

Promises and perils of assessing character and social and emotional learning

Author: Clark McKown (ORCID: 0000-0001-9694-1179)

Publication Year: 2017

Journal: Journal of Character Education

Publisher: <https://www.infoagepub.com/about.html>

Grant #: R305A160053

PROMISES AND PERILS OF ASSESSING CHARACTER AND SOCIAL AND EMOTIONAL LEARNING

Clark McKown

Rush University Medical Center

Dr. Card's article provides useful and clear standards for determining what good measurement is, identifies important problems in the measurement of constructs important to character education, and provides clear guidance about measurement development. In this article, I raise 4 points that build on Dr. Card's work. First, advancing assessment in character education requires a vigorous pursuit of conceptual clarity. Second, the field will benefit from efforts specifically to create assessments designed for practice, and those efforts should include consideration of how assessment data are interpreted and used. Third, I highlight the importance in practice of being clear about the purposes for assessing character. And finally, I argue that the method of assessment is a critical but underappreciated consideration, because different methods of assessment are suited to measuring different dimensions of character. The article provides examples from the field of social and emotional learning that are relevant to the adjacent field of character education.

If it is true that “What gets assessed gets addressed,” a natural corollary is this: If we want character and social and emotional learning skills to be addressed in schools and youth development program, we had better assess those skills. Dr. Card's article provides useful and clear standards for determining what good measurement is in terms of reliability, validity, and measurement equivalence. Dr. Card points out important problems in the measurement of constructs such as those reflected in the field of character education—problems such as when

there are a limited number of widely used measures or, conversely, when there are no widely used measures. Dr. Card's article provides clear guidance to the field about the planning, execution, and dissemination of research on measurement. If researchers in the fields of character education and social and emotional learning follow this guidance, we can make great progress in developing rigorous and useful assessments.

In this response, I raise four points that build on Dr. Card's work as it relates to the

• **Correspondence concerning this article should be addressed to:** Clark McKown, Clark_A_McKown@rush.edu

field of character development and social and emotional learning. First, advancing assessment in the field requires a vigorous pursuit of conceptual clarity. Second, the field will benefit from efforts specifically to create assessments designed for practice, and those efforts should include consideration of how assessment data are interpreted and used. Third, I highlight the importance in practice of being clear about the purposes for assessing character and social and emotional learning. And finally, I argue that the method of assessment is a critical but underappreciated consideration, because different methods of assessment are suited to measuring different dimensions of character and social and emotional learning. My own work focuses on children's social and emotional learning, and so the examples I offer are drawn from that body of work. My intention, however, is that all of the points in this response are also relevant to the adjacent field of character education.

CONCEPTUAL CLARITY IN A WORLD OF FUZZY BOUNDARIES

In his article, Dr. Card speaks of “fuzzy boundaries” referring to the often unclear conceptual border between one construct and another. However, the metaphor is relevant to the entire field. A thought experiment will illustrate how: I would wager that if ten scientists and practitioners were asked to define character or social and emotional learning, ten distinct definitions would emerge, with fuzzy boundaries between them. So the broader fuzzy boundary problem is that there is a substantial lack of clarity about what constitutes character or social and emotional learning.

This is consequential. It is the origin of what I see as a kind of measurement paralysis in the field wherein there are not many robust measurement development efforts because funders and scientists are waiting for clarity before committing the considerable resources needed to build sound measurement systems. In addition, fuzzy boundaries levy an implicit

tax on the field. Lack of clarity interferes with communication when we use different terms to mean the same thing, or the same term to mean different things; it impedes the accumulation of scientific knowledge when different researchers define the construct differently; and it undermines practice when different programs with different content and unequal effectiveness are described with the same language. Without clarity, researchers and practitioners can spend a lot of resources purchasing, creating, or adapting measures, but make little progress. Some might argue that in this imperfect world of social science and its fascinating subjects, some fuzziness is inevitable. In general, I would agree. But greater clarity in the field is possible, and indeed essential for its healthy forward momentum.

There are many excellent and useful models for defining social and emotional learning. The Collaborative for Academic Social and Emotional Learning (CASEL) defines social and emotional learning as “the process through which children and adults acquire and effectively apply the knowledge, attitudes, and skills necessary to understand and manage emotions, set and achieve positive goals, establish and maintain positive relationships, and make responsible decisions” (CASEL, 2017). Another model by Stephanie Jones and Suzanne Bouffard identifies critical cognitive, emotional, and social and interpersonal skills, along with contexts that influence the development of those skills (Jones & Bouffard, 2012). Another model emphasizes cognitive, intrapersonal, and interpersonal skills (National Research Council, 2012).

This is by no means a comprehensive listing of frameworks. Each has strengths and can serve as to organize thinking and work in the field of character education and social and emotional learning. However, their large numbers reflects a struggle for clarity, and each researcher, practitioner, and policy maker interested in the field is well advised to come to grips with the question of what it is they mean when they talk about character or socio-emotional learning. Often, this will involve

adopting a model in its entirety. In the measurement development arena, however, this may be difficult, because existing models generally cover a vast conceptual landscape that may be difficult to operationalize in a way that lends itself to measurement development.

To address this challenge, for example, my colleagues and I have been working on building scalable, web-based systems to measure social and emotional skills in the elementary grades (McKown, Russo-Ponsaran, Allen, Johnson, & Russo, 2016). We divide those skills into specific thinking skills, like the ability to understand another person's thoughts and feelings and solve social problems, and behavioral skills, like the ability to join an ongoing group and help someone in need. We also include self-control, which has both mental and behavioral components. Identifying those three broad domains to operationalize social and emotional learning provided us a point of entry for identifying crucial component skills that are measurable, meaningful, and malleable.

In our effort to capture what is most important, we have made commitments about what is and what is not included in the social and emotional arena, which has given us the clarity of purpose needed to build robust measurement systems that largely meet the standards articulated by Dr. Card. Our commitment to a model very specifically and strongly influenced assessment design considerations. We do not claim to have a perfect answer. However, it is surely a good sign that colleagues from different "camps" who care about children's social and emotional development have asked us to partner with them to provide measurement and assessment support. I urge scientists and practitioners alike to be diligent about clarifying precisely what is being measured. This will stimulate the adoption, adaptation, and development of assessments that are sound and useful.

PRACTICAL ASSESSMENT AND ITS CONSEQUENCES

Dr. Card's article focuses largely on assessment for science. There is also an urgent need

for good assessments for practice. In addition to the aspects of validity that Dr. Card described, for practice, assessments should demonstrate what Samuel Messick called "consequential validity," which refers to the ways in which test scores are interpreted as a basis for action, and the consequences, both intended and unintended, of those actions (Messick, 1995). If this sounds esoteric, a real-life example will show that it is not. The CORE districts is a consortium of 10 large school districts in California who have been using self-report measures of self-efficacy, social awareness, mindsets, and self-management as part of their accountability system. I believe what they are doing is a bold and important experiment—using measures of these skills to determine how well schools are doing their jobs. But not many months ago, a very public controversy unfolded on the pages of the *New York Times*, with prominent figures in the field criticizing this endeavor, arguing that the measures did not have the qualities that justified their use for accountability (Duckworth, 2016).

At issue in the CORE districts was consequential validity, with the key question being this: Are the measures of character chosen by the CORE districts appropriate indicators of school performance and are the scores they yield a reasonable basis for accountability-related consequences? This very important question highlights the importance of the consequential validity of all measures of character and social and emotional learning (and, by the way, achievement). For any measure, consequential validity can be only partly evaluated by rigorous study of the measure's technical properties. At least some of a measure's consequential validity is a matter of social values and the decisions and actions people take on the basis of assessment results. In considering the validity of measures, if we are being complete in our work, we cannot therefore be totally insulated from the vicissitudes of social values and our historical moment in its glorious complexity.

SENSE OF PURPOSE IN ASSESSMENT

A clear intention can place constructive boundaries on the consequences of assessment. Contrast fictitious Programs A and B. Leaders of Program A have decided to measure many dimensions of character and determine the use of those measures afterwards. In contrast, leaders of Program B have decided to measure particular social and emotional skills specifically and exclusively to inform instructional planning. In Program A, how assessment data will be interpreted and used is unclear. Therefore, the possibility that it will be used for non-valid purposes is high. In addition, in Program A, because no one is clear about the goals and therefore payoff of assessment, it is likely that considerable resources will be expended on assessment that will not yield any benefit.

In contrast, in Program B, because the purpose of assessment is clear, training in the interpretation and use of assessment data can be focused and practical. This will increase the odds the data will be used as intended. In Program B, all players know the purpose of assessment. Therefore, they will expect the data to be used in a particular way, increasing the likelihood that it will be used as intended and will be beneficial. Equally important, in Program B, all player understand a large number of decisions that will *not* be informed by the data—school and teacher accountability, special education placement, et cetera. Therefore, after data are collected, constituents will be less anxious that data may be used against them. It is still of course possible that in Program B, assessment data will have negative unintended consequences, but the range of those negative consequences has been significantly reduced.

As practitioners consider implementing assessments, it is important to note that at the present moment, the purposes for which social and emotional assessment can be fully used are limited. Good character and social and emotional learning assessments can help clarify

student need to inform instruction. In other words, the current state of the art supports, in my opinion, high-quality formative assessment. In addition, existing assessments are promising for program evaluation purposes. However, fewer character and social and emotional learning assessments have the rigorous psychometric properties, well-articulated in Dr. Card's article, commonly demanded of assessments used for high-stakes accountability purposes or student placement.

WISELY SELECTING METHODS OF ASSESSMENT

Finally, Dr. Card referred to a rarely-considered but critical consideration in the assessment of character and social and emotional learning. Specifically, in the best of all worlds, the method of assessment should be matched to what is being measured. By method of assessment, I am referring formally to the procedure through which an assessment samples behaviors hypothesized to reflect an underlying character or social and emotional learning skill. In discussions of assessment, surveys are often given as examples. However, there are many other methods of measurement. Observation, direct behavior ratings (<http://dbr.education.uconn.edu/>), and direct assessments, in which children demonstrate their skill through solving challenging problems (McKown et al., 2016), are all viable options.

Here is the important part: no single method can measure everything well and each method is better suited to measuring some things than others. To assess how well a child reads, we can ask her to fill out a self-report questionnaire. But a sound direct assessment of reading—in which she reads something and answers questions about what she read, for example—is likely to provide more useful and valid data. Similarly, to measure how well children read facial expressions, we can ask them to rate their skill level. But I would venture to say that direct assessment, in which children look at faces and indicate what emo-

tion the faces reflect, is more valid. To measure behavior, teacher report is probably better than self-report, and certainly more practical than observation. To measure peer acceptance and networks, peer nominations are superior to teacher report and other methods. Reasonable people can disagree about what method is best-suited to measuring what construct. What is important is that researchers and practitioners seriously consider what method of assessment is best for what they want to assess.

THE STAKES

The stakes are high. Yes, the scientific study of character education, which is the focus of Dr. Card's article, depends heavily on developing some consensus about what to measure and how to measure it. I would argue that no less than the survival of the character education and social and emotional learning enterprises—from policy to practice to research—depends on our ability to assess these skills well. How else can we know what children's strengths and needs are and therefore how to target instruction to foster character? That is formative assessment. How else can we know if a set of practices intended to foster character worked? That is program evaluation. How else can we know to what heights of character development students have risen? That is perhaps summative assessment. How else can we know if our system of education has met state standards (assuming such standards apply to the education of character)?

These are not idle questions. If nature abhors vacuums, educational fads feast on them. Without evidence, rooted in good measurement, the pendulum tends to swing from one fad to another. All of us—scientists, practitioners, and policymakers alike—should hope that the very best evidence of what works will be used to spur the evolution of effective educational and youth development programs and practices. Good measurement is foundational to collecting such evidence. If, however, we do not measure character and social and

emotional learning skill well, these fields will be buffeted by the winds of fad and polemics and they risk ending up on the dust pile of bygone movements.

In summary, in addition to Dr. Card's thoughtful and useful recommendations, there are four important considerations: getting to conceptual clarity; designing assessment for practice; being clear about the purposes of assessment; and selecting the method of assessment best suited to what it is we want to measure. The field is in an excellent position to translate these imperatives to functional, technically sound assessment systems. To do so will, in my opinion, require sustained collaborative effort, financial support, and cooperation between university researchers, educators, policy makers, and the private sector. It is heartening that these considerations are being deeply considered by many in the field. For example, under the leadership of Roger Weissberg and Jeremy Taylor from CASEL, a diverse collaborative is working to advance the field of social and emotional assessment. The stakes are high, and we would do well to move with all deliberate haste toward the development of practical, useful and scientifically sound assessment systems.

Acknowledgment: This work was supported by Institute of Education Sciences through Grants R305A110143 and R305A140562 to Rush University Medical Center. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education.

REFERENCES

- Collaborative for Academic Social and Emotional Learning. (2017). Core SEL Competencies. Retrieved July 1, 2017, from <http://www.casel.org/core-competencies>
- Duckworth, A. (2016, March 26). Don't grade schools on grit. *The New York Times*. Retrieved from <https://www.nytimes.com/2016/03/27/opinion/sunday/dont-grade-schools-on-grit.html>

- Jones, S. M., & Bouffard, S. M. (2012). Social emotional learning in schools: From programs to strategies. *SRCD Policy Report*, 26, 1–33.
- McKown, C., Russo-Ponsaran, N. M., Allen, A. A., Johnson, J., & Russo, J. (2016). Web-based direct assessment of children's social-emotional comprehension. *Journal of Psychoeducational Assessment*, 34, 322–338. doi:10.1177/0734282915604564
- Messick, S. (1995). Validity of psychological assessments: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 9, 741–749. doi:0003-066X/95
- National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press.

Copyright of Journal of Character Education is the property of Information Age Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.