



CENTER FOR THE ECONOMICS
OF HUMAN DEVELOPMENT
THE UNIVERSITY OF CHICAGO

CONFERENCE ON MEASURING AND ASSESSING SKILLS REPORT

October 1-2, 2015, Chicago, IL

5 OPEN QUESTIONS

How predictive are measurements of skills for short and long-run outcomes?

Two presentations, by Tim Kautz and John Eric Humphries, considered the importance of both cognitive and non-cognitive skills for predicting lifecycle outcomes. They both showed that cognitive and non-cognitive skills predict long-term outcomes but their relative predictive power depends on the outcome. Oliver John presented some correlations showing the predictive power of the Big Five personality traits with a range of later-life outcomes.

Does environment effect the predictive accuracy of skill measurements?

Estimates presented by Patrick Kyllonen and Armin Falk both show cross-national differences in preference and Big 5-based measures of socio-emotional skills but discuss methodologies that help to elicit measurements robust to cultural and national differences. Patrick Kyllonen showed that forced choice questionnaires and anchoring vignettes can improve the predictive power of self-reported personality across cultures.

How significantly do changes in incentives and/or context effect measurements of skills?

Big 5 traits were less predictive of decision making ability in an experiment presented by Daniel Silverman, while self-reported Big 5 personality traits were found to be less predictive of social behavior in incentivized games than experimentally elicited altruism and trust parameters in findings presented by Aldo Rustichini.

Are there separate components of skills, and how can they be identified?

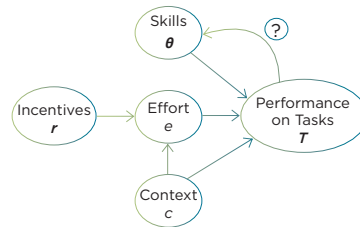
A major component of a person's general ability and skill is their personality, which can impact how two individuals of equal cognitive ability may react differently to environmental and social stimuli. Several presenters discussed the difficulty in isolating particular components of skills. James Heckman presented a task-based framework that illustrated the underlying identification problem. Andrew Caplin presented a model showing the difficulty of separately identifying knowledge and test-taking skill in the context of a simple, multiple-choice test. Different fields address this issue in different ways. John Eric Humphries and Tim Kautz both showed how single measures can be allowed to depend on multiple skills.

Are measurements of skills comparable across elicitation strategies?

Achievement and IQ tests explained less of the variance in high school graduation and college enrollment rates than observational measures like grades in findings presented by Tim Kautz and John Eric Humphries. Armin Falk discussed the relationship between economic preference parameters and personality traits. There are few empirical relationships between these different systems. Michelle Lamar and Bob Mislevy showed how games and simulations can be used to track learning, but there remain open questions about how those elicitation schemes are linked to traditional measures.

OVERVIEW

The Conference on Measuring and Assessing Skills was held to compare methods for measuring individual's skills and to explore how those skills are related to life-cycle outcomes, such as educational attainment, employment, health, mental health, earnings, crime, citizenship, family life, and many others. Both hard skills, like cognitive ability and soft (or socio-emotional) skills are considered as important inputs for determining labor, education, and other life outcomes. The multiplicity of influences on performances on tasks (T) is illustrated schematically in the below figure.



Skills (θ), effort (e), and context (c) shape performance on tasks. Both incentives (r) and context (c) shape effort (e). For all tasks, the challenge is to infer skills from tasks. This is the inferential step denoted by $\theta?$ in the diagram. To infer any component of vector θ , one needs to control for all other components as well as context and effort.

All measurement schemes claim to solve this identification problem. Measures that include components of both intellectual ability and socio-emotional skill, such as IQ combined with personality inventories or an individual's grades, are better predictors of short- and long-run outcomes than using IQ alone. This conference aimed to discuss specific ways this is done in each of the three approaches:

1. Tests, Observer & Self Reports

Self-reports are a major source of information as in the Big Five inventories, often used to assess personality. So are assessments by parents, teachers, supervisors, and peers. These assessments are elicited by surveys administered to subjects. IQ scores often serve as the best dedicated measure for cognitive ability, while dedicated measures of socio-emotional ability are unreliable.

2. Observed Behaviors

Eliciting measurements of skills directly through observed behaviors helps address the lack of standardization associated with test and interview based measures. This sometimes involves using early measures of behavior to predict later behaviors, an approach used by empirical economists. Empirically observed behaviors provide the benefit of being the most accurate representations of how agents actually respond to "real world" tasks.

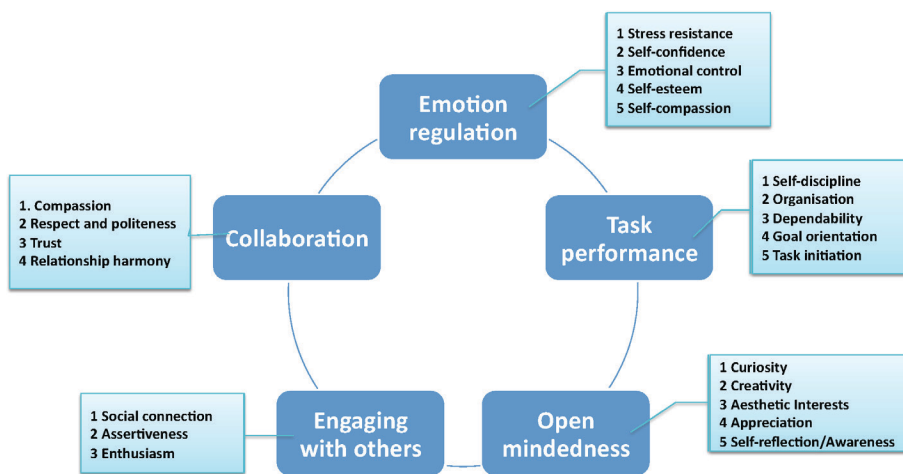
3. Games and Laboratory Experiments

Games and other laboratory experiments have found a growing importance in both measuring and improving skills. Dictator and centipede games have been employed by economists as a tool for measuring an individual's degrees of altruism and self-interest. Game-based approaches for measuring skills benefit from being incentivized, with incentivized behavior being more representative of real world responses as compared to tests or interviews which often have only arbitrary meaning to the subject.

PERSONALITY TRAITS AND PREFERENCES AS COMPONENTS OF SKILLS

A major component of a person's general ability and skill is their personality, which can impact how two individuals of equal cognitive ability may react differently to environmental and social stimuli. The Big 5 personality traits have become a standard measure of personality and are correlated with many outcomes, both short- and long-term. Internal locus of control and grit have also been shown to predict outcomes better than IQ measures alone. Oliver John presented correlations between the Big 5 traits and labor, family, and educational outcomes among a group of college-educated women. Conscientiousness and openness have been found to be strongly correlated with performance on math, reading, writing, and spelling tests, which are in turn often used as measures of cognitive ability themselves. Extraversion is correlated with occupational achievement and satisfaction with one's job, while openness is predictive of occupational and artistic creativity.

FRAMEWORK OF SOCIO-ECONOMIC SKILLS



Several preferences have been frequently identified as predictors of one's behavior in experimental tasks and in life-cycle outcomes within the economics literature, including reciprocity, patience, risk-aversion, and altruism. Patience, (or inter-temporal preference) and risk-aversion (or preference for or against uncertainty), have been found to have predictive power for outcomes like labor earnings, educational attainment, and entrepreneurial behavior as well as social outcomes like civic participation and philanthropic activity.

Correlation of skill measures and elicited preferences^a

	Trust	Altruism
Agreeableness	-.36	-.15
Neuroticism	.44	.25
Extraversion	.81*	.61
Conscientious	.52	.23
Openness	-.96*	-.35
IQ	.21***	.13*
Risk Pref.	.07	-.09

savings behavior, while risk aversion is a predictor of entrepreneurial choices and tobacco use in a global survey representative of the population in 96 countries.

Falk, et al. (2015) shows that patience is a statistically significant predictor of an individual's income, saving, and education choices, and when aggregated, also correlates strongly with national GDP. Presentations by Armin Falk and Daniel Silverman reviewed findings from recent efforts to combine experimental and survey methods for measuring individual preferences, including patience, risk aversion, reciprocity, altruism, and trust. Patience is found to be significantly correlated with educational attainment and

FACETS OF THE BIG 5

Openness

- Strong intellectual interest
- High aesthetic sensitivity
- Imaginative & original

Conscientiousness

- Disciplined
- Perseverant
- Tidy & reliable

Extraversion

- Sociable, outgoing, & assertive
- Enthusiastic & active

Agreeableness

- Compassionate
- Trusting
- Respectful of others

Neuroticism

- Emotional stability
- Self-confident
- Calm & resilient
- Non-temperamental

ECONOMIC PREFERENCES

Altruism

- The degree to which a person derives benefit from the well being of others, through self-sacrificing and philanthropic behaviors.

Reciprocity

- Positive reciprocity can be understood as how willing a person is to return a favor.
- Negative reciprocity is their willingness to punish or take revenge when a favor is not returned.

Risk Preference

- A person's willingness to accept risk relative to certainty.

Patience

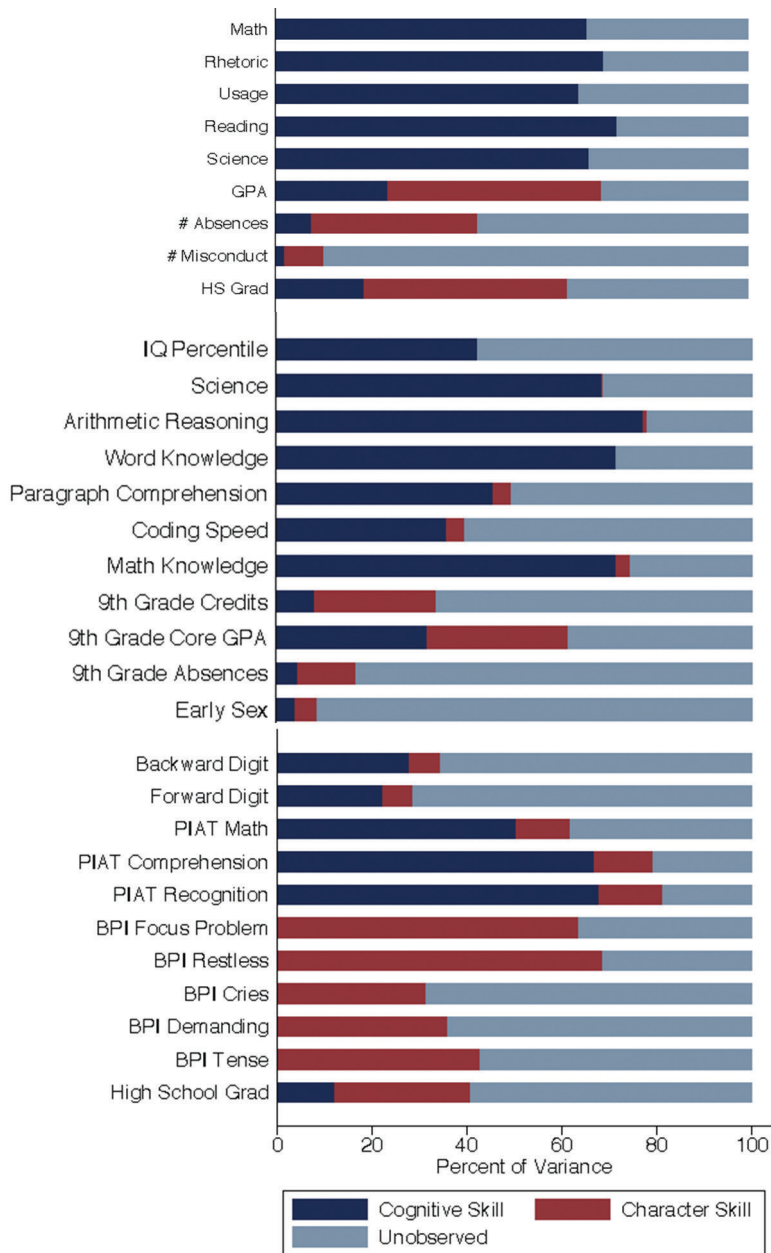
- A person's ability to wait and delay receiving benefit until the future.

Trust

- Trust or a person's belief that others have only good intentions, while not strictly a preference, is an important input for a person's social behavior.

^aSource: Aldo Rustichini, Personality and Social Behavior. Estimates based on behavior in a monetary exchange game conducted in a laboratory setting. Stars represent (90%*), (95%**), and (99%***) significance levels.

Explained variance by cognitive and non-cognitive factor for different measurements and outcomes^b



CHALLENGES TO THE MEASUREMENT OF SKILLS

The correlation between measures of cognitive skills with non-cognitive ones poses challenges to interpreting test scores as the measurements of a person's underlying abilities or skills. Many of the empirical skill measures common in the literature are correlated with both socio-emotional skills and cognitive ability, which poses problems for evaluating how differences in outcomes are determined by relatively fixed cognitive ability versus more malleable socio-emotional skills. Math and reading test scores and other test scores in general are often assumed as dedicated measures for cognitive ability. However, they are also correlated with non-cognitive measures, including Big 5 traits and previously observed behavior, such as the different indices of childhood behavioral problems. Observed prior behavior, such as truancy and discipline problems, can be useful measurements of skill in general, but are influenced by both intelligence and socio-emotional skills. Grades and GPA are often one of the best predictors of future outcomes and best measures of overall ability and skill. However, grades are impacted by intelligence, social interactions, and emotional behavior in the classroom.

One solution, proposed in a presentation by Robert Mislevy, is the movement towards interpreting tests as representative of cognitive processes, which can be context and environment specific. In turn, measurements can be better designed to capture these complex processes and their relationship to a person's overall ability. Michelle LaMar extends this idea in a presentation on using Markov decision processes as a model for measuring ability as a cognitive process rather than an underlying trait. Peg solitaire can be used as a method for eliciting both capability and motivation and measuring how they impact a player's strategic thinking. More complex environments using computer games can be used to measure general ability based on performance on a complex task, such as managing a simulated city's pollution and energy resources. This approach applies a Bellman Equation to model a student's intertemporal decision making process and can identify several related parameters.

Definition of Personality:

“Personality traits are the relatively enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances.”

- Roberts (2009, p. 140) from “Back to the Future: Personality and Assessment and Personality Development”.

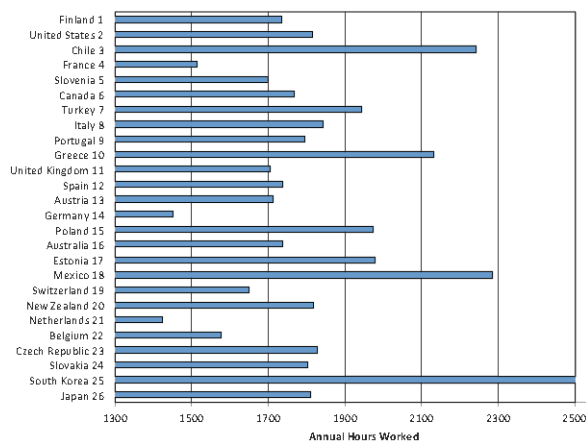
^bSource: Tim Kautz, “Using School Administrative Data to Measure Non-Cognitive Skills”

IMPROVING THE VALIDITY OF SKILL MEASUREMENTS

Personality traits have typically been studied through self-reports and observer reports, and the accuracy of these methods, which can suffer from individuals trying to manage their reputation, introduces concerns regarding inaccuracy and bias. Concerns about deceptive or biased responses in self-reports can be corrected for or reduced by using observer reports, such as those by teachers or parents, in conjunction with self-reports. Patrick Kyllonen presented two methods for improving the validity and reliability of self-reports, 1) anchoring vignettes, which require respondents to rate hypothetical people or situations before completing a self-rating and 2) forced choice methods which present two or more statements which respondents must choose from with respect to how well they describe the respondent. Self-reported measures of Big 5 traits and mathematics proficiency have been found to have stronger correlation with outcome measures when anchored using vignette methods than with unanchored Likert scale measures, while Big 5 personality traits correlate stronger with development outcomes across different countries and cultures when measured with forced choice methods.

Experimental validation methods can be used to better design survey questionnaires.

Correlation of skill measures and elicited preferences^e



Incentivized experiments are assumed to be more accurate in capturing an individual's true response, and survey questions can be selected so they have the greatest predictive power for behavior in the experimental task. Experimental settings offer more opportunity for eliciting skills and exploring effects from context and environment. Validating the comparability of measurements elicited with questionnaires to those elicited by experimental methods helps address concerns about inaccuracies inherent to survey methods, while avoiding the cost and feasibility issues that prevent experimental elicitations from being used on a large scale. Decision-making quality is considered to be representative of rational decision making. Findings demonstrate that higher quality decision making is tied to wealth using web-based experiments in the Netherlands and US piggy backed to survey based panels.

^cSource: Tim Kautz, "Using School Administrative Data to Measure Non-Cognitive Skills".

^dSource: Patrick Kyllonen, "Anchoring Vignettes, Ranking Methods, and Situational Judgment Testing for Improving the Quality of Self-Reports".

^eSource: Armin Falk, "The Nature and Predictive Power of Preferences: Global Evidence". Estimates based on a survey conducted on a globally-representative population of individuals in 76 countries.

CROSS-CULTURAL VALIDITY

Big 5 Trait ^d	Human Development Index Likert Scale (Forced Choice)	Global Competitive Index
Agreeableness	.09(.57)	.39(.58)
Emotional Stability	.07(.27)	.50(.53)
Extraversion	.41(.76)	.20(.46)
Conscientious	-.46(-.08)	-.40(.21)

Preferences ^e	Between Country (% variation)	Within Country (% variation)
Patience	14.3	85.7
Risk Pref.	9.9	90.1
+ Reciprocity	11.4	88.6
- Reciprocity	7.6	92.4
Altruism	12.3	87.7
Trust	8.3	91.7

Differences in Preferences Between OECD and Non-OECD Countries

