <sup>iii</sup> Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, 82(6), 903-918.
- <sup>iv</sup> West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli C. F. O., & Gabrieli J. D. E. (2016). Promise and paradox: Measuring non-cognitive traits of students and the impact of schooling. *Educational Evaluation and Policy Analysis*, 38(1), 148-170.
- <sup>v</sup> Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of educational psychology*, 104(2), 439.
- <sup>vi</sup> For more information on the CORE Districts, which include Fresno, Garden Grove, Long Beach, Los Angeles, Oakland, Sacramento, San Francisco, Sanger, and Santa Ana Unified School Districts, see <http://coredistricts.org/why-is-core-needed/core-districts/>.
- <sup>vii</sup> West, M. R., Scherer, E., & Dow, A. (2016). Measuring social-emotional skills at scale: Evidence from California's CORE Districts. Paper presented at the American Education Finance and Policy Annual Conference.
- <sup>viii</sup> We observe broadly similar relationships for elementary and high schools.
- <sup>ix</sup> The fact that the dispersion of school-average test scores in Figure 2 is higher among schools with higher average SEL skills reflects the fact that students in smaller schools tend to report higher SEL skills.
- <sup>x</sup> We obtain a comparable pattern of results for GPA and math test scores.
- <sup>xi</sup> Due to recent changes in federal policy, the identification of schools for comprehensive support and improvement under the Every Student Succeeds Act is likely to be based on performance metrics available statewide, rather than on the SQII.