

## **Measurement as Leverage between Research and Educational Practice.<sup>1</sup>**

**John Q. Easton, National Forum on Education Statistics. Washington, D.C. July 15, 2013.**

A little over one year ago I gave a talk at the National Council on Measurement in Education meetings in April 2012,<sup>2</sup> where I extolled the “power of measurement.” In that speech I urged the psychometricians in the audience to help educators solve real-world problems. I gave three examples: improved student assessments, help with teacher and principal evaluations, and the measurement of non-cognitive skills. This third topic is one that I have been reading and thinking about a lot lately. Even though it may not be on the top of your own agendas, I am going to use it as an example for how good measurement can be a lever for moving research into practice through partnerships. Measurement is usually defined as the process of quantifying an underlying construct, but another way of thinking about it is turning data into information, an idea you are very familiar with.

In this talk I am trying to weave three ideas together: the importance and potential positive impacts of good measurement; the importance of non-cognitive skills for children and adults; and, finally, the importance of partnerships between researchers and practitioners.

---

<sup>1</sup> This is a revised and abridged version of a talk originally presented at the University of Virginia on June 20, 2013. <http://ies.ed.gov/director/pdf/Easton062013.pdf>

<sup>2</sup> <http://ies.ed.gov/director/pdf/Easton041412.pdf>

But I'm going to start on a cautionary note. A couple of months ago I picked up a classic book that I read many years ago called *The Mismeasure of Man* by Stephen Jay Gould, written in 1981<sup>3</sup>. Gould describes a century of scientific endeavor centered on the use of ".....numbers to rank people in a single series of worthiness, invariably to find that oppressed and disadvantaged groups—races, classes, or sexes – are innately inferior and deserve their status."<sup>4</sup> Hence the title of his book, *The Mismeasure of Man*. The book describes how scientists first measured heads, then bodies, and then they hit upon IQ. Over a century or so, most of the studies ended up "proving" the superiority of white northern Europeans over just about everyone else.

Much of Gould's book is about the idea of reification—also called the reification fallacy. This is when we treat an abstract idea as if it were a real tangible thing. This often happens after we've learned to measure something. After Binet invented the intelligence test, a generation of scientists reified intelligence. They arrived at a point where they could measure intelligence better than they could define it and better than they understood it. The IQ score itself took on a meaning of its own, a life of its own, one much more precise than the meaning of the underlying concept that it was meant to measure.

I was director of research and evaluation for Chicago Public Schools in the mid-1990's and was responsible for administering, scoring and reporting the *Iowa Tests of*

---

<sup>3</sup> Gould, Stephen Jay. 1981. *The Mismeasure of Man*. New York: W. W. Norton & Company

<sup>4</sup> Gould, Page 25.

*Basic Skills.* What really struck me then, when accountability on this scale was unprecedented, was how the scores themselves took on meanings of their own – as if 3.8 or 6.3 or 5.5 meant something real. I was always astounded at how little interest people had in the actual content of the test. What were the knowledge, skills and abilities that students were being asked to demonstrate?

In the new accountability era of the mid-1990's many school districts stopped talking about student learning and talked about test scores instead. The scores themselves became more important than what they were supposed to be measuring. This is another example of reification, when the ease of the quantification overtakes the meaning and importance of the underlying concepts. It is easier to talk about test scores than it is to talk about what we want students to know and do.

It's plain that measurement isn't always beneficial– but it can be. Let me turn to some examples of good measurement and explore different ways that it can help us. Good measurement can help us in at least three different ways: It can facilitate scientific inquiry and add to scientific knowledge as we turn data into information; it can improve communication between different stakeholders; and it can catalyze productive and positive responses or behaviors.

I have talked and written a lot about a leading indicator of high school graduation that many colleagues and I developed in Chicago more than ten years ago. We called

this the “on-track to graduate” indicator.<sup>5</sup> It is a very simple binary indicator created from administrative records. A student is either on-track or not, depending on whether he or she had accumulated at least five full credits and had no more than one semester F in a core subject at the end of the freshman year. This isn’t a very high bar, yet students who are on-track are four to five times more likely to graduate from high school in four years. What’s especially important here is that the on-track indicator is a better predictor of high school graduation than students’ demographic characteristics, including their age, prior grade retention, school mobility history, and their prior achievement including 8th grade test scores.

We did a lot of digging around to figure out why this indicator was so important. We asked how are student behaviors (e.g., attendance, homework, or classroom engagement) associated with increasing the likelihood of being on-track for graduation? Are the predictors of being on track freshman year different than the predictors of high GPA? And, what elements of school and classroom environments are associated with student behaviors that lead to improved class performance?

Our on track indicator is a reliable measure with predictive validity. It is simple, easy to understand, easy to calculate and easy to talk about. Good measurement brings about conceptual clarity by helping us to build and test models for the linkages among

---

<sup>5</sup>Allensworth, E. & Easton, J.Q. 2005. *The On-Track Indicator as a Predictor of High School Graduation*. Chicago: Consortium on Chicago School Research; Allensworth, E. & Easton, J.Q. 2007. *What Matters for Staying On-Track and Graduating in Chicago Public Schools*. Chicago: Consortium on Chicago School Research. .

being on-track and school attendance, classroom engagement, doing homework and other relevant behaviors. It facilitates communication among researchers and the wide range of stakeholders who are seeking ways to improve student high school graduation rates. A common vocabulary makes communication across stakeholder roles much easier. Parents, teachers, principals, researchers, administrators and newspaper reporters understand what it means. And finally, in this case, the measurement also spurred to a lot of productive action. Schools across the city began watching student engagement from the first day of freshman year and intervened when they saw signs that students appear likely to fall off track. My colleague Elaine Allensworth just published a paper on the use of on-track indicators in improving high school graduation rates in Chicago.<sup>6</sup>

In the last year or so, I've been corresponding with a cochlear implant surgeon at the University of Chicago, Dana Suskind. Outside of her surgery, Dana has embarked on an ambitious project called *The Thirty Million Words Project*,<sup>7</sup> which was recently described in a New York Times blog called *The Power of Talking to Your Baby*.<sup>8</sup> This project is named for the estimated gap in the number of words that affluent three-year olds have heard spoken and the number of words that poor three-year olds have heard. Suskind works with poor mothers to help them increase verbal interactions with their

---

<sup>6</sup> Allensworth, E. (2013). The use of ninth-grade early warning indicators to improve Chicago schools. *Journal of Education for Students Placed at Risk*. 18:68-83.

<sup>7</sup><http://itm.uchicago.edu/2011/07/29/closing-the-30-million-word-gap/>

<sup>8</sup><http://opinionator.blogs.nytimes.com/2013/04/10/the-power-of-talking-to-your-baby/>

infants. The curriculum that she developed begins by describing growth mindset – the idea made known by Carol Dweck that brains grow in response to exercise.<sup>9</sup> It and then provides mothers with tips on how to talk to their children and ask and answer questions.

Measurement comes in to play with a device called LENA – Language Analysis System. It is an electronic device strapped on to the infant to record his or her sound environment. What I find very cool about LENA is how it can provide mothers with feedback, which in turn can motivate positive behavior on their part by encouraging them to increase their verbal interactions. Suskind says, “LENA provides concrete feedback in an easy to understand format.” One mother says, “getting the feedback from LENA was very motivating.” Another, “every week I was waiting for the report to come back ... as the weeks went on I would challenge myself.”<sup>10</sup>

Two months ago I bought myself a Fitbit hoping that it would help motivate me to get some more exercise. I didn’t know that there is actually published research on the benefits of pedometers and the feedback they provide until I read a paper on the effects of LENA feedback on adult verbal production by Dana Suskind and several of her colleagues<sup>11</sup>. Suskind’s paper drew the analogy between the quantitative linguistic

---

<sup>9</sup> Dweck, C. 2006. *Mindset: The new psychology of success*. New York: Ballantine Books.

<sup>10</sup> All quotes from video at <http://itm.uchicago.edu/2011/07/29/closing-the-30-million-word-gap/>

<sup>11</sup> Suskind, D., Leffel, K.R., Hernandez, M.W., Sapolich, S.G., Suskind, E., Kirkham, E. & Meehan, P. 2013. An exploratory study of “quantitative linguistic feedback”: Effect of LENA feedback on adult language production. *Communication Disorders Quarterly*. Published online February 1 2013.

feedback from LENA and the information that pedometers and more sophisticated devices like the Fitbit provide hopeful.

Let me move on now to talk about non-cognitive factors. Everybody hates this term but everyone knows roughly what you mean when you use it and no one has a much better alternative. I've been thinking a lot about these for the past two years, especially as the country has been engaged in vociferous debates about the appropriate use of student achievement test scores in teacher and principal evaluations. I've worried that in the past decades we've become so consumed with achievement test scores that we've marginalized the many other skills that are so crucial for children's healthy development and for their success both in and out of school. These include motivation, determination, perseverance, study skills, ability to work with others, resilience, metacognition, mindset, grit, social and emotional skills and many others. I'm going to remain indiscriminate about these various skills and terms, knowing that they are very different from each other and some are either more important or more malleable than others.

In the public sphere, we hear more and more about the importance of non-cognitive skills as well. In addition to reading their academic publications, you can see and hear famous scholars discussing their work for a public audience on youtube. James Heckman is on youtube talking about the influence of noncognitive skills on life

outcomes and how “skills beget skills.”<sup>12</sup> Carol Dweck is on talking about the positive effects of growth mindset on academic outcomes.<sup>13</sup> Angela Duckworth is there on the importance of grit in making it through West Point, in winning spelling bees and in graduating from high school in Chicago. In a recent TED talk, she advocates developing growth mindset as the best way to make kids grittier.<sup>14</sup>

Both Angela Duckworth and Carol Dweck were at a meeting about growth mindset in Washington sponsored by the White House in the middle of May. Top officials in the government believe the evidence is pretty strong that growth mindset can be relatively easily taught at a large scale.

Last year, the journalist Paul Tough wrote a beautifully compelling book called *How children succeed: Grit, curiosity and the hidden power of character*.<sup>15</sup> Tough argues that our national obsession with test scores is misguided and that schools have actually been emphasizing the wrong skill sets. His book provides many thorough anecdotal descriptions of programs and schools that emphasize the development of non-cognitive skills.

This is a hot topic. I know of four independent review papers on the importance of non-cognitive factors to children’s success in and out of school that have appeared in the last twelve months.

---

<sup>12</sup> <http://www.youtube.com/watch?v=eAp1Q4ezAXY>

<sup>13</sup> <http://www.youtube.com/watch?v=ICILzbB1Obg>

<sup>14</sup> <http://www.youtube.com/watch?v=H14bBuluwB8>

<sup>15</sup> Tough, P. 2012. *How children succeed: Grit, curiosity, and the hidden power of character*. New York: Houghton Mifflin Harcourt Publishing Company.

You may be familiar with an influential 2011 paper by Raj Chetty, John Freidman and Jonah Rockoff called “The long-term impacts of teachers; Teacher value-added and student outcomes in adulthood.”<sup>16</sup> These researchers found that students who had one high value add teacher (+1 SD above average) in reading or math in grades 4 to 9 were more likely to attend college at age 20, have steeper earning trajectories and reduced likelihood of having children as a teenager. This study provides additional evidence that good teachers (as measured by value-add) matter, not just in the short run, but in the longer run as well.

What I find fascinating about this study is that the high value-add teachers gave students a bump up in their test scores and in their learning rates, but 2/3 of the bump faded out after a few years. As you know, this is a common finding in education research. An intervention that may look good in the short term loses its impact over the long term. But here we have the fade out in test scores yet we still see the improvements in very important distal outcomes, like going to college, earning money and avoiding teenage births.

What’s going on here? I think that the effective teachers are indeed boosting their students’ achievement, but they are also boosting other important skills that aren’t measured in this study. I am betting that they could be psychological constructs like grit,

---

<sup>16</sup> Chetty, R., Friedman, J.N. & Rockoff, J.E. 2011. The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. National Bureau of Economic Research. Working Paper 17699.

perseverance, self-control, engagement, emotional intelligence, social emotional learning, or sense of mastery.

In a December 2012 NBER working paper, a young economist at Northwestern University named Kirabo Jackson analyzed the impact of teachers on students' cognitive ability as measured by state achievement test scores, and on their non-cognitive ability as measured by absences, suspensions, grades and grade progress, all obtained from administrative records.<sup>17</sup> Jackson asserts that this is the first paper to comprehensively analyze teacher effects on both cognitive and non-cognitive outcomes. The paper specifically addresses the question of whether teachers can improve skills that are not measured on achievement tests.

Jackson says, "This paper presents the first evidence that teachers have meaningful effects on non-cognitive outcomes that are strongly associated with adult outcomes *and are not captured by test scores.*"<sup>18</sup> While test scores may pick up some skills beyond achievement, there are many others that are not being measured. The implications of this finding are pretty obvious: an overemphasis on test scores may steer teachers away from helping students develop non-cognitive skills (another narrowing of the curriculum) and many highly effective teachers will not be identified by a sole focus on student test scores.

---

<sup>17</sup> Jackson, C. K. (2012). Non-cognitive ability, test scores and teacher quality: Evidence from 9<sup>th</sup> grade teachers in North Carolina. NBER Working Paper 18624.

<sup>18</sup> Page 4, emphasis in original

Jackson uses econometric modeling that is far beyond my technical capacity. But his basic method was to develop and test a model on NELS:88 data and then apply it to data from about 350,000 middle and high school students and their 3,500 Algebra I and 4,300 English I teachers in North Carolina. The NELS analysis shows that non-cognitive abilities are associated with improved adult outcomes and that they are at least as important as the cognitive abilities. The North Carolina data show two important findings: that teachers have larger impacts on students' non-cognitive abilities than on their cognitive abilities and that teacher cognitive effects are only weakly correlated with teacher non-cognitive effects. The implications are pretty stark and startling. If Jackson's findings are correct and it's true that the words of another group of economists -- "Success in life depends on personality skills that are not well captured by achievement scores. Conscientiousness, perseverance, sociability, and curiosity matter."<sup>19</sup> then we need to think hard about what schools and teachers should be doing.

We probably agree that motivation and the so-called non-cognitives are important. We may not know which are the most important but we can probably agree that we should focus on the most salient, actionable and malleable skills rather deeply ingrained stable traits.

---

<sup>19</sup> Boghans et al, 2012,

It's clear that we have some measurement issues to deal with. There are no easy measures here, no simple on-track to graduate indicator, or count of adult verbal interactions, grade equivalent scores, steps taken or stairs climbed to give us feedback and nudge us on. But there are opportunities for us in administrative data, including absences, grades and behavior.

We also know that that some of our current measures of these non-cognitive skills can be problematic. We have some good measures for research purposes, but I wonder if many of them are of much use beyond research. Self-report Likert items can be problematic. How can we measure the right things in the right way so that good measurement can lead to the three things that I said they can? How do we make sure that we don't end up reifying a non-cognitive variable just because we can measure it? We're not so naïve that we will reify the GRIT scale, Hands Toes Elbows Knees, the KIPP Character report card, or Galvanic Skin Response. But measurement really is the key for us if we want to move these non-cognitive skills on a large scale for a lot of kids.

I want to wrap by talking about partnerships between researchers and practitioners. Before I moved to Washington four years ago I lived and worked in Chicago for more than thirty years. For the last 12 years I worked at the Consortium on Chicago School Research at the University of Chicago. I've said many times how much I loved working at the Consortium with the school district. I enjoyed the research itself, but I particularly loved being part of a broader community that wanted to help make the

Chicago Public Schools better for students, their families and for their teachers. From the earliest days, we researchers at CCSR wanted our research to matter, and importantly, we believed that it could. I believe that good measurement facilitated our partnership with CPS and it can facilitate other partnerships as well. Everyone can be on the same page when they know what it is they are talking about. It could be on-track to graduate, kindergarten readiness, college readiness, resilience, perseverance, growth mindset or any of dozens of other important concepts.

IES has taken up the idea of partnerships in several ways because we believe that research conducted in these partnerships is more likely to be relevant and to get used. Our new programs, Researcher Practitioner Partnerships and Continuous Improvement Research in Education share the belief that the equal partnerships will require researchers to focus on topics of interest to the educators and that by having some skin in the game the educators will make use of the findings. We are really promoting research use, but not in a unidirectional “research to practice” sense but in a more reciprocal “practice to research” pathway. You will hear more about the REL’s Research Alliances this afternoon.

We just awarded six new Researcher Practitioner Partnership grants and we are competing a new topic called Continuous Improvement Research Education – the RFA was posted in early May and proposals are due in September. For the Continuous Research in Education program we are looking for established partners to work together

to fine tune programs, interventions or regimens of activities through iterative processes that rely heavily on measurement, quick studies and refinement. Some people have called this “empirical tinkering” and likened it to an engineering process. To bring together some of this new work with a partnership focus, we are also competing a new research and development center on knowledge utilization. We expect this center to study how educators use research and how researchers conduct usable research.

I believe that through partnerships and with the aid of good measurement education researchers can have a great impact on school improvement efforts by conducting usable research that is meaningful to educators. I also believe that the whole terrain of non-cognitive skills is extremely important for the healthy development and future success of our school children and youth. Good measurement can facilitate these partnerships, it can help us focus on and talk about important skills, keep them in proper perspective, and at the same time help us build a body of scientific knowledge.

Thank you.