K-12 Center at ETS

Invitational Research Symposium on
Technology Enhanced Assessments

Measurement of 21st Century Skills Within the Common Core State Standards

Patrick C. Kyllonen

May 7–8, 2012

Measurement of 21st Century Skills Within the Common Core State Standards

Patrick C. Kyllonen
Educational Testing Service

For the past 10 years or so, Educational Testing Service has supported the Center for New Constructs, which has had as its mission the identification and evaluation of constructs and measures other than those typically measured by content tests, such as mathematics, English language arts, and science achievement tests. Since then the center has been renamed the Center for Academic and Workforce Readiness and Success (CAWRS) to highlight ETS's recent foray into the workplace, but the work of the center has remained the same—to explore the foundations, rationale, and feasibility of measuring personal characteristics that are important for educational success but that are typically ignored in educational assessment. During this period CAWRS staff have explored this area through literature reviews, collaborations, and data collections ranging from a few students to more than 100,000, including some in K-12 education settings, in community college and college settings, in graduate and professional school, and more recently, in the workplace. CAWRS staff have explored a wide array of constructs, ranging from ones that overlap considerably with what is already tested on standardized tests, such as critical thinking, to ones that are quite distinct, such as teamwork and work ethic. CAWRS also has explored a wide variety of measurement methods, such as self-rating and others' rating scales, situational judgment tests, anchoring vignettes, conditional reasoning, the implicit association test, and others. The purpose of this paper is to review some of this work, particularly as it pertains to the measurement of 21st century skills in K-12 education. I will touch on other areas, but that is the focus. This paper begins with a brief history and justification for why there is interest especially now in 21st century skills, and what is meant by that term.

In the 1990s the discussion in education concerned the growing gap in the college-high school earnings differential and there was a strong sentiment at the time that this was due to increasingly complex world where the value of cognitive skills grew greater and greater (Murnane, Willett, & Levy, 1995). This was a time of editorials in the *Wall Street Journal* on "the math gap" (Murnane, Willett, Duhaldeborde, & Tyler, 2000) and the time of Hernstein and Murray's (1994) controversial best seller, *The Bell Curve*, which argued in the spirit of Jensen (1998) that cognitive skills were more important than socioeconomic status and educational attainment for all varieties of life outcomes ranging from wages and unemployment, to staying out of jail and avoiding teenage pregnancy, and that growing inequality could be attributed directly to a cognitive ability gap between the cognitive haves and have nots.

Meanwhile, others were noting that the data did not suggest that cognitive skill differences per se were responsible for growing income inequality (Howell, 2002) and that our faith in standards based education, might not be justified (Levin, 1998). It clearly has been the case that educational attainment per se is important. There is a reliable decrease in the likelihood of unemployment and a concomitant increase in earnings expected as one goes up the educational attainment ladder, a relationship that has held up over time (Levine, 2004). But at the same time it has been known that schooling's contribution to economic success is only partly explained by the cognitive skills acquired in school (Bowles & Gintis, 1976; Bowles, Gintis, & Osborne, 2001).

Consequently, for the past 10 years or so, there has been a growing appreciation for the importance of skills other than the cognitive skills typically tested by standardized tests. In psychology and education, for example, there have been several meta-analyses showing the importance of personality, attitudes, and study skills on academic achievement. A meta-analysis of the correlations between personality factors and academic achievement (grades) found that in K-12 personality factors, typically measured by self- or teacher ratings, were as predictive of grades as IQ scores were (Poropat, 2009). Similar meta-analytic studies that have focused on higher education outcomes (Crede & Kuncel, 2008; Noftle & Robins, 2007; Robbins et al., 2004) and workforce outcomes (e.g., Dudley, Orvis, Lebiecki, & Cortina, 2006) have shown similar findings.

In economics, there have been studies that show that the correlation with noncognitive factors does not stop at school, but continues through to one's transition to the workforce. For example, Segal's (in press) dissertation, based on data from the National Educational Longitudinal Study (NELS:88) showed that misbehavior in 8th grade, based on teacher "yes-no-don't know" ratings on five items (whether the student rarely completes homework, is consistently inattentive, or frequently absent, tardy, or disruptive) predicted workforce earnings 20 years later, controlling for educational attainment and achievement test scores. A study of over 14,000 military enlistees in Sweden (Lindqvist & Vestman, 2011) found that noncognitive factors measured in a 30 minute clinical interview at age 18, including responsibility, independence, outgoingness, persistence, emotional stability, and initiative, were at least as important in predicting employment and earnings, 20 years later, as were cognitive skills, measured by a 2-hour standardized test, and were much more important in predicting chronic unemployment.

The studies cited above show clearly that noncognitive skills are important predictors of educational and workforce outcomes. But a question is whether such skills can be deliberately developed in school. Two recent meta-analyses of social-emotional learning interventions suggest that they can. Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011) conducted a meta-analysis of 213 school-based, social and emotional learning (SEL) programs involving over 270,000 K-12 students. Students participating in the SEL interventions compared to controls were shown to have demonstrated significantly improved social and emotional skills, attitudes, and behavior. In addition, participating students experienced significant achievement gains. In these programs, administration was by school teaching staff. The study also identified several effective practices, and ways to avoid implementation problems, which they characterized with the acronym SAFE. Effective programs were ones emphasizing *sustained or sequential* activities rather than unstructured one-off opportunities; relied on *active* student involvement, *focused* on personal and social skills by setting aside time specifically to develop those skills, and *explicitly* identified which skills they developed. Durlak et al. (2011) also suggested that the evidence base was now sufficient to justify the incorporation of socio-emotional learning programs

in school. A second meta-analysis, conducted a few years earlier on after-school programs (Durlak & Weissberg, 2007), had produced very similar findings. Noncognitive skills are important and can be developed in school.

A Note on Terminology

This article uses the terms cognitive skills, noncognitive skills, 21st century skills, socio-emotional learning, interpersonal skills, and intrapersonal skills. *Cognitive skills* have come to mean those skills developed through schooling, usually in content areas such as mathematics and English language arts, that are readily measured with standardized tests, such as those used in accountability testing, in large scale domestic and international assessments, such as the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA), or in admissions (e.g., the SAT® and ACT), or selection (the Armed Services Vocational Aptitude Battery).

Noncognitive skills have come to mean all other skills developed through schooling that are not reflected in cognitive test scores. The nature of these other skills is to be determined, and that is part of the focus of this paper. Sometimes noncognitive is confused with personality, but noncognitive is intended to be a broader term, with personality perhaps being a part of noncognitive skills.

The concept of *21st century skills* does not have a precise definition, but is intended to convey the idea that changes in technology and culture are leading to changing demands in the workplace, and so the skills that are required in today's and the future workplace are different from those required in the past (Autor, Levy, & Murnane; 2003; Levy & Murnane, 2004). One way in which the demands will be different is that there may be a greater emphasis on interpersonal (social, communicative) and intrapersonal (self-regulatory) skills, as well as broader cognitive skills (National Research Council [NRC], 2011). If the requirements of the 21st century workplace are changing, there may be increased pressures on the educational system to produce the skills that are emerging in importance. Social-emotional learning (SEL), which has come to stand for in- and out-of-school programs that develop "core competencies to recognize and manage emotions, set and achieve positive goals, appreciate the perspectives of others, establish and maintain positive relationships, make responsible decisions, and handle interpersonal situations constructively" (Durlak et al., 2011) is one response to this demand. But there may be additional responses in that the *21st century skills* concept encompasses a broad spectrum of skills, perspectives, capabilities, and competencies that undoubtedly will evolve further with continued changes in technology and culture. We now turn to an analysis of what various groups have said about the importance of these so-called 21st century skills.

What Are the 21st Century Skills?

The Conference Board and others sponsored a survey of more than 400 U.S. employers across a range of sectors, which included items asking employers how important various "applied" and "content" skills were for incoming employees, either those entering with a high school or a 2-year or 4-year college degree (Casner-Lotto & Benner, 2006). The results (for high school and 4-year college) are shown in Table 1. There are two useful conclusions to be drawn from these findings. First, some of what they called *applied skills*, or what might be referred to here as 21st century skills—specifically, oral and written communication, teamwork and collaboration, professionalism and work ethic, and critical thinking and problem solving—were rated "very important" for college graduates by over 90% of the employers. This is a higher percentage than the "very important" percentages received by the content

skills—English language (88%), mathematics (64%), and science (33%). Second, this pattern was true for both college and high school graduates, although for high school graduates, the percentages were a bit lower across the board, and, interestingly, the skill receiving the highest number of "very important" ratings was work ethic.

Table 1. Percentage of Employers Rating Skill as "Very Important"

	2-year college graduates	High school graduates
Oral communications	95%	70%
Teamwork/collaboration	94	75
Professionalism/work ethic	94	80
Written communications	93	53
Critical thinking/problem solving	92	58
English language	**88**	**62**
Ethics/social responsibility	86	63
Leadership	82	29
Information technology	81	53
Creativity/innovation	81	36
Lifelong learning/self-direction	78	43
Diversity	72	52
Mathematics	**64**	**30**
Science	**33**	**9**

SOURCE: Data from Casner-Lotto & Benner (2006).
Note. Content skills in **boldface**.

Other surveys have shown similar findings. For example, at an National Academy of Sciences (2011) workshop, Boisvert reported on a survey conducted by the Boston Advanced Technological Educational Connection (BATEC) in 2008 that asked employers what skills were most important in the workforce. The survey identified communication skills (oral and written), ability to work productively in teams and groups (teamwork skills), customer and business focus (understanding the big picture), ability to listen for meaning and comprehension, ability to prioritize work and self evaluate (self-reflection and time management), development of original solutions to novel problems (problem solving), and ability to lead and act responsibly (leadership and ethics). McKinsey Global Institute recently conducted a survey of 2,000 employers in an attempt to find out what the reasons were for difficulties in finding qualified applicants (Manyika, Lund, Auguste, Mendonca, Welsh, & Ramaswamy, 2011). Among the top reasons were "work habits are not suitable" (33%), "insufficient problem solving skills" (19%), and "poor communication ability," all of which might be considered 21st century skills. Interestingly, in that same survey, only 6% of the employers cited "poor reading or math" as a reason for workers failing to be qualified for the job.

A Common Taxonomy of 21st Century Skills[1]

Several organizations have made attempts to assemble lists of 21st century skills, through literature reviews, the consideration of survey results such as those just reviewed, expert panels, and the like. The Partnership for 21st Century Skills (P21, 2002), founded in 2002 by Ken Kay and Diny Golder-Dardi with support from AOL, Cisco, Microsoft, and the U.S. Department of Education, among others was one of the first of these groups. It has had as its mission "to position 21st century readiness at the center of US K12 education by building collaborative partnerships among education, business, community, and government leaders." P21, as it is called, proposed a set of 21st century student outcomes, including both core subjects (the 3 R's); and new ones, including learning and innovation skills (the 4 C's: critical thinking, communication, collaboration, creativity); information, media, and technology skills; and life and career skills. P21 also proposed that these 21st century student outcomes could affect standards and assessment, curriculum and instruction, professional development, and learning environments.

Other groups and organizations have proposed similar taxonomies. The Assessing and Teaching of 21st Century Skills (ATC21S) organization, sponsored by Cisco, Intel, and Microsoft, commissioned a white paper to define 21st century skills (Binkley, Erstad, Herman, Raizen, & Ripley, 2010), which used an expert group to propose a taxonomy organized as follows:

- Ways of Thinking (creativity and innovation; critical thinking, problem solving, and decision making; learning to learn and metacognition)
- Ways of Working (communication; collaboration and teamwork)
- Tools for Working (information literacy; information technology and communication literacy)
- Living in the World (life and career; personal and social responsibility)

The organization also commissioned additional white papers, on methodological issues (Wilson et al., 2012); technology issues (Csapó, Ainley, Bennett, Latour, & Law, 2010); learning environments (Scardamalia, Bransford, Kozma, & Quellmalz, 2010); and policy frameworks (Darling-Hammond, 2011).

A Dutch public educational organization, Kinnesnet, which supports Dutch primary, secondary, and vocational institutions in use of ICT, conducted an analysis of the extant theoretical frameworks for 21st century skills, focusing on definition, implementation, and assessment. Kinnesnet was particularly interested in noting the diversity of definitions and methods from an international perspective. Their approach (Voogt & Pareja Roblin, 2010) was to review 59 documents on 21st century skills, from the organizations just mentioned (P21 and ATC21S) and others, including Engage, NETS, the U.S. Department of Education's National Assessment of Educational Progress (NAEP), the European Union, OESO, and UNESCO. Their identification of 21st century skills was designed to inform what it is that students should learn, pedagogical approaches, and learning outcomes.

A slightly different approach was taken by Finegold and Notabartolo (2008), in an interdisciplinary literature review sponsored by the Hewlett foundation. Their work, relative to the

[1] See Dede (2010) for a complementary effort in synthesizing approaches and results for identifying 21st century skills.

others, was more influenced by a focus on requirements of the future workplace. In the end, they developed a skills taxonomy similar to the others, and they made several helpful suggestions for future research in this area. This included a broad charge to identify the relationship between 21st century competencies and outcomes and to explore the mechanisms underlying the development of 21st century skills. They also suggested focusing on the "demand side" (i.e., employment requirements), recognizing domain specificity in areas such as critical thinking, understanding that some competencies such as problem solving may simply reflect more advanced levels of competence rather than be separate competencies. They also suggested that attention be given to new competencies, such as cross-cultural fluency, systems thinking, and financial literacy. They pointed out that the United States was lagging behind other countries in recognizing the importance of these skills, particularly the United Kingdom, Australia, and New Zealand.

In addition to the expert group studies, and the interviews, there have been related attempts to identify the importance of new skills, or 21st century skills, in high school, and in higher education. For example, in a couple of studies sponsored by the College Board, researchers at Michigan State University (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Schmitt et al., 2007) inspected mission statements from 35 universities, treating them as evidence for the kinds of skills colleges purport to develop in their students. They sorted statements, and through that process identified the 12 most popular themes:

- Knowledge, learning, mastery of general principles
- Continuous learning, intellectual interest, and curiosity
- Artistic cultural appreciation and curiosity
- Multicultural tolerance and appreciation
- Leadership
- Interpersonal skills
- Social responsibility, citizenship, and involvement
- Physical and psychological health
- Career orientation
- Adaptability and life skills
- Perseverance
- Ethics and integrity

A similar study of high school mission statements was recently completed by Stemler, Bebell, and Sonnabend (2011).

A related methodology was employed by Shultz and Zedeck (2011) to examine lawyering effectiveness, although their method was more dependent on interviews with law firms and partners. They developed the following list:

- Intellectual and cognitive (analysis and reasoning; creativity/innovation; problem solving; practical judgment)
- Research and information gathering (researching the law; fact finding; questioning and interviewing)
- Communications (influencing and advocating; speaking; writing; listening)

- Planning and organizing (strategic planning; organizing and managing self and others)
- Conflict resolution (negotiation skills; seeing the world through others' eyes)
- Client and business relations (networking and business development; advice/counseling; relationships with clients)
- Working with others (developing relationships with the legal profession; evaluating, development, mentoring)
- Character (community involvement, integrity/ honesty, stress management, passion, diligence, self-development)

Although the studies just reviewed relied on a wide variety of methodologies, and addressed groups ranging from primary school students to college students, lawyers to various entry level and advanced workforce entrants, there is considerable overlap in the conclusions regarding the nature of 21st century skills. There is overlap among lists of 21st century skills. Clearly there is a great deal of consensus on the skills that are likely to emerge as most important in school and in the workplace.

The NRC and National Academy of Sciences (NAS) has conducted several studies of 21st century skills over the past several years, beginning in 2005 with workshops convened to discuss the role of K-12 education in developing 21st century skills (NRC, 2008), following up with workshops on the role of 21st century skills in science education (NRC, 2010), one focusing specifically on assessment issues (NRC, 2011), and then a recent effort on deeper learning and 21st century skills (NRC, 2012). Over the course of the workshops, a tripartite organization of 21st century skills has proven to serve as a useful organizational framework. Adopting that organization and inserting that into the components identified here, we get something like the following:

- Cognitive skills—critical thinking, problem solving, creativity
- Interpersonal skills—communication skills, social skills, teamwork, cultural sensitivity, dealing with adversity
- Intrapersonal skills—self management, self regulation, time management, self development (lifelong learning), adaptability, executive functioning

Of course, there is overlap among these categories. For example, communication skills have both cognitive and interpersonal components, and other skills similarly could be cross classified. But this is a useful framework for helping think about assessment and other issues.

Common Core State Standards and 21st Century Skills

Two recent developments in American education are having a major effect on educational policy, curriculum, assessment, and accountability in the United States. One is the adoption by many states of the common core state standards for mathematics and English language arts, along with the NRC'S K-12 science standards, and the other is the development of two multistate consortia—the SMARTER Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers (PARCC) —in response to Obama's race to the top (RTT) legislative initiative. These two activities are related in that a focus of RTT efforts will be on curricular development, teacher professional development, and accountability and assessment efforts that support the common core state standards. A recent study by the NRC (2012) compared the standards documents to identify overlaps and uniquenesses between standards and 21st century skills, defined along the lines of those

described here. They found substantial evidence for overlap in the cognitive domain, involving for example, critical thinking and non-routine problem solving, but less overlap in the interpersonal and intrapersonal domains.

Presentations by leaders of the two consortia at a recent meeting at the University of Southern California's for Enrollment Research, Policy, and Practice 2012 conference, 21st Century Knowledge and Skills: The New High School Curriculum and the Future of Assessment, came to a similar conclusion. There were some overlaps between the skills targeted by the consortia and 21st century skills. For example Jones (2012), representing the PARCC group, pointed to the 21st century nature of standards such as "communicating research findings" with respect to speaking and listening skills, and "making arguments and critiquing the arguments of others" as examples of higher order critical thinking skills, but there were fewer examples in the interpersonal and intrapersonal areas, although "persevering in solving problems" was cited as an exception. Mitchell (2012), representing the SBAC group, cited "Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others," and "Students can employ effective speaking and listening skills for a range of purposes and audiences," as examples of cognitive and interpersonal category overlaps but these examples appeared to be exceptions.

It may be that the importance of 21st century skills in the United States anyway, will emerge only in the next generation of the common core state standards. Some signs of this potential future interest include an Illinois learning standard called "Social/Emotional Learning" and legislation in the U.S. House of Representatives, HR2437: Academic, social, and emotional learning act, sponsored by Judy Biggert. Other signs of changes that may emerge is the inclusion in the next cycle of the Programme for International Student Assessment (PISA) of a collaborative problem solving task. PISA is a large-scale international comparative assessment of 15-year olds from over 60 countries, which has traditionally focused on comparisons in mathematics, reading, and science achievement. The inclusion of a collaborative problem-solving measure in PISA 2015 may be a signal of the interest around the world in skills that go beyond those measured by traditional standardized tests in the mathematics, reading, and science content areas.

Measurement Issues

Assessment of 21st century skills may require different approaches from those that have dominated assessment systems until now. For example, multiple-choice; short, constructed-response; or essay tests may not prove sufficient for measuring many of the 21st century skills, such as the interpersonal skills of teamwork, collaboration, leadership, and communication, or some of the hard-to-measure cognitive skills, such as creativity, or some of the intrapersonal skills, such as self-regulation, time management, and adaptability. It may be that some of the reluctance apparent in the failure of current standards efforts to embrace 21st century skills education goals reflects a distrust in the capability of current educational measurement science to measure such skills in a reliable and valid way. And it may be that future adoption will depend on the field's ability to develop creative and innovative solutions to the problem of measuring 21st century skills.

In this section I review the current state of the art in measuring 21st century skills. Ratings—both self-ratings, and ratings by others—are probably the most popular ways to measure the interpersonal and intrapersonal skills. But there are techniques designed to improve the validity of these

ratings, through behavioral anchors, and other means. In addition there are an increasing number of performance measures of interpersonal and intrapersonal skills. This is likely to be a growing area of innovation and development as we learn to use advances in technology to capture interpersonal and intrapersonal skills.

Self-Ratings

Self-ratings are used widely in educational research and applications. Uses include program evaluation, background questionnaires on student, teacher, and school surveys, such as NAEP and PISA, self-improvement, and engagement surveys, such as the High School Survey of Student Engagement (HSSSE; Yazzie-Mintz, 2010). For example, consider the evaluations of the effectiveness of social emotional learning (SEL) programs reported on in Durlak et al. (2011). They included various measures of social and emotional skills, attitudes, positive social behavior, conduct problem, and reducing emotional stress. Self ratings can be found for any 21st century skills construct.

Consider some example items from such surveys. The HSSSE asks students how much time they spend on various activities, such as working for pay, surfing the web, or doing volunteer work, and how important such activities are to them, allowing them to choose "not at all," "a little," "somewhat important," "very important," or "top priority." PISA similarly uses Likert scales to ask students to rate themselves on factors such as their motivation and engagement levels, whether they attribute their successes to effort or other factors, how they use their time, and other many other factors. There is growing interest in having NAEP used more extensively than it has been used in the past to measure student factors such as these (Smith et al., 2012).

There are many other studies that measure student noncognitive factors for purposes of monitoring the development of student skill. For example, in a CAWRS project (Burrus & Roberts, 2012; see also, Burrus, Roberts, Brenneman, & Lipnevich, 2012) we administered a wide ranging battery to more than 2,000 6th to 8th graders from the Elementary Schools Research Collaborative (ESRC), a group of private schools across the United States. ESRC members are interested in comparing members on their students' noncognitive skills. The battery included measures of the following factors (with example items in parentheses; items include a Likert scale response format):

- cooperation ("I like to work with people")
- leadership ("I like leading groups")
- on time ("I complete my language arts homework on time")
- planning ("I like to make lists of things to do for school")
- anxiety ("I give up easily")
- emotion stability ("I remain calm under pressure")
- openness ("I enjoy homework")
- sympathy ("I think it is important to help people")
- math attitude ("I am determined to become good at math")
- creativity ("I think outside the box")
- teamwork negative emotion
- teamwork positive emotion
- avoidance coping ("I get some sleep")

- task coping ("I take control and say to myself: I can do this!")
- emotion coping ("I wish that I could change what was happening or how I feel")

In addition, teachers also rated students on factors including *being on time* ("She is organized with her schoolwork"), *cooperation* ("She takes time to help others"), *emotional stability* ("She overcomes challenges and setbacks"), and *creativity* ("She has a good imagination").

Behaviorally anchored rating scales. A problem with ratings is that the commonly used Likert scale may mean different things to different people. Labels indicating evaluation or agreement such as "highly characteristic of me," or "uncharacteristic of me," "strongly agree" or "disagree," or labels indicating frequency, such as "often," "seldom," or "rarely" are all subjective labels and are consequently open to wide ranging interpretations by different people, making the results difficult to interpret. A clear example can be seen in PISA where some of the countries with the lowest proficiency scores tend to show the highest positive attitudes towards the subject matter and have the highest self concepts about their abilities. One way to avoid some of the subjectivity of rating scales is through what are known as behaviorally anchored rating scales (BARS).

A BARS is a scale, like a Likert scale, but with behavioral descriptions at various locations along the scale. An example, taken from a study of lawyering effectiveness (Shultz & Zedeck, 2011), is shown in Figure 1. This is a 6 point scale (0 to 5), which provides examples of lawyering effectiveness ranging from very good ("Extracts the essence of complex issues and doctrines quickly and accurately") to very poor ("Over-simplifies arguments, misses possible sub-issues and nuances, and fails to anticipate the opposing side's points"). A colleague or supervisor (or teacher) could rate a student, or a lawyer for that matter, on these kinds of scales. Similar kinds of scales have been developed for students in business school (e.g., see Hedlund, Wilt, Nebel, Ashford, & Sternberg, 2006) and college (e.g., Oswald et al., 2004; Schmitt, Oswald, Kim, et al., 2007; Sternberg et al., 2000).

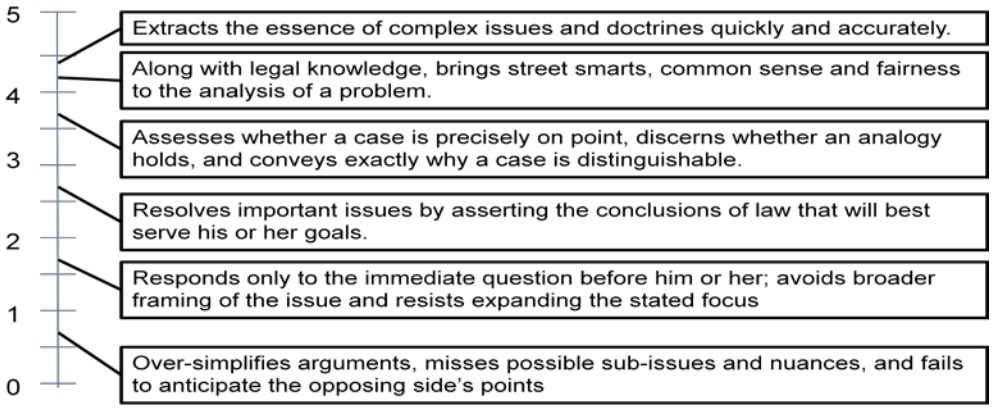

Figure 1. Behaviorally Anchored Rating Scale (BARS) for "Analysis and Reasoning" factor for lawyering effectiveness (Shultz & Zedeck, 2011).

BARS scales are related to *scale anchoring*, a method used in large-scale assessments to describe a score by relating it to what students achieving that score can do, such as what problem most of them can solve (e.g., Beaton et al., 2011). However BARS are used for rating purposes. BARS development typically begins with a critical incident analysis (Flanagan, 1949), in which students or job

incumbents are interviewed and are asked to identify examples where the construct to be measured, such as teamwork, or perseverance, is on display. These are often given in the form of problem-plus-solution description. These incidents are then rated by a group of incumbents, or students, or experts, on the target scale (e.g., 0 to 5) and some summary of those ratings (e.g., the average) is placed on the scale. The incidents then serve as the anchors for the BARS, which then can be used as a rating scale.

Anchoring vignettes. Anchoring vignettes are a relatively new method for improving the validity of ratings. An anchoring vignette item asks respondents to rate several hypothetical situations or persons on the construct of interest, then it uses those ratings as anchors around which to place a respondent's actual ratings (Hopkins & King, 2010). For example, a respondent might be told to rate three people who vary on a construct such as teamwork, then to rate themselves. The technique concerns recoding one's own rating to the scale established from one's ratings of the other three people. We have recently used this method in the analysis of data from the PISA 2012 questionnaire, and preliminary results suggest that it increases the validity of the questionnaire quite substantially, essentially solving the subjective rating scale problem mentioned above.

Forced choice. Another method designed to increase the comparability of ratings across people is the forced choice or preference method. This is a technique in which instead of asking a student to rate on a 5-point Likert scale how much they like mathematics, for example, the student is asked which they like more, mathematics or science (or English language arts). Or, whether they characterize their own classroom as one that more encourages cooperation or competition, for example. Forced choice methods were also experimented with in the PISA 2012 field trial, and they worked quite successfully. Forced choice methods have been used for a long time, but there has been recent psychometric research that has shown that they can be applied more broadly than had been the case previously (Brown & Maydeu-Olivares, 2011; Stark, Chernyshenko, & Drasgow, 2004).

Others' Ratings

Probably just about any construct that can be measured through self ratings can also be measured through ratings by others. Certainly, a rater does not have access to the personal experiences and innermost thoughts and feelings of a ratee, but the rater may have a better and less biased sense of where the ratee stacks up on a particular dimension than a ratee does. Research suggests in fact that others' assessments of a wide variety of personal factors are more accurate, less biased, and more predictive of future outcomes than self ratings are (Connelly & Ones, 2010). The better one knows another, the more accurate the ratings are, of course, but improvements over self ratings can be obtained even with observers who do not know the subject all that well. For example, the Swedish enlistment study cited above (Lindqvist & Vestman, 2011) relied on ratings by a trained professional who only knew the subject for 30 minutes, and yet the predictiveness of the ratings for labor market outcomes 20 years later was substantially higher than is typically the case with self ratings, which is a major claim made by the authors of that report.

With younger grades, teacher ratings are probably especially critical as individual self reports are likely to be unreliable at that age, but teacher ratings are useful across the grade spectrum. In NELS:88 (Ingels, et al., 1990), for example, teachers rated 8th grade students on items such as whether the student "performs below ability," "rarely completes homework," "is frequently ..." "absent," " tardy," "disruptive," "inattentive in class," and "is excessively passive or withdrawn," (yes, no, don't know). As discussed above,

some of these items were the basis for the finding that teacher ratings of misbehavior predicted labor market outcomes years later (Segal, in press).

PISA includes a parent questionnaire, NAEP includes a teacher questionnaire, and other large-scale assessments include one or the other or both. Providing multiple perspectives is useful in questionnaires because parent, teacher, and student perspectives are correlated, but do not completely overlap. We have conducted several studies using self, parent, and teacher ratings (e.g., MacCann, Wang, Matthews, & Roberts, 2010). In all of these we have found that both parent and student (or teacher and student) questionnaires add predictive validity to the other in predicting various student outcomes. This finding is consistent with the meta-analytic findings (Connelly & Ones, 2010) that self and other ratings both add predictive validity to each other.

ETS recently launched an others' rating system, the ETS® Personal Potential Index (ETS® PPI), designed to measure personal skills that have shown to be critical for success in graduate school. The factors measured are (a) knowledge and creativity, (b) communication skills, (c) teamwork, (d) resilience, (e) planning and organization, (f) ethics and integrity, along with an overall evaluation. A report documents the background for the instrument, and provides data on its psychometric properties (Kyllonen, 2008). Details on the system can be found at the PPI website: http://www.ets.org/ppi. Another version of the PPI, called the ETS® Collegiate Personal Potential Index, or ETS® CPPI, is intended to be used by high school students applying for college. Figure 2 presents a screen capture, and Figure 3 presents the scales and items, which are similar to those of the PPI, but are tailored to high school students.

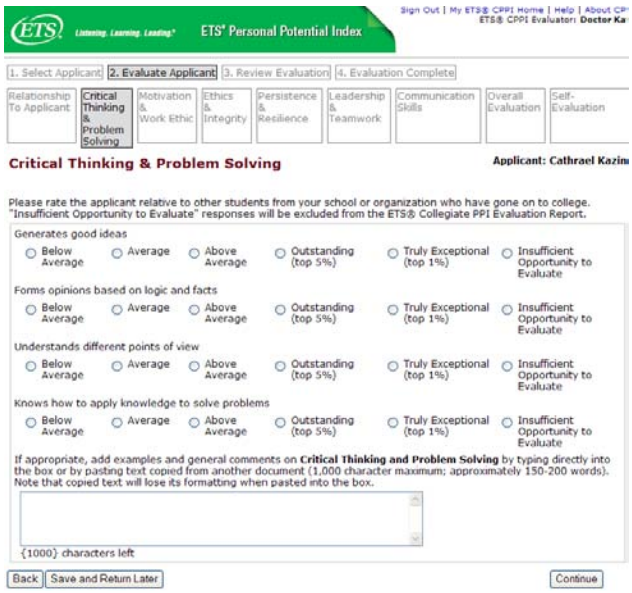

Figure 2. Screen capture of items from the "critical thinking & problem solving" scale of the ETS Collegiate Personal Potential Index (ETS CPPI).

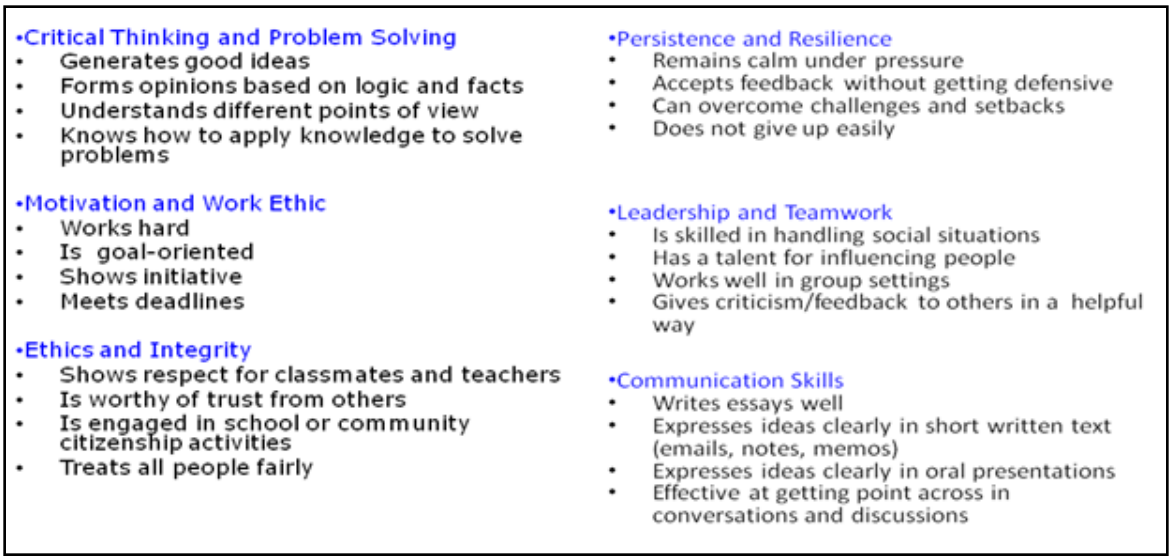

Figure 3. Scales (6) and Items Within Scales (4 Each) for the ETS Collegiate Personal Potential Index (ETS CPPI).

Situational Judgment Tests

Situational judgment tests (SJT) are ones with items that present situations, and ask respondents for the best (or their typical) response to the situation. The responses are most often multiple choice, or Likert scale. An example developed by the ETS CAWRS group in the area of teamwork is as follows (MacCann, Fogarty, Zeidner, & Roberts, 2011; MacCann & Roberts, 2012):

> Your team is writing a business case for creating a new flavor of soda. You have a tight deadline, and everyone is really busy. You have been assigned to write the section of the report that describes the results of taste-test research. You are a new employee, and you are not sure if your section of the report is clear. What would you do? (in some situational judgment tests the question might alternatively be, "What would be the best course of action to take?)
>
> A. Work as hard as you can on your section until the deadline.
>
> B. Write several versions of your section and submit them all to the team.
>
> C. E-mail a draft of your section to the entire team for their comments.
>
> D. Ask an experienced team member for advice on your draft.

There has been considerable research on these kinds of items, which are used in industrial/organizational psychology (Weekly & Ployhart, 2006) and in education (e.g., Sternberg et al., 2000). SJTs are popular assessment methods for various reasons, including (a) they can measure otherwise hard-to-measure constructs, such as teamwork, communication, and other 21st century skills, (b) they are often based on critical incidents studies of student or job incumbent performance, and so

the validity argument is developed in the course of developing the assessment, (c) they blend the measurement of cognitive and noncognitive skills, and (d) although it takes longer to get an acceptably reliable measurement, SJTs typically have relatively high predictive validity, according to meta-analyses on their validities and moderating factors (e.g., McDaniel, Hartman, Whetzel, & Grubb, 2007).

Situational judgment tests have been developed for measuring student 21st century skills in college (e.g., Oswald et al., 2004; Schmitt et al., 2007; Sternberg et al., 2000), business school (Hedlund et al., 2006), and law school and practice (Shultz & Zedeck, 2011). The ETS CAWRS group has developed SJTs in a variety of settings, including for business school, for practicing business analysts (to measure critical thinking), for the military (to measure emotional intelligence), for a medical client (to measure communication skills), for prospective competitive college athletes (to measure noncognitive skills associated with success in sports), and in other settings.

Biodata

Biodata refers to the collection of activities (biographical, or bio activities) that provide evidence for a student's having acquired skill or demonstrated competence or service in some area. Biodata can be thought of as the kind of information that typically appears on a resume, but implied is an additional component of capturing the information in a standardized way, and some way to summarize or score that data so that it can be treated alongside other data on the student such as standardized test scores. There have been attempts over the years to capture documented accomplishments at ETS (Baird & Knapp, 1981; Stricker, Rock, & Bennett, 2001). For example, Stricker et al.'s measure produced scores for Academic Achievement, Leadership, Practical Language, Aesthetic Expression, Science, and Mechanical. A typical item might simply be one that asked whether a student demonstrated leadership by serving on a committee, and then asked for details for verification purposes. In their assessment of college applicants, Oswald et al. (2004) and Schmitt, et al. (2007) included biodata measures designed to capture many of their 21st century skills. A recent study evaluating a new college admissions system for universities in Chile made extensive use of biodata measures (Santelices, Ugarte, Flotts, Radovic, & Kyllonen, 2011). In all these studies there has been demonstrations that biodata measures add to other measures in predicting educational outcomes.

Interest in biodata measures seems to be on the increase due to a couple of new developments. One is in the area of what have come to be known as badges. A badge is a demonstration and credit for an accomplishment that occurs outside of school, but that can be used by schools or by employers as an indication of some kind of skill or competency. Badges are receiving considerable attention in the higher education world, and there are innovative programs underway to develop badge concepts, such as what kinds of skills to measure, and what kinds of accomplishments to treat as evidence for possession of that skill (Carey, 2012; Young, 2012). The other development is in the use of so-called "big data" technology to scan biodata information obtained from social media including job sites (e.g., Monster), Facebook, LinkedIn, Twitter, Google+, and MySpace. It is likely that these kinds of efforts will grow in the coming years and that we may see increased use of what has traditionally been called biodata.

Performance Tests

Much of 21st century skills measurement relies on rating scales (self and other), which is a subjective form of measurement. There is a certain unease with subjective forms of measurement: For example, imagine the public's acceptance of self or teacher ratings of students' mathematics

achievement as compared with their actual mathematics achievement test scores. Although methods such as anchoring vignettes and forced choice techniques seem to reduce the subjectiveness of ratings to some extent, there is nevertheless interest in more objective measurement. Situational judgment tests represent an improvement along these lines, usually, and can be developed for many 21st century skills. However, ultimately it would be desirable to develop performance tests for 21st century skills as a way to reduce subjectivity and to provide an additional basis for inferences about a student's level of 21st century skills in areas such as teamwork, collaboration, communication skills, and creativity.

There is a long history in psychology of attempting to measure noncognitive constructs with objective measures going back at least to Cattell (1957, 1973). More recently, within the field of emotional intelligence, there is a raging debate between proponents of ratings, and proponents of performance measures (see, for example, Matthews, Zeidner, & Roberts, 2004). A self-assessment of emotional intelligence might include items asking a student (or teacher) to rate his or her own competencies in interpreting and understanding emotions, whereas a performance measure might present a photograph of a face, and ask the viewer to judge the emotion expressed. The Mayer-Salovey-Caruso (2004) Emotional Intelligence Test (MSCEIT) consists of various such emotional intelligence performance tests.

There are many opportunities to develop new measures of 21st century constructs that might rightfully be called performance tests rather than ratings. Several of these are reviewed below.

Collaborative problem solving. Collaborative problem solving is a performance activity requiring groups of students to work together to solve problems. There are a number of collaborative problem solving activities already available. For example, Vanderbilt University's Jasper Woodbury Problem Solving Series (Cognition and Technology Group at Vanderbilt, 1992) include "adventure" problems in various content areas designed to be solved by students working together collaboratively. Similar collaborative problem solving activities have been developed for the ATC21S activity (reviewed earlier as a source for the nature of 21st century skills), and other groups. There also has been discussion that PISA 2015 will include collaborative problem solving as an additional domain area (besides mathematics, reading, and science).

A recent study on collaborative problem solving among college-age students suggests that collaborative problem solving might measure skills that are not measured in individual problem solving (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). The study consisted of two experiments. In the first, 40 three-person groups were administered a battery of cognitive tasks, including a figural reasoning test (Raven's Progressive Matrices), a creativity test (how many ways can a brick be used?), a moral reasoning test, planning a shopping trip, and a group typing test. They also were given a computer checkers test as an outcome measure. In the second study, 152 students in two-to-five person groups were given similar measures, along with an intelligence test, a word completion test, incomplete words test, quantity estimation test, and art copying test. They were given an architectural design task as an outcome. In both studies, some of the tests were given both individually and in the group setting (e.g., Raven's progressive matrices). There were several interesting findings. First, groups varied systematically in how well they solved problems—some groups were consistently better than others across the whole set of measures. The authors argued that this reflected a "collective intelligence" or "c" factor—some groups are simply better than others at solving problems. The second finding was what makes groups better or worse has little to do with the abilities of the individual members. Groups with

more intelligent members (as measured by the individual test performances) were not necessarily the groups that performed the best. Instead, other predictors played a role. In particular, the main predictors of group performance were (a) a social sensitivity, or emotional intelligence measure, (b) a process measure, the amount of turn taking, and (c) whether there were more females on the team.

There are several implications of this study. The most important is that collaborative problem solving might indeed prove to be a unique skill, and one not readily measured by standardized tests taken individually. A second is that the components of collaborative problem solving, such as turn taking, and being socially sensitive, would seem to be ones amenable to instruction.

Creativity. It has long been thought that creativity is a unique aspect of human ability, and in fact Carroll's (1993) meta-analysis of performance on cognitive tests identified creativity as a separate ability, an instance of idea production. There have been a variety of creativity measures used in educational and psychological measurement over the years, including tests requiring examinees to think of various uses for common objects, tests requiring the writing of a caption for a cartoon or title for a story, tests of the consequences of unusual events ("what would be the consequences if the entire United States west of the Mississippi turned overnight into an arid desert?"), creating a story from a picture, and thinking of ways to group objects by creating various rules. Carroll suggested that in defining what makes a task a good measure of creativity was "The common element [in good measures of the creativity factor] is that they require examinees fairly quickly to think of…a series of responses fitting the requirements of the task…furthermore, … it is difficult and challenging to think of responses beyond the more obvious commonsense ones" (Carroll, 1993, p. 428).

ETS developed several creativity measures in a series of studies (Bennett & Rock, 1995; Frederiksen & Ward, 1978), and evaluated them as experimental measures within an operational GRE® test administration. One test was called *Formulating Hypotheses*, and it gave examinees 2 minutes (per problem) to generate possible explanations for events such as a drop in car sales over a 10-year period, or a failure to observe children playing in a playground on one's daily walk at a time and place where children are usually observed playing. Students are incredibly poor at such tasks, generating only a bit over two explanations per 2 minutes, on average. Students were followed up a year later (after their first year in graduate school) and asked about various measures of student success. The most important finding of the study was that *controlling for GRE Verbal and Quantitative test scores*, psychology students who generated more explanations were found to be more knowledgeable about experimental design, had participated in more professional activities and had attended more professional meetings, had more publications, and were more likely to be involved in collaborative research.

A similar result was found in a study of army officers using a test called *Consequences* (Mumford, Marks, Connelly, Zaccaro, & Johnson, 1998). Consequences presents hypothetical situations to examinees (e.g., "what would happen if people no longer needed to eat food in order to survive?"), and examinees were given two minutes to generate as many consequences as they could think of. Results were similar to those found in the GRE study. Mumford et al. (1998) administered a five-item version of *Consequences* to a group of approximately 1,800 U.S. army officers, who were additionally given measures of verbal reasoning, and leadership expertise. *Consequences'* predictive validity was evaluated with respect to (a) quality and (b) originality scores from a complex situational problem solving task, (c) a biographical leadership assessment requiring one-or-two-paragraph candidate self-reports on displaying leadership skills, (d) rank, and (e) awards and appraisals. The main finding was that

Consequences predicted all of the outcomes, even after controlling for cognitive ability and expertise. Taken together these studies present remarkable findings in that it is rare to find incremental validity over general cognitive ability in predicting broad, real-world outcomes (e.g., Humphreys, 1986; Ree & Earles, 1991).

Summary and Next Steps

The purpose of this article was to justify the assessment (and implied development) of 21st century skills in education, particularly K-12 education, to identify the most important 21st century skills, and to suggest a variety of approaches for measuring them. It is clear that educators and employers claim that 21st century skills are important for the schools to develop, and for students to possess in order to be successful in the 21st century workplace. It is also clear that prediction studies show that these so called 21st century skills are as important, and in some cases more important for success than the content skills that are the focus of much current attention in large-scale standardized testing.

A review of various sources for identifying 21st century skills suggests that there are common themes running through all the various lists. A framework adopted in several publications by the NRC and NAS serves as a nice summary. The framework suggests that 21st century skills can be organized into the areas of cognitive skills (e.g., critical thinking, problem solving, creativity), interpersonal skills (communication skills, social skills, teamwork, cultural sensitivity, dealing with adversity), and intrapersonal skills (self management, self regulation, time management, self development, lifelong learning, adaptability, executive functioning). Furthermore, 21st century skills can serve as student learning outcomes, curriculum can be built around developing them, teacher professional development can emphasize such instruction, and various learning environments could be developed to promote them.

A review of assessment methods suggested that currently rating scales (self, teacher, and parent) are widely used in the assessment of 21st century skills, and that there have been significant innovations in rating scales (e.g., anchoring vignettes, forced choice). Situational judgment tests have increasingly been used to measure what might otherwise prove to be hard-to-measure skills such as teamwork, and communication skills. There are performance measures for some of the 21st century skills, such as collaborative problem solving, and creativity, and developments in these areas are likely to prove increasingly important in the short- and medium term, as users demand more objective measures of these very important skills.

References

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics, 118*(4), 1279–1334.

Baird, L. L., & Knapp, J. E. (1981). *The inventory of documented accomplishments for graduate admissions: Results of a field trial study and its reliability, short-term correlates, and evaluation.* Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED214940)

Beaton, A. E., Rogers, A. M., Gonzalez, E., Hanly, M. B., Kolstad, A., Rust, K. F., & Jia, Y. (2011). *The NAEP primer* (NCES 2011-463). Retrieved from the U.S. Department of Education, National Center for Education Statistics website: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011463

Bennett, R. E., & Rock, D. A., (1995). Generalizability, validity, and examinee perceptions of a computer-delivered formulating hypotheses test. *Journal of Educational Measurement, 32,* 19-36.

Binkley, M., Erstad, O., Herman, J., Raizen, S., & Ripley, M (with M. Rumble). (2010). *Defining 21st century skills and assessments* (Draft White Paper 1). Melbourne, Australia: Assessment and Teaching of 21st Century Skills. Retrieved from http://cms.education.gov.il/NR/rdonlyres/19B97225-84B1-4259-B423-4698E1E8171A/115804/defining21stcenturyskills.pdf

Bowles, S., & Gintis, H., (1976). *Schooling in capitalist America: Educational reform and the contradictions of economic life.* NewYork, NY: Basic Books.

Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature, 39*(4), 1137–1176.

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*(3), 460–502.

Burrus, J., & Roberts, R. D. (2012). *Elementary schools research collaborative mission skills assessment: fall 2011 assessment report.* Princeton, NJ: Educational Testing Service, Center for Academic and Workforce Readiness and Success.

Burrus, J., Roberts, R. D., Brenneman, M., & Lipnevich, A. (2012). *Elementary schools research collaborative preliminary report of first wave mission skills assessment.* Princeton, NJ: Educational Testing Service, Center for Academic and Workforce Readiness and Success.

Carey, K. (2012, April 8). A future full of badges. *The Chronicle of Higher Education.* Retrieved from http://chronicle.com

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press.

Casner-Lotto, J., & Benner, M. (2006). *Are they really ready for work? Employers' perspectives on the basic knowledge and applied skills of new entrants into the 21st century workforce.* New York: The Conference Board, Inc., the Partnership for 21st Century Skills, Corporate Voices for Working Families, & the Society for Human Resource Management.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York, NY: Harcourt, Brace, & World.

Cattell, R. B. (1973). *The scientific analysis of personality.* Weinheim, Germany: Beltz.

Cognition and Technology Group at Vanderbilt. (1992). The Jasper experiment: An exploration of issues in learning and instructional design. *Educational Technology Research and Development, 40*(1), 65–80.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136,* 1092–1122.

Crede, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3*(6), 425–453.

Csapó, B., Ainley, J., Bennett, R., Latour, T., & Law, N. (2010). *Technological issues for computer-based assessment* (Draft White Paper 3). The University of Melbourne, Australia. Retrieved from the http://atc21s.org/wp-content/uploads/2011/11/3-Technological-Issues.pdf

Darling-Hammond, L. (2011). Effective teaching as a civil right: How building instructional capacity can close the achievement gap. *Voices in Urban Education, 31,* 44–58.

Dede, C. (2010). Comparing frameworks for 21st century skills. In J. Bellanca & R. Brandt (Eds.)*, 21st century skills* (pp. 51–76). Bloomington, IN: Solution Tree Press.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*(1)*,* 40–57.

Durlak, J. A., & Weissberg, R. P. (2007). *The impact of after-school programs that promote personal and social skills.* Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., &, Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405–432.

Finegold, D., & Notabartolo, A. S. (2008). *21st century competencies and their impact: An interdisciplinary literature review.* Retrieved from the William and Flora Hewlett Foundation website: http://www.hewlett.org/library/grantee-publication/21st-century-competencies-and-their-impact-interdisciplinary-literature-review

Flanagan, J. C. (1949). A new approach to evaluating personnel. *Personnel, 26*, 35–42.

Frederiksen, N., & Ward, W. C. (1978). Measures for the study of creativity in scientific problem-solving. *Applied Psychological Measurement, 2*, 1–24.

Hedlund, J., Wilt, J. M., Nebel, K., Ashford, S. E., & Sternberg, R. J. (2006). Assessing practical intelligence in business school admissions: A supplement to the Graduate Management Admissions Test. *Learning and Individual Differences, 16*(2), 101–127.

Hernstein, R., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York, NY: Free Press.

Hopkins, D., & King, G. (2010). Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly, 74*(2), 201–222.

Howell, D. R. (2002). Increasing earnings inequality and unemployment in developed countries: Markets, institutions, and the "Unified Theory." *Politics & Society, 30*(2)*,* 193–243.

Humphreys, L. G. (1986). Commentary. *Journal of Vocational Behavior, 29*(3), 421–437

Ingels, S. J., Abraham, S. Y., Rasinski, K. A., Karr, R., Spencer, B. D., Frankel, M. R., & Owings, J. A. (1990). *National Education Longitudinal Study of 1988. Base year: Teacher component data file user's*

manual(NCES 90-484). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* London, England: Praeger.

Jones, A. (2012). *The Partnership for Assessment of Readiness for College and Careers (PARCC).* Paper presented at the 21st century knowledge and skills: The new high school curriculum and the future of assessment conference*,* University of Southern California, Los Angeles.

Kyllonen, P. C. (2008). *The research behind the ETS personal potential index (PPI).* Princeton, NJ: Educational Testing Service.

Kyllonen, P. C., Lipnevich, A. A. Burrus, J., & Roberts, R. D. (2008). *Personality, motivation, and college readiness: A prospectus for assessment and development.* Princeton, NJ: Educational Testing Service.

Kyllonen, P. C. (2012, January). *The role of noncognitive skills in academic success*. Paper presented at the conference of 21st century knowledge and skills: the new high school curriculum and the future of assessment, University of Southern California, Los Angeles.

Levin, H. M. (1998). Educational performance standards and the economy. *Educational Researcher, 27*(4), 4–10.

Levine, L. (2004). *Education matters: Earnings by educational attainment over three decades* (Federal Publications Paper 245). Retrieved from the Cornell University ILR School website: http://digitalcommons.ilr.cornell.edu/key_workplace/245

Levy, F., & Murnane, R. J. (2004). *The new division of labor: How computers are creating the next job market*. Princeton, NJ: Princeton University Press.

Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economical Journal: Applied Economics, 3*(1), 101–128.

MacCann, C., Fogarty, G. J., Zeidner, M., & Roberts, R. D. (2011). Coping mediates the relationship between emotional intelligence (EI) and academic achievement. *Contemporary Educational Psychology, 36*, 60–70.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion, 8,* 540–551.

MacCann, C., & Roberts, R. D. (2012). Just as smart but not as successful: Obese students obtain lower school grades but equivalent test scores to non-obese students. *International Journal of Obesity*, 1–7.

MacCann, C., Wang, L., Matthews, G., & Roberts, R. D. (2010). Emotional intelligence and the eye of the beholder: Comparing self- and parent-rated situational judgments in adolescents. *Journal of Research in Personality, 44,* 673–676.

Manyika, J., Lund, S., Auguste, B., Mendonca, L., Welsh, T., & Ramaswamy, S. (2011). *An economy that works: Job creation and America's future.* Retrieved from the McKinsey Global Institute website: http://www.mckinsey.com/Insights/MGI/Research/Labor_Markets/An_economy_that_works_for_US_job_creation

Matthews, G., Zeidner, M., Roberts, R. D. (2004). *Emotional intelligence: Science and myth*. Cambridge, MA: MIT Press.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 15,* 197–215.

Mitchell, C. (2012, January). *SMARTER Balanced Assessment Consortium (SBAC)*. Paper presented at the 21st century knowledge and skills: The new high school curriculum and the future of assessment conference, University of Southern California, Los Angeles.

Mumford, M. D., Marks, M. A., Connelly, M. S., Zaccaro, S. J., & Johnson, J. F. (1998). Domain-based scoring of divergent-thinking tests: Validation evidence in an occupational sample. *Creativity Research Journal, 11,* 151–163.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management, 19*(4), 547–568.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics, 77*(2), 251–266.

National Research Council. (2008). *Research on future skill demands: A workshop summary.* Washington, DC: The National Academies Press.

National Research Council. (2010). *Exploring the intersection of science education and 21st century skills: A workshop summary.* Washington, DC: The National Academies Press.

National Research Council. (2011). *Assessing 21st century skills: Summary of a workshop.* Washington, DC: The National Academies Press. Retrieved from http://ctl.sri.com/publications/downloads/21st_century_skills.pdf

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* Washington, DC: The National Academies Press.

Noftle, E. E., & Robins, R. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93*, 116–130.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89,* 187–207.

Partnership for 21st Century Skills. (2002). *Framework for 21st century learning*. Retrieved from http://www.p21.org/overview/skills-framework

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

Ree, M.J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology, 44*, 321-332.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skills factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130,* 261–288.

Santelices, M. V., Ugarte, J. J., Flotts, P., Radovic, D., & Kyllonen, P. (2011). *Measurements of new attributes for Chile's admission system to higher education* (ETS Research Report No. RR-11-18). Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RR-11-18.pdf

Scardamalia, M., Bransford, J., Kozma, B., & Quellmalz, E. (2010). *New assessments and environments for knowledge building* (Draft White Paper 4). The University of Melbourne, Australia. Retrieved from http:/./ikit.org/fulltext/2010_NewATKB.pdf

Schmitt, N., Oswald, F. L., Kim, B. H., Imus, A., Merritt, S., Friede, A., & Shiypuri, S. (2007). The use of background and ability profiles to predict college student outcomes. *Journal of Applied Psychology, 92*, 165–179.

Segal, C. (in press). Misbehavior, education, and labor market outcomes. *Journal of the European Economic Association*.

Shultz, M. M., & Zedeck, S. (2011). Predicting lawyer effectiveness: Broadening the basis for law school admission decisions. *Law and Social Inquiry, 36*(3), 620–661.

Smith, M. S., Chudowsky, N., Ginsburg, A., Hauser, R., Jennings, J., & Lewis, S. (2012). *NAEP background questions: An underused national resource. A report to the National Assessment Governing Board by the expert panel on strengthening the NAEP background questions*. Washington, DC: National Assessment Governing Board.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Reconsidering the use of personality tests in personnel selection contexts: Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84*(5), 1041–1053.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item/test functioning (DIF/DTF) on selection decisions: When are statistically significant effects practically important? *Journal of Applied Psychology, 89,* 497–508.

Stemler, S. E., Bebell, D., & Sonnabend, L. (2011). Using school mission statements for reflection and research. *Educational Administration Quarterly, 47*(2)*,* 383–420.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M,….Grigorenko, E. L. (2000). *Practical intelligence in everyday life.* New York, NY: Cambridge University Press.

Stricker, L. J., Rock, D. A., & Bennett, R. E. (2001). Sex and ethnic-group differences on accomplishment measures. *Applied Measurement in Education, 14,* 205–218.

Voogt, J., & Pareja Roblin, N. (2010). *21st-century skills: Discussion paper.* Retrieved from http://www.internationalsymposiumoneducationalreform.com/storage/21st%20Century%20Skills.pdf

Weekly, J., & Ployhart, R. (2006). *Situational judgment test: Theory, measurement and application*. Mahwah, NJ: Lawrence Erlbaum Associates.

Weekly, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. Weekly & R. Ployhart (Eds.), *Situational judgment test: Theory, measurement and application* (pp. 157–182). Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, M., Bejar, I., Scalise, K., Templin, J., Wiliam, D., & Irribarra, D. T. (2012). Perspectives on methodological issues. In P Griffin, B. McGaw, & E. Care (Eds.), A*ssessment and teaching of 21st century skills* (pp. 67–141). Dordrecht, Germany: Springer.

Yazzie-Mintz, E. (2010). *Charting the path from engagement to achievement: A report on the 2009 High School Survey of Student Engagement*. Bloomington, IN: Center for Evaluation & Education Policy.

Young, J. R. (2012, January 8). “Badges” earned online pose challenge to traditional college diplomas. *The Chronicle of Higher Education.* Retrieved from http://chronicle.com/article/Badges-Earned-Online-Pose/130241/

Invitational Research Symposium on

Technology Enhanced Assessments

The Center for K–12 Assessment & Performance Management at ETS creates timely events where conversations regarding new assessment challenges can take place, and publishes and disseminates the best thinking and research on the range of measurement issues facing national, state and local decision makers.