A Rosetta Stone for Noncognitive Skills

Understanding, Assessing, and Enhancing Noncognitive Skills in Primary and Secondary Education

January 2015

Asia Society

Professional Examination Service

Richard D. Roberts, Jonathan E. Martin, and Gabriel Olaru

Professional
Examination Service

A Rosetta Stone for Noncognitive Skills

Understanding, Assessing, and Enhancing Noncognitive Skills in Primary and Secondary Education

January 2015

Richard D. Roberts, Jonathan E. Martin, and Gabriel Olaru

ASIA SOCIETY

Asia Society is the leading educational organization dedicated to promoting mutual understanding and strengthening partnerships among peoples, leaders, and institutions of Asia and the United States in a global context. Founded in 1956 by John D. Rockefeller 3rd, Asia Society today is a global institution—with offices throughout the United States and Asia—that fulfills its educational mandate through a wide range of cross-disciplinary programming. Across the fields of arts, business, culture, education, and policy, the Society provides insight, generates ideas, and promotes collaboration to address present challenges and create a shared future. Asia Society's education department partners with education leaders, primary and secondary education schools, non-formal learning programs, and education systems in the United States and Asia to ensure that students and young leaders develop global competence as the foundation for understanding between people in the Asia Pacific region and the United States, and throughout the world.

PROFESSIONAL EXAMINATION SERVICE

Professional Examination Service (ProExam), founded in 1941, provides a full range of assessment and advisory services to organizations across a broad range of professions in support of professional licensure and certification, training, and continuing professional education programs. As a 501(c)(3) not-for-profit organization, ProExam specializes in delivering tailored services to build and enhance credentialing programs. ProExam's Center for Innovative Assessments helps clients develop new ways to accurately measure noncognitive skills. The Center serves ProExam's traditional clients, as well as new markets that traverse the educational and workplace sectors.

ABOUT THE AUTHORS

Richard D. Roberts, Ph.D. is currently Vice President and Chief Scientist, Center for Innovative Assessments, Professional Examination Service. A former National Research Council Fellow who conducted research at Brooks Air Force Base (1996-1998), he has also been a Senior Lecturer at The University of Sydney (1998-2003) and a Managing Principal Research Scientist at the Educational Testing Service (2003-2014). His main area of specialization is assessment, with a special emphasis on developing and researching innovative new items types for the measurement of both cognitive and noncognitive factors, as well as constructs that have elements that straddle these two domains (e.g., critical thinking and emotional intelligence). Dr. Roberts has published over a dozen books and 200 peer-review articles or book chapters on these topics in diverse sub-disciplines (including, education, psychology, business, medicine, and wind engineering), with nearly 400 presentations across the globe.

Jonathan E. Martin has 15 years of experience as a school principal and is a consultant and writer on 21st century learning and assessment. He is lead consultant to the Secondary School Admission Test Board's Think Tank on the Future of Assessment and is Project Manager for SSATB's program developing a new noncognitive assessment for admission selection purposes. He is the author of the "OECD Test for Schools (based on PISA) User's Guide" and of the "Mission Skills Assessment User's Guide and Toolkit."

Gabriel Olaru is currently a Graduate Research Assistant in the Department of Differential Psychology and Psychological Diagnostics at the University of Ulm. His main area of specialization is personality and assessment, with a main focus in research on the Big Five and current methods in evaluating and improving psychological measurement instruments.

AUTHORS' NOTE

The impetus for this paper came from Asia Society, whose Global Cities Education Network (GCEN) is examining international best practices for advancing social, emotional, and personal competencies at the systems level. The GCEN is a network of urban school systems in North America and Asia that explores common challenges, benchmarks international best practices, and supports adaptation and implementation of these practices within local contexts.

Tony Jackson and Alexis Menten on Asia Society's staff helped to frame the paper, and the Charles Stewart Mott Foundation provided resources to support the publication and dissemination of this paper to members of the Global Cities Education Network and other stakeholders.

The views and opinions expressed in this article are solely those of the authors and do not purport to represent the views of the Asia Society, Professional Examination Service, or any of the organizations the authors are variously affiliated with.

The authors would like to thank Robert Block, Simmy Ziv-el, Carolyn MacCann, Ralf Schulze, Christina Cappiello, Sagar Athota, Selina Weiss and Johanna Hartung for support and/or assistance in the preparation of this manuscript.

CONTENTS

INTRODUCTION

A commitment to developing character, social and emotional skills, and 21st century competencies can be found in the mission statements of many schools across the globe (Stemler & Bebell, 2012), and in national policy statements worldwide (Torrente, Alimchandani, & Aber, 2015). In addition to delivering academic learning, schools proclaim their commitment to developing students to become life-long learners, skillful collaborators, moral individuals, confident and persistent problem-solvers, organized and conscientious leaders, innovative thinkers, and much more. These types of outcomes, however, are rarely intentionally inculcated through primary and secondary education teaching and learning. As a result, such skills are often fostered through informal means—as a byproduct of good teaching or good parenting—or through non-formal programs, including extracurricular activities and programs organized by community-based organizations. This means that some students benefit from the opportunity to develop these skills, while others do not. In fact, they are as important predictors of success in school and careers as academic abilities, and thus essential for all students.

At the same time, policymakers and employers around the world are realizing the mismatch between the outcomes promoted by their education systems and the skills required for work and life in the 21st century. Recognizing that more careful attention needs to be given to their place in primary and secondary educational practice, various organizations have identified as essential many of these types of skills in recent years. Work ethic, teamwork, oral communication, leadership, creativity, and life-long learning topped the list of items most sought after by employers in the US in a report commissioned by a consortium that included the Partnership for 21st Century Skills, Conference Board, Society for Human Resources, and Corporate Voices for Working Families (Casner-Lotto & Barrington, 2006). While also valued, cognitive skills ran a distant second. More alarmingly perhaps, the report also noted that noncognitive skills were precisely the skills most often found lacking in new employees joining the workforce (i.e., former students of the primary and secondary education system).

What is needed is an evidence-based framework to help primary and secondary education policymakers and educators make sense of the myriad skills beyond academics that are critical for 21st century success, along with strategies and approaches to effectively teach and reliably assess these skills. This paper puts forward one such framework, the Big Five personality factors, which can act as a Rosetta Stone to "translate" the various concepts and terms used among and between researchers and practitioners, economists and businesspeople, and policymakers in education systems in different countries.

Interpreting critical noncognitive educational outcomes through the lens of the Big Five tethers them to the hundreds, thousands even, of psychological research studies conducted in the past two decades. The conclusion of this research is clear and compelling: these traits matter. Research has shown that several of these traits are as important for academic performance as cognitive ability is, and that these traits positively predict performance, behavior, and satisfaction in work life. Where once social and emotional learning appeared problematic, or at best, only relevant to early childhood and primary education, there now appears a solid evidentiary base showing it is not only plausible, but also credible, through secondary and even post-secondary education. In the context of fairly recent studies showing that personality can change over the lifespan, research suggests that noncognitive factors could and should play a more pivotal role in educational policy and practice than hitherto realized.

This paper describes the Big Five factors, how they were determined, and how they have been demonstrated to be universal across different ages and consistent across different countries and cultures. It concludes by reviewing many of the approaches to assessment of these skills, and the related challenges and solutions. As many other reports and organizations have noted (e.g., Oxford Research, 2010; EFA Global Monitoring Report Team, 2012; Asia-Pacific Education Research Institutes Network, 2014), effective implementation strategies need to be further documented, shared, and evaluated for education systems to commit to systematic and intentional support of noncognitive outcomes for all students.

THE BIG FIVE FACTORS

At the turn of the millennium, the Partnership for 21st Century Skills (P21) launched its work with a long list of the skills that are now required for success in the 21st century. Then, recognizing the value of simplicity, P21 decided to put their focus on four, which they call the 4 C's: critical thinking, communication, collaboration, and creativity (though "self-regulation" is something of an honorary fifth in their publications). Others have attempted to distill the essence of character, social and emotional skills, and 21st century competencies into a manageable few. For example, the Collaborative for Academic, Social, and Emotional Learning (CASEL) selected five "competency clusters": self-awareness, self-management, social awareness, relationship skills, and responsible decision-making.

Tony Wagner's (2010) bestseller, *The Global Achievement Gap*, highlights seven "Survival Skills": problem-solving and critical thinking, collaboration across networks and leading by influence, agility and adaptability, initiative and entrepreneurship, effective written and oral communication, accessing and analyzing information, and curiosity and imagination. Another bestseller, Paul Tough's (2013) *How Children Succeed* champions grit, curiosity, and the hidden power of character in its subtitle. Indeed, many, many different terms have been used by educators, practitioners, policy makers, and researchers. We distill a subset of these in Table 1 that follows.

It is no simple task for educators to narrow down the above list, prioritize what is most important, and develop these skills in young people. It is easy to see why: they are all good things, at least in moderation, and their instruction is not self-evident. In 2012, the National Research Council, in its landmark report *Education for Life and Work*, attempted to cut through the morass by declaring three clusters of competencies: the Cognitive, the Interpersonal, and the Intrapersonal. They argue that the many terms inside of each cluster may or may not be meaningfully differentiated, and it would seem more important to keep the highest attention to the three clusters and to worry less about parsing them in much greater detail.

Table 1: Terms describing key noncognitive skills

Assertiveness	Adaptability	Cheerfulness	Collaboration
Collegiality	Communications	Confidence	Coping with Stress
Creativity	Cultural Competence	Curiosity	Dependability
Determination	Effortful Control	Enthusiasm	Entrepreneurialism
Ethical Behavior	Fairness	Friendliness	Generosity
Grit	Growth Mindset	Honesty	Imagination
Innovation	Integrity	Kindness	Leadership
Liveliness	Moderation	Optimism	Organization
Patience	Persistence	Planning	Professionalism
Punctuality	Resilience	Responsibility	Self-Consciousness
Self-Esteem	Self-Regulation	Sociability	Teamwork
Time Management	Tolerance	Trustworthiness	Work Ethic

A different approach is offered here. Rather than creating a new and original taxonomy, and instead of looking for something alliterative (the 4C's, the Seven Survival Skills), our approach instead draws upon an already well-established taxonomy of personality traits, called the Big Five personality factors, from which can be drawn the set of noncognitive attributes most significant for life and work. Decades of psychological research have found that the power of the Big Five factors lies precisely in the fact that these are the set of constructs that are

most strongly differentiated, non-overlapping, and predictive of valued societal outcomes across domains of school, work, and leisure (Kyllonen, Lipnevich, Burrus, & Roberts, 2014; MacCann, Duckworth, & Roberts, 2009; Poropat, 2009; Schmidt & Hunter, 1998). They also have the benefit of being already demonstrated as universal across the lifespan (Roberts & DelVecchio, 2000; Roberts, Walton, & Viechtbauer, 2006) and relatively consistent across cultures (De Raad & Perugini, 2002; McCrae & Terracciano, 2005; Schmitt, Allik, McCrae, & Benet-Martinez, 2007). Scientific evidence documents that the Big Five factors are reliable when measured in a wide variety of ways and in differing contexts (Jackson, Wood, Bogg, Walton, Harms, & Roberts, 2010; Lipnevich, MacCann, & Roberts, 2013; McCrae & Terracciano, 2005; Schmitt et al., 2007).

A ROSETTA STONE FOR NONCOGNITIVE SKILLS

The Big Five factors, were not "invented" per se (by researchers combing society, the economy, and citizenship to determine which traits are most important, and then determining how they might be best labeled and differentiated); rather they were in effect "discovered." Under the assumption that all important matters in life have been named and are thus represented in our language, Allport and Odbert (1936) searched Webster's New International Dictionary from 1925 for English words that described human characteristics. In total, 18,000 English words were selected, with 4,500 being classified as descriptions of stable personal traits. Cattell (1943) applied factor analytic procedures to reduce the massive list of traits by analyzing the underlying patterns among them. He then studied personality data from different sources (e.g., interpersonal ratings, objective measures of daily behavior, and questionnaire results), and measured these traits in diverse populations to arrive at 16 major personality factors (Cattell, 1957, 1973). He was not able to replicate his 16 factors again, but the modern psycholexical approach was born and applied by many researchers to come (Fiske, 1949; Norman, 1963, 1967; Tupes & Christal, 1961), resulting in five factors. In these analyses, initial semantic judgments are made on the extent to which terms refer to similar perceptible variations in performance and appearance between persons or within individuals over time and varying situations, and how these terms can be usefully clustered into groups. The validity of these associations and groupings are then tested by observing in representative populations of individuals whether these ways of discriminating between personality traits actually align with the way people classify other people they know. These analyses consistently yielded five factors that were labeled Extraversion, Agreeableness, Conscientiousness, Neuroticism/Emotional Stability, and Intellect (which was later relabeled Openness).

Extraversion describes a person's likelihood to engage in social interaction, but also a propensity towards sensation seeking. Less extraverted persons are more reserved and less socially dominant. Like Extraversion, Agreeableness is a trait primarily influencing interaction with others. A very agreeable person may be described as friendly, helpful, and empathic. A person with low Agreeableness is considered to be cold and unfriendly, but certainly not naive. Conscientiousness primarily describes achievement-related traits. A person high in Conscientiousness can be described as very organized, diligent, and perfectionistic. Emotional Stability, which is often referred to by its opposite pole Neuroticism, describes a person's capability to cope with stressful situations and emotions. Low Emotional Stability (i.e., Neuroticism) is often accompanied by feeling depressed, stressed, anxious, or worried. Highly emotional stable persons, however, also tend to be less cautious. Openness is somewhat related to cognitive ability (Ackerman & Heggestad, 1997) and can best be described as a person's interest in, and acceptance of, "new" cultures, ideas, values, artistic endeavors, and even feelings. The opposite pole of Openness is thought to comprise conservatism.

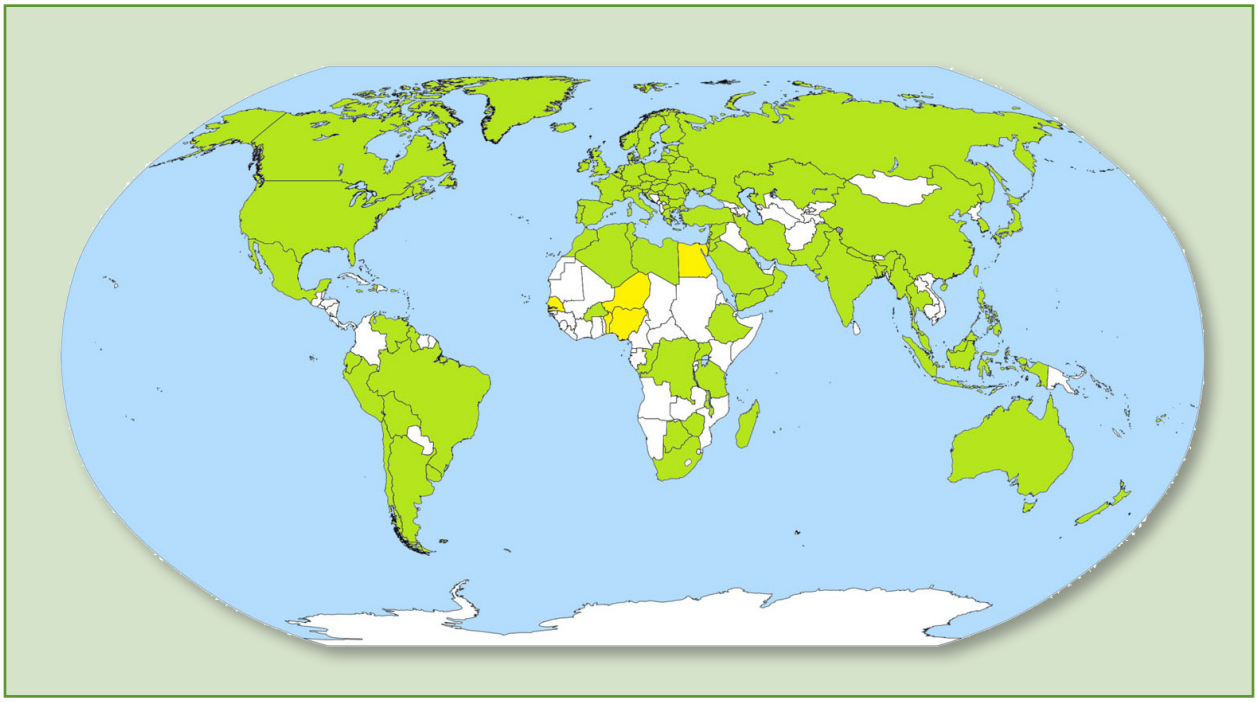
In contrast to centuries of personality research and models, starting back in ancient times with Hippocrates (ca. 400 BC) and Galen (ca. 150 AD), the Big Five emerged from a statistical analysis of natural language, instead

of theoretical assumptions or causal explanations. In effect, it can be argued that these constructs existed a priori of research (and researcher's egos).

Even though they were first discovered in the English language, replication studies, either involving the full psycholexical approach (De Raad, 2000), or applying translations of established Big Five personality inventories, resulted in the same five factors (McCrae & Terracciano, 2005; Schmitt et al., 2007). Indeed, this research has proven the Big Five's universality in the vast majority of countries, cultures, and languages across the world. In short, the Big Five plays an important role in human nature, independent of the environment writ large. Figure 1 displays the countries of the world where the Big Five has thus far been replicated.

The popularity and expansion of the Big Five depicted in, for example, Figure 1, emerged because it is the best compromise between an easy to understand model and an exhaustive representation of all aspects comprising a human's personality. This means that these broad Big Five factors measure personality relatively efficiently, but with relatively low fidelity (Soto & John, 2009). Thus, the search for "facets" below the higher order Big Five factors began (see e.g., Costa & McCrae, 1995; Christal, 1994; DeYoung, 2007; John & Srivasta, 1999; Soto & John, 2009). These facets help measure desired traits with much higher precision, and the appropriate facets can be picked for every situation, increasing time and cost efficiency.

Figure 1: Map of the globe showing where measurements of the Big Five model of personality have been applied and replicated (green = compelling evidence; yellow = suggestive evidence; white = inconclusive replication studies).



Seeing that the Big Five factors are broad and multifaceted, it becomes clear why newer, less well-established constructs (e.g., grit, growth mindset) can be integrated into the Big Five. Comparing the myriad of constructs given in Table 1 with the Big Five shows near perfect overlap (see Table 2); it is only in the interests of space that we do not classify all such constructs.

Table 2: Terms describing key noncognitive skills realigned by the Big Five factors

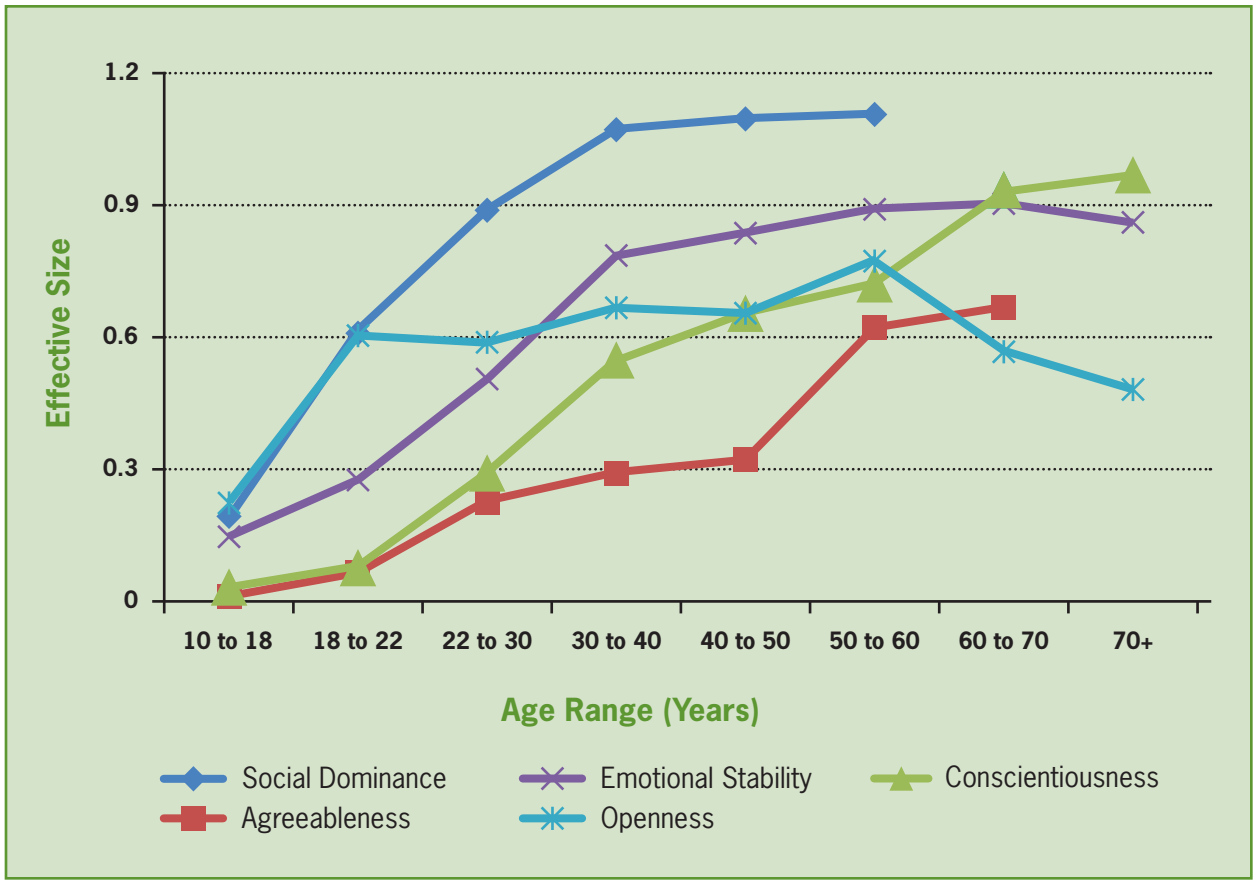
Conscientiousness	Agreeableness	Emotional Stability	Openness	Extraversion
Dependability	Collaboration	Confidence	Creativity	Assertiveness
Grit	Collegiality	Coping with Stress	Curiosity	Cheerfulness
Organization	Generosity	Moderation	Global Awareness	Communication
Persistence	Honesty	Resilience	Growth Mindset	Friendliness
Planning	Integrity	Self-Consciousness	Imagination	Leadership
Punctuality	Kindness	Self-Esteem	Innovation	Liveliness
Responsibility	Trustworthiness	Self-Regulation	Tolerance	Sociability

The Big Five can therefore be considered as something of a Rosetta Stone for understanding noncognitive skills and their typology. The Rosetta Stone, by offering the same content presented in the words of three different languages, allowed archaeologists to understand how each language related to the others on the stone, and how different words in different languages mean the same underlying thing. Using the Big Five factors, we can take concepts expressed as time management in one list, grit in another, and responsibility in still a third, and understand their connectedness by seeing them all as manifestations of Conscientiousness, at least in significant measure.

There is another aspect of the Big Five model that should not go unnoticed by educators. Until recently it was thought that personality was "set in stone" (McCrae & Costa, 1994). More recently, two meta-analyses have questioned this assertion and suggest instead that personality develops over a lifespan. In the first of these studies, Roberts and DelVecchio (2000) examined 152 longitudinal studies to show that the rank-order consistency of personality was fairly moderate: .31 in childhood, .54 in college, .64 by age 30, and .74 by ages 50-70 (values much closer to one would have supported the idea of personality being immutable). In a follow-up study, Roberts et al. (2006) examined mean-level change in personality over a lifespan. They found that individuals became more socially dominant (i.e., extraverted), conscientious, agreeable, and emotionally stable throughout a lifespan, particularly in adolescence and early adulthood. The effects were not slight: Change over a lifespan was up to a full standard deviation (see Figure 2).

There are two important corollaries related to this rather compelling research showing that personality can change. The first is that it frees up the potential for educational interventions: Where once social and emotional learning appeared problematic, or at best, only relevant to early childhood and primary education, there now appears a solid evidentiary base showing it is not only plausible, but also credible through secondary and even post-secondary education. Coupled with their high valuation by educators and employers alike, this research suggests that noncognitive factors could and should play a more pivotal role in educational policy and practice than hitherto realized. Put simply, there is a very high potential payoff from investment in the development of noncognitive skills.

Figure 2: Meta-analytic evidence showing that personality does change over a lifespan



RELEVANCE FOR EDUCATION SYSTEMS

As mentioned earlier, interpreting critical noncognitive skills through the lens of the Big Five aligns them to myriad other research studies. The skills and attributes thus move from the observations, opinions, and speculations of the past few years to the deeply grounded evidence of decades of psychological research. This allows educators to make far better use of existing research on what is important, what can be changed, and how meaningful change can happen. Table 3 gives a summary of the research that has thus far been conducted on the relationship between the Big Five factors and academic performance, with an accumulated sample size of over 70,000 students (Poropat, 2009). It displays the correlations between the Big Five factors and cognitive ability with academic performance (in this case, as measured by grade point average).

Table 3: Correlation of the Big Five and cognitive ability with grade point average in primary, secondary, and tertiary educational sectors as determined by a meta-analysis of over 70,000 students

	Educational Level		
	Primary	**Secondary**	**Tertiary**
Conscientiousness	.28	.21	.23
Agreeableness	.30	.05	.06
Emotional Stability	.20	.01	-.01
Openness	.24	.12	.07
Extraversion	.18	-.01	-.03
Cognitive Ability	*.58*	*.24*	*.23*

In the area of immediate interest to this readership, primary and secondary education, the research is compelling: these traits matter, most particularly in the primary grades, but also to a substantial extent in secondary and post-secondary educational environments. In secondary and tertiary education, Conscientiousness is nearly as important for academic performance as is cognitive ability, yet it receives less attention in large-scale group score assessments with policy impact (from the Program of International Student Assessment, or PISA, to the new assessments of the Common Core in the United States). Little attention, therefore, is paid to how these types of skills might be enhanced during a student's school career.

The last point in particular is apposite. Research in this field has led to the empirically founded assumption that cognitive ability may not be changed easily (Kyllonen, Roberts, & Stankov, 2008), but personality traits can be, and in fairly brief and sometimes innocuous ways (Dweck, 2012). Even so, more focused interventions may have especially powerful ramifications. Another recent meta-analysis on afterschool programs sheds light on how interventions related to these noncognitive factors play out in the one place they can be targeted in the United States without getting stonewalled by state and national educational policy or related large-scale assessments.

Summarizing the results of over seventy-five studies, and especially those afterschool programs where social and emotional skills are inculcated, Durlak, Weissberg, and Pachan (2010) note that these non-formal learning programs had an overall positive and statistically significant impact on the youth who participated. These changes did not occur in all domains, but rather in three main areas: feelings and attitudes, indicators of behavioral adjustment, and school performance. In particular, there were significant increases in youths' self-perceptions, bonding to school, positive social behaviors, school grades, and achievement test scores. Significant reductions also appeared for problem-related behaviors. In addition, certain programs that used a protocol focused on sequenced, active, focused, and explicit programming (that the authors describe at length) were associated with practical gains in participants' test scores of 12 percentile points between the afterschool and control group, a result that is similar to, or better than, those obtained by many other evidence-based interventions for school-aged populations. Durlak et al. (2010) concluded that current findings for afterschool programs "merit support and recognition as an important community setting for promoting youths' personal and social well-being and adjustment" (p. 302).

Even if our focus is exclusively upon academic achievement, supporting the development of a student's Conscientiousness, Agreeableness, Emotional Stability, Openness, and Extraversion (particularly in the early years) is of enormous importance. The story gets far more interesting if one moves beyond academic achievement, as the preceding passages suggest, to consider other outcomes of great merit in their own right. An especially important case in point is retention. Recently, Burrus et al. (2013) undertook a comprehensive review of the higher education literature to ascertain those factors that might readily be the most important to focus upon for student retention. Again, it may not come as a startling revelation that the Big Five personality factors play a significant role.

RELEVANCE FOR WORKFORCE SYSTEMS

The value of the Big Five as a Rosetta Stone does not end with education. The Big Five factors are also increasingly viewed as important for a variety of uses in the workforce: employment selection, career training, outcomes assessment, professional development/enhancement, and succession planning, to name a few. In this domain, they are being widely adopted as a near consensual model for behavioral economics, psychology (particularly, industrial, organizational, and military psychology), and, in the United States, Department of Labor policy and practice.

Consider the United States Department of Labor's Occupational Information Network (O*Net), which provides occupational definitions to help job seekers and businesses, and the human resource specialists who are pivotal in staffing these businesses, understand the world of work. It contains numerous resources, including lists of worker characteristics for each of approximately 1,102 occupations. The model ostensibly guiding these worker characteristics should come as no surprise: It is the Big Five.

In what is likely to prove an immensely influential piece, Sackett and Walmsley (2014) have gone through the entire O*Net database to highlight those noncognitive skills valued by each profession. Table 4 lists the importance of the Big Five to 23 job families representing over one thousand occupations distilled in O*Net. For the educator, interested in their student's eventual life beyond school and college, the results are telling.

Table 4: The relative importance of Big Five facets and factors to 23 job families representing approximately 1,102 occupations in O*Net (Big Five factors are in brackets)

	O*NET Worker Style Characteristic		
Job Family	**Top Ranked**	**Second Ranked**	**Third Ranked**
Architecture and Engineering	Analytical Thinking (O) / Dependability (C)	Integrity (C)	Initiative (C)
Arts, Design, Entertainment, Sports, and Media	Dependability (C)	Adaptability and Flexibility (N)	Initiative (C) / Stress Tolerance (N)
Building, Grounds Cleaning, and Maintenance	Dependability (C)	Cooperation (A)	Self-Control (N)
Business and Financial Operations	Integrity (C)	Dependability (C)	Analytical Thinking (O) / Cooperation (A)
Computer and Mathematical	Analytical Thinking (O)	Dependability (C)	Cooperation (A
Construction and Extraction	Dependability (C)	Cooperation (A)	Self-Control (N)
Education, Training, and Library	Dependability (C)	Integrity (C)	Self-Control (N)
Farming, Fishing, and Forestry	Dependability (C)	Self-Control (N)	Independence (C)
Food Preparation and Serving	Cooperation (A)	Dependability (C)	Self-Control (N)
Healthcare Support	Dependability (C)	Concern for Others (A)	Integrity (C)

Table 4 (continued)

	O*NET Worker Style Characteristic		
Job Family	**Top Ranked**	**Second Ranked**	**Third Ranked**
Healthcare Practitioners and Technical	Integrity (C)	Dependability (C)	Concern for Others (A)
Installation, Maintenance, and Repair	Dependability (C)	Integrity (C)	Cooperation (A)
Legal	Integrity (C)	Dependability (C)	Analytical Thinking (O)
Life, Physical, and Social Science	Integrity (C)	Analytical Thinking (O)	Dependability (C)
Management	Dependability (C)	Integrity (C)	Leadership (E, A)
Office and Administrative Support	Dependability (C)	Integrity (C)	Cooperation (A)
Personal Care and Service	Dependability (C)	Self-Control (N)	Integrity (C)
Production	Dependability (C)	Cooperation (A)	Integrity (C)
Sales and Related	Dependability (C)	Integrity (C)	Self-Control (N)
Transportation and Material Moving	Dependability (C)	Self-Control (N)	Integrity (C)

Well and good one might claim, but does this translate into meaningful information about career readiness and workplace success? Indeed, as is the case with its relation to educational outcomes, the influence of the Big Five on work-related outcomes and behavior has been assessed in many studies, which cumulatively account for several hundreds of thousands of persons (Barrick, Mount, & Judge, 2001; Berry, Ones, & Sackett, 2007; Judge, Rodell, Klinger, Simon, & Crawford, 2013; Sackett & Walmsley, 2014). Table 5 displays the correlations between the Big Five and a number of work-related outcomes, including overall job performance (usually supervisor ratings), task performance, organizational citizenship behavior, and counterproductive workplace behaviors.

Table 5: Relationships of the Big Five and a variety of workplace outcomes as determined by various meta-analyses totaling over 190,000 workers

	Job Performance	Task Performance	Organizational Citizenship Behavior	Counter-Productive Work Behavior
Conscientiousness	.33	.31	.40	-.40
Agreeableness	.22	.13	.23	-.51
Emotional Stability	.13	.11	.21	-.31
Openness	.10	.14	.04	-.08
Extraversion	.26	.15	.28	-.04

As in the educational context, greater Conscientiousness positively predicts performance and behavior in work life, and also increases job satisfaction (Sackett & Walmsley, 2014). One should not, however, underestimate the influence of the other Big Five factors. Agreeableness is not only something consistently shown to be of value in Table 5, it actually does a better job of predicting counterproductive workplace behaviors, which can cost the labor market millions (perhaps billions) of dollars per year, than any other skill. These Big Five factors have been shown to be related to more specific job performance: Managers perform better when being more extraverted, higher Emotional Stability and Agreeableness improve teamwork and interaction with others (e.g., clients), whereas higher Openness enhances training proficiency.

BIG FIVE PERSONALITY ASSESSMENT

As fascinating as this information about the Big Five might be, it is not valuable without effective ways to teach and measure these attributes. Neither researchers nor educators could determine which programs, approaches, and interventions are most effective; evaluate their success at their initiatives; and identify which students most needed this support without meaningful assessment techniques. And so the question becomes: Can the Big Five be measured? Several methods exist, as summarized below and then explained in further technical detail in the Appendix.

SELF-REPORTS

Self-reports have been used in noncognitive research for decades and have proven to be very efficient in gathering a lot of information over a brief period of time. Surveyed persons are asked to indicate their agreement with a number of different statements (e.g., "I like to work hard at school"). In order to gain more detailed information, respondents are not just answering whether they agree or not, but instead to report their level of agreement via Likert-type scales, which provide anywhere between four and seven response options that represent increasing grades of agreement. This type of assessment is preferred in environments when there are no stakes for the self-assessor and faking is not expected (Lipnevich et al., 2013). Respondents, however, may fake their responses on self-assessments to avoid having to attend training programs or to appear more attractive to a prospective school admissions officer, university system, or employer (Zickar, Gibby, & Robie, 2004). Fortunately, researchers have identified several promising methods for collecting data through self-reports while reducing fakability. These include giving real-time warnings, using a forced-choice format, and using one's estimates of how others will respond to help control for faking (Ziegler, MacCann, & Roberts, 2011).

FORCED-CHOICE ASSESSMENT

This procedure has many different aspects, including pair comparisons, rank-ordering, and multidimensional forced-choice. In pair comparisons¸ the test-taker must choose between two statements (e.g., which is more like you: "I work hard" or "I enjoy working in teams"?). In rank-ordering, test-takers must rank a series of equally desirable statements in order from "most like me" to "least like me." In multi-dimensional forced-choice assessments, test-takers are presented with a dichotomous quartet of four different traits, in which two socially desirable statements are paired with two socially undesirable statements (Jackson, Wroblewski, & Ashton, 2000). There is compelling evidence to suggest that forced-choice tests are less fakeable than standard rating scales, and show stronger relationships with performance outcomes (Drasgow et al., 2012; Jackson et al., 2000). An empirically-based procedure for item selection and test development, combined with new statistical modeling

techniques, seems to produce the best of all worlds: fake-proof normative tests that can also tell the individual how they score relatively on each dimension. To date, forced-choice assessments have not been widely used in primary and secondary education, although in principle they should work, certainly for high school students. While they may have acceptable psychometric properties, and even predict valued educational outcomes, they are often distinct from the actual processes and procedures that one might wish to see in an educational context. The situation resonates with a similar instance in cognitive testing: Although Analogies tasks were demonstrable predictors of first year college GPA, the College Board replaced these in the SAT® with essays, because essays, not analogies, are tasks students actually must complete successfully while at college.

SITUATIONAL JUDGMENT TEST

A Situational Judgment Test (SJT) is a type of test where test-takers are presented with variegated situations, each with several possible responses that must be evaluated (see Table 6 for an example). SJTs represent fairly simple, economical simulations of relevant school-, home-, or job-related tasks. This methodology is suitable for virtually any noncognitive skill (Lipnevich et al., 2013; MacCann & Roberts, 2008; Wang, MacCann, Zhuang, Liu, & Roberts, 2009). SJTs can be text-based or presented through multimedia, and responses can be multiple choice (i.e., pick the best response) constructed response (i.e., provide a response to this situation), or ratings (i.e., rate each response for its effectiveness on a Likert-type scale) (Lievens & Coestsier, 2002; Lievens & Sackett, 2006).

Table 6: A situational judgment test item measuring cooperation at the facet level, and Agreeableness at the Big Five factor level

You are part of a group assignment that your teacher has set on [any potential school topic]. As you are all dividing up the workload, it becomes clear that both you and another student are interested in the same topic. Your colleague has received good marks on this topic in the past, but you have been extremely excited about working on this part of the project for several months, even though it is new to you.

What would be the best response? (The participant is also asked what is the worst response immediately after this question, which is entered into the scoring model.)

(a) Toss a coin to determine who gets to work on that particular aspect of the project.

(b) Insist that, for the good of the group, you should work on that aspect of the project because your interest in the area means you will do a particularly good job.

(c) Insist that, for the good of the group, you should work on that aspect of the project because you are able to give it more time.

(d) Ignore your own desires and allow the other person to work on that aspect of the project.

(e) Choose a different group member to work on that aspect of the project so that no one person is privileged over another.

SJTs have several advantages over the traditional self-assessment instruments more commonly used to measure the noncognitive skills described throughout this paper. First, SJTs may be developed to reflect both general and more subtle and complex judgment processes than are possible with conventional tests. Second, SJTs appear to be associated with a less adverse impact on ethnic minorities, which may be of relevance to subgroup differences in any population under consideration. Third, SJTs can be re-purposed as formative assessments, so as to provide a student with feedback on his or her competencies in the domain of interest. Fourth, SJTs appear to be less susceptible to faking, compared to self-assessments, where the improvement due to incentives can be up to a full standard deviation. Fifth, SJTs have the advantage of face validity: participants in these surveys can quickly and easily see what they are measuring, and it makes intuitive sense to them that being able to answer these type of questions would correspond to the skills being assessed. Last but not least, students report them as engaging and worth completing (Lipnevich et al., 2013), which better supports multiple administrations and retains student "buy-in" to the ongoing process of noncognitive assessment.

BIOGRAPHICAL DATA

Biographical data (also known as biodata) is another approach. In this paradigm, individuals are asked standardized questions about their past behaviors, activities, or experiences (e.g., "During the past month, how many times have you been involved in group projects?). Respondents are offered multiple-choice answer options or are requested to answer questions in an open format (e.g., state frequency of behaviors). Biodata may offer a less fakeable method of assessment than standard self-assessment scales, as there are several test characteristics that can be implemented to minimize faking (Schmitt, Oswald, Kim, Gillespie, & Ramsay, 2003). These include asking individuals to elaborate on the biodata details (e.g., "What was the name of the last group project you did?") or combining results obtained with alternative measurement approaches (e.g., self-assessments). It is noteworthy also that the biodata approach has variously survived legal challenges in high stakes (i.e., selection) contexts (http://www.state.nj.us/csc/msb/decisions04/2004September/pdf/FAguannoEtAl.pdf).

OTHERS-RATINGS

This is an assessment type in which others (e.g., teachers, supervisors, trainers, colleagues, friends, faculty advisors, coaches, etc.) rate individuals on various noncognitive skills. This method has a long history, and numerous studies have been conducted that employed this methodology to gather information. Others-ratings have an advantage over self-assessments, in that they preclude socially desirable responding, although they are prone to rating biases. Self- and other-ratings do not always converge, but other-ratings have been demonstrated often to be effective in predicting a range of educational outcomes (Kenny, 1994; Wagerman & Funder, 2006).

Letters of recommendation represent a specific form of other-ratings and have been extensively used in a broad range of academic and workplace contexts (Arvey, 1979). Letters of recommendation provide stakeholders with detailed information about the applicants' past performance, with the writers' opinion about the applicant being presented in the form of an essay. One major drawback of letters of recommendation is that they are not in standardized format: Different letter-writers may include or exclude qualitatively different types of information, such that it is difficult to judge one letter against another. A standardized format for such letters can be developed to prompt faculty members to respond to specific items using a Likert scale, in addition to eliciting open-ended comments (Kyllonen, 2008). Developing a standardized format for the primary and secondary education community, using constructs thought highly relevant, would seem feasible, and could indeed supplement the information provided by self-assessments that the individual might take.

TOWARDS A COMPREHENSIVE NONCOGNITIVE ASSESSMENT SYSTEM

A recent RAND corporation research report found that assessments can influence and improve teaching and student outcomes when certain conditions are met, and though this study was addressing cognitive testing, it is likely the findings would be pertinent to the noncognitive domain as well. These conditions include "Teachers receive training and support to interpret scores effectively; the test scores 'matter' but important consequences do not follow from test scores alone; they are a part of an integrated assessment system that includes formative and summative components; and they are one component of a broader systemic reform effort." It is also important that the assessments have "face validity"; in other words, that when teachers and administrators view the assessment tool, they can recognize and respect it as an effective technique for its purposes and goals (Faxon-Mills, Hamilton, Rudnick, & Stecher, 2013).

Among the social and emotional learning assessment tools currently available, one of the most highly regarded is the Mission Skills Assessment (MSA) of the Index group, which is currently only available for measuring students (by group, not individually) in grades six, seven, and eight. A (different) RAND Corporation research report, published in Fall 2013, entitled "Measuring 21st Century Competencies," reviewed the tools currently available for that purpose, and singled out the MSA as standing above the others for its ease of use, cost effectiveness, reliability, validity, and resistance to faking (Soland, Hamilton, & Stecher, 2013). It makes use of many of the assessment techniques previously discussed—self-report, bio-data, situation judgment tests, and others-rating—to arrive at measures of the key noncognitive constructs (with various facets of the Big Five represented quite extensively).

Many educators at schools using the MSA are committed to using the tool as a foundation and springboard for improving student outcomes in the skills measured by the MSA. Some of the ways in which they are doing so include aligning their internal assessments—project rubrics, disciplinary sheets, student-led conferences, report cards, and much more—with the MSA noncognitive skills; organizing professional development for faculty and parent education on the skills; and infusing their curriculum with lessons about the skills (Martin & Pullman, 2014). Other schools are intentionally shifting their instructional strategies in ways that they think should improve their student outcomes in low-performing areas, such as a school with poor performance in curiosity and teamwork adapting a project-based learning methodology that emphasizes real-world connections, inquiry driven learning, and group-based collaborative learning.

As such, the MSA represents an early proof of concept regarding the power of a comprehensive noncognitive assessment system, but it is limited in not fully covering important facets of Conscientiousness (currently it focuses almost exclusively on time management and organization), containing far too few situational judgment test items, and not providing students, parents, and teachers with feedback and action plans that might point them directly to educational interventions. A truly impactful noncognitive assessment system would also need to be extended down (into the early grades) and up (into the latter grades) with modern equating techniques, so as to systematically influence policy and practice. Because no comprehensive noncognitive assessment system has yet been translated or localized beyond English speaking students and the American context, it is our hope that a more global and expansive noncognitive assessment system is created in the not too distant future, with scientists, educators, and policymakers all at the table to influence its content, breadth, and reach into the real world.

CONCLUSION

The value of noncognitive skills could not be clearer, whether by the research evidence of hundreds of studies or by the common sense and intuition of nearly every classroom teacher. The Big Five personality factors model, widely recognized and trusted among psychology researchers, and increasingly appreciated by economists, policy makers, and educators, offers a foundation for organizing the enormous list of desirable skills and traits commonly discussed when promoting noncognitive skills. Furthermore, reliable and meaningful assessments now exist.

The rest remains for educators and education decision makers. The excuse that "we can't measure that" no longer exists when it comes to critical noncognitive attributes. Educators already understand the importance of the value, but major gaps persist in awareness of the strategies they might use to translate noncognitive measurement data into action, including techniques for teaching, learning, and assessment. If we are to prepare our current generation of primary and secondary education students for the demands of a vastly more competitive global economy, it is time to make the commitment, choose the framework, embed them in education practice, implement the assessments, and use them to drive improvement of outcomes for our students.

REFERENCES

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, 121, 219-245.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47 (1).

Arvey, R. D. (1979). Unfair discrimination in the employment interview: Legal and psychological aspects. *Psychological Bulletin, 86*, 736-765.

Asia-Pacific Education Research Institutes Network (2014). Integrating Transversal Competencies in Education Policy and Practice (Phase I). *ERI-Net Regional Policy Study Series Vol. 1*.

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences, 40*, 1235-1245.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*, 9-30.

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology, 92*, 410-424.

Burrus, J., Elliott, D., Brenneman, M., Markle, R., Carney, L., Moore, G., Betancourt, A., Jackson, T., Robbins, S., Kyllonen, P. C., & Roberts, R. D. (2013). Toward a comprehensive understanding of student persistence and goal attainment: "Putting and keeping you on track." *Educational Testing Service Research Report No: RR-13-14*. Princeton, NJ: Educational Testing Service.

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century* U.S. New York, NY: The Conference Board. Retrieved from http://www.conference-board.org/pdf_free/BED-06-Workforce.pdf

Cattell, R. B. (1943). The description of personality: The foundations of trait measurement. *Psychological Review, 50*, 559-594.

Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York: World Book.

Cattell, R.B. (1973). *Personality and mood by questionnaire*. San Francisco: Jossey-Bass.

Christal, R. E. (1994). *The Air Force self-description inventory. Final R&D status report*. Armstrong Laboratory, Brooks AFB, Texas.

Costa Jr, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64*, 21-50.

De Raad, B. (2000). *The Big Five Personality Factors: The psycholexical approach to personality*. Ashland, OH: Hogrefe & Huber Publishers.

De Raad, B., & Perugini, M. (Eds). (2002). *Big five assessment*. Ashland, OH: Hogrefe & Huber Publishers.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*, 880.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling: A Multidisciplinary Journal*, 13, 440-484.

Drasgow, F., Stark, S., Chernyshenko, O. S., Nye, C. D., Hulin, C., & White, L. A. (2012). *Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions*. Fort Belvoir, VA: US Army Research Institute for the Behavioral and Social Sciences.

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of After-School Programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology, 45*, 294-309.

Dweck, C. S. (2012). *Mindset: How you can fulfill your potential*. New York: Constable & Robinson Limited.

EFA Global Monitoring Report team (2012). *The education for all global monitoring report*. Paris: UNESCO.

Faxon-Mills, S., Hamilton, L., Rudnick, M., & Stecher, B. M. (2013). New Assessments, Better Instruction? Designing Assessment Systems to Promote Instructional Improvement. Santa Monica, CA: RAND Corporation.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology, 44*, 329-344.

Harzing, A-W. (2006). Response styles in cross-national survey research. *International Journal of Cross-Cultural Management*, 6, 243-266.

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Do forced-choice offer a solution? *Human Performance, 13*, 371-388.

Jackson, J. J., Wood, D., Bogg, T., Walton, K. E., Harms, P. D., & Roberts, B. W. (2010). What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness (BIC). *Journal of Research in Personality, 44*, 501-511.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research, 2*, 102-138.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology, 98*, 875-925.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). Attitude measurement. In D. Albarracin, B. T. Johnson, & M. P. Zanna (Eds.), Handbook of attitudes and attitude change. Mahwah, NJ: Erlbaum.

Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2014). Personality, motivation, and college readiness: A prospectus for assessment and development. *Educational Testing Service Research Report No: RR-13-14*. Princeton, NJ: Educational Testing Service.

Kyllonen, P. C. (2008). *The research behind the ETS Personal Potential Index*. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Products/PPI/10411_PPI_bkgrd_report_RD4.pdf on January 15, 2009.

Kyllonen, P. C., Roberts, R. D., & Stankov, L. (Eds.) (2008). *Extending intelligence: Enhancement and new constructs*. New York: Routledge.

Lievens, F., & Coestsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, 10, 245-257.

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*, 1181- 1188.

Lipnevich, A. A., MacCann, C., & Roberts, R. D. (2013). Assessing noncognitive constructs in education: A review of traditional and innovative approaches. In D. H. Saklofske, C. B. Reynolds, & V. L. Schwean (Eds.), *Oxford handbook of child psychological assessment*. (pp. 750-772). Cambridge, MA: Oxford University Press.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: theory and data. *Emotion, 8*, 540.

MacCann, C., Duckworth, A., & Roberts, R. D. (2009). Identifying the major facets of Conscientiousness in high school students and their relationships with valued educational outcomes. *Learning and Individual Differences, 19*, 451-458.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics. *Journal of Cross-Cultural Psychology, 23*, 498-509.

Martin, J., & Pullman, L. (2014). Mission Skills Assessment User's Guide and Toolkit. Index Group. Retrieved from http://www.indexgroups.org/msa/docs/MSA-Toolkit-Interactive.pdf

McCrae, R. R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: data from 50 cultures. *Journal of Personality and Social Psychology, 88*, 547.

National Research Council. (2012). Education for life and work: developing transferable knowledge and skills in the 21st century. Committee on Defining Deeper Learning and 21st Century Skills. J. W. Pellegrino & M. L. Hilton (Eds), *Board on Testing and Assessment on Science Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academies Press.

Norman, W.T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology, 66*, 574-583.

Oxford Research (2010). Transversal Analysis on the Evolution of Skills Needs in 19 Economic Sectors. *European Community Programme for Employment and Social Solidarity*.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*, 322-338.

Rammstedt, B., & Krebs, D. (2007). Does response scale format affect the answering of personality scales? Assessing the big five dimensions of personality with different response scales in a dependent sample. *European Journal of Psychological Assessment, 23*, 32-38.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*, 3-25.

Roberts, B. W., Walton, K. & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*, 1-25.

Sackett, P. R., & Walmsley, P. T. (2014). Which personality attributes are most important in the workplace? *Perspectives on Psychological Science, 9*, 538-551.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Schmitt, D., Allik, J., McCrae, R. R., & Benet-Martinez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self–description across 56 nations. *Journal of Cross-Cultural Psychology, 38*, 173-212.

Soland, J., Hamilton, L., & Stecher, B. (2013). *Measuring 21st Century Competencies: Guidance for Educators*. Santa Monica, CA: RAND Corporation.

Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., & Ramsay, L. J. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. *Journal of Applied Psychology, 88*, 979-988.

Soto, C. J., & John, O. P. (2009). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality, 43*, 84-90.

Stemler, S. E., & Bebell, D. (2012). *The school mission statement: Values, goals, and identities in American education*. New York: Taylor and Francis.

Torrente, C., Alimchandani, A., & Aber, L. (Forthcoming 2015). International perspectives on social-emotional learning. In J. Durlak, C. Domitrovich, R. Weissberg, & T. Gullotta (Eds.), *Handbook of Social Emotional Learning*. New York: The Guilford Press.

Tough, P. (2013). *How children succeed: Grit, curiosity, and the hidden power of character*. New York: Houghton Mifflin Harcourt.

Tupes, E. C., & Christal, R. C. (1961). Recurrent personality factors based on trait ratings. *USAF ASD Technical Report, 61-97*, U.S. Air Force, Lackland Air Force Base, Texas.

Wagerman, S. A., & Funder, D. C. (2006). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41*, 221-229.

Wagner, T. (2010). *The Global Achievement Gap: Why even our best schools don't teach the new survival skills our children need–and what we can do about it*. New York: Basic Books.

Wang, L., MacCann, C., Zhuang, X., Liu, L., & Roberts, R. D. (2009). Assessing teamwork and collaboration in high school students: A multimethod approach. *Canadian Journal of School Psychology*, 24, 108-124.

Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods, 7*, 168-190.

Ziegler, M., MacCann, C., & Roberts, R. D. (Eds.) (2011). *New perspectives on faking in personality assessment*. New York: Oxford University Press.

APPENDIX: ADDITIONAL INFORMATION ON ASSESSMENT APPROACHES

A.1 SELF-REPORTS

Many issues must be taken into account when developing a psychometrically sound self-report assessment, and there is a large literature on a wide variety of such topics. The optimal number of points on a scale, scale point labels, the inclusion of a neutral point, alternative ordering, and other test characteristics have been widely analyzed and examined (Krosnick, Judd, & Wittenbrink, 2005). For instance, studies reveal that response scale format influences individuals' responses (Rammstedt & Krebs, 2007), while the inclusion of negatively-keyed questions (to avoid acquiescence or "yay-saying") is considered controversial (DiStefano & Motl, 2006). Respondents vary in their use of the scale—for example, young males tend to use extreme answer categories (Austin, Deary, & Egan, 2006), as do Hispanics (Marin, Gamba, & Marin, 1992), and in general, there are demonstrable cultural effects in response style (Harzing, 2006).

A.2 FORCED-CHOICE ASSESSMENT

Of all of these methods, forced-choice measurement appears the most promising in reducing fakeability of self-reports. The procedure has many different aspects, including pair comparisons, rank-ordering, and multidimensional forced-choice. In pair comparisons, the test-taker must choose between two statements (e.g., which is more like you: "I work hard" or "I enjoy working in teams"?). In rank-ordering, test-takers must rank a series of equally desirable statements in order from "most like me" to "least like me." Both of these methods require that the paired statements included in any one item are carefully matched for social desirability, so that test-takers cannot use the evaluative aspects of the statements in their responses. In multi-dimensional forced-choice assessments, test-takers are presented with a dichotomous quartet of four different traits, in which two socially desirable statements are paired with two socially undesirable statements (Jackson, Wroblewski, & Ashton, 2000). For example, a test-taker would be asked to select which is "most like you" and which is "least like you" from the following four statements: (a) "I work hard," (b) "I lose my temper," (c) "I love to help others," and (d) "I cannot deal with change." The statement selected as "most like you" would be scored "+1," the statement selected as "least like you" would be scored "-1," and the statements that were not endorsed at all would be scored zero.

There is compelling evidence to suggest that forced-choice tests show stronger relationships with performance outcomes and are less fakeable than standard rating scales (Drasgow et al., 2012; Jackson et al., 2000).Forced-choice measures, however, have what is known as ipsative properties. That is, scores on forced-choice measures may be appropriate for comparing the relative level of different traits within an individual, but they are inappropriate for comparing the relative level of a trait across different people. Essentially, personality dimensions lose their independence when measured this way: one cannot be high on them all. This poses a problem for test-takers who really are high on multiple personality dimensions, or for test users who want to select individuals based on high (or low) scores on one or more personality dimensions.

Professional
Examination Service