# Predictive Validity of Soft Skills Measures: A Research Synthesis

Smita Das, Steven Glazerman, Savanna Henderson, Sarah Kabay

Innovations for Poverty Action

## Abstract

Soft skills are believed to be important contributors to labor market success, but there is little consensus on how to measure them and whether any measures can predict longer term outcomes like employment, earnings, and job performance. This study synthesizes four decades of interdisciplinary research to address the critical question of whether particular soft skill measurement approaches (e.g., self-reports, observer-reports, or task-based assessments) and skill types reliably predict labor outcomes. Our comparative analysis focuses on 50 quality studies primarily conducted in Western contexts, complemented by early findings from ongoing studies of African youth, a group largely absent from prior research.

The findings reveal that quantitative evidence on soft skills and labor market outcomes is limited in quantity, geography, skill type, and measurement type. Certain skill types—such as aspirations, higher-order thinking, grit, and responsibility—show more consistent associations with labor market success. However, these results are derived from a relatively small number of studies. In contrast, there is no observed association between anxiety-related measures and labor outcomes, and evidence is mixed for other skills, underscoring the variability across both measurement approaches and contextual factors. Our analysis finds that no single type of measurement consistently predicts outcomes across diverse settings, raising concerns about the generalizability of existing tools and findings. Our findings have significant implications for research and practice. They caution against assuming that interventions enhancing measured soft skills will automatically improve labor market outcomes and highlight the need for employers and policymakers to rigorously validate soft skill measures before using them for recruitment or lending. These results underscore the importance of refining soft skill measurement and exploring contextual variations, particularly in low- and middle-income settings. This work contributes to global development by offering actionable insights for improving soft skill assessment and aligning it with labor market needs.

Keywords: soft skills, socio-emotional skills, noncognitive, labor, employment, validity

## I. Introduction

There has been growing consensus that soft skills have an important role to play in addressing global poverty. Investing in human capital extends beyond traditional academic skills like literacy and mathematics to a wide range of traits and abilities, which can include interpersonal skills, positive mindsets, emotional regulation, creativity, and critical problem-solving. Social innovators have fielded hundreds of different soft-skills interventions aimed at helping poor, vulnerable, or marginalized groups to become more employable, better entrepreneurs, and empowered to live better lives (see for example, Puerta, Valerio, & Gutiérrez Bernal 2016; Corcoran, et al. 2018).

In this paper, we focus on the relationship between soft skills and labor because of its increasing policy relevance: soft skills are transferable across occupations and relevant for both workplace success and household decision-making; their importance is expected to grow with automation, artificial intelligence and remote work; and with a burgeoning literature examining the causal impact of soft skills programming on labor in developing countries, soft skills interventions are increasingly highlighted in policy design.

It is still difficult to test and improve soft skills interventions, in part, because soft skills are difficult to measure and there is little evidence on the *predictive validity* of measures. Concurrent validity, the relationship between skills and outcomes at the same time point, is indicative of promising measures. However, predictive validity, which is the ability of a measure to forecast future outcomes, is a more demanding criterion to satisfy and a powerful tool for program design. If a measure is linked to long term outcomes, such as education and employment, the measure can be used as a short-term outcome for policy development and data-driven program design. Predictive validity can contribute to key concerns such as whether measures (i) are measuring the intended construct, (ii) capture changes in skill levels, and thus capture the effectiveness of policies and programs, (iii) indicate which skills require training to improve outcomes such as earnings, productivity, wellbeing, education, and health, and (iv) are suitable as selection criteria in college and job aptitude tests, hiring decisions, program admission, or loan approvals. However, the application of predictive validity to selection and hiring must require intensive scrutiny to ensure its use does not increase illegal or unethical discrimination or bias, particularly because the validity of soft skills measures often varies greatly across cultures and target groups (Laajaj et al., 2019).

Predictive validity evidence is somewhat rare because of the data that this analysis demands. It requires measuring the construct at one point in time and then allowing enough time to pass to observe its relationship with long-term outcomes. Moreover, the skills and skill measures vary so widely that it is difficult to generate comparable evidence across studies and target populations.

We address this gap in knowledge by conducting an extensive literature review of studies examining predictive validity of soft skills measures for labor outcomes. We list the most prevalent categories of skills and measurement types in this literature and identify which instruments have been shown to predict employment, earning, and job performance. As much of this literature is based in Western contexts, we then describe results from four recent, thus far unpublished studies in Africa.

### A. Background on Soft Skills Measurement

We use the term "soft skills" to capture any skills other than traditional academic competencies like math, reading, writing, and other subjects typically taught in school curricula. Several other terms are often used interchangeably with "soft skills": "life skills", "21st century skills", "non-cognitive" skills, and "socio-emotional" skills or "socio-emotional learning." The Collaborative for Academic, Social, and Emotional Learning (CASEL), defines socio-emotional learning as the "process through which children and adults acquire and effectively apply the knowledge, attitudes, and skills necessary to understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions" (CASEL, n.d.). We take an inclusive approach to expand our search beyond this list of skills to include personality traits, beliefs, attitudes, and character for individuals of all ages.

In theory, soft skills could improve labor outcomes at several points along the employment pathway: skills such as grit, listening, and the ability to ask questions may contribute to learning that better positions one for employment; awareness of strengths and weaknesses may contribute to confidence and targeting the types of work where one can be most successful; social awareness and proactivity may give individuals a larger menu of income-generating opportunities; positive mindset, self-control, perseverance, and networks may be closely tied to job search, job retention or business longevity by giving individuals the tools to overcome financial constraints, the psychology of poverty, social norms and pressure, and trauma; negotiation and communication skills can enhance an individual's capacity to cultivate social support, gain access to resources and household allocations of time and money, or obtain better prices to earn higher profits. At each of these steps, bridging the inaction-action gap may involve soft skills such as positive mindset, perseverance, personal initiative, and developing supportive networks.

The transferable nature and importance of these skills across occupations and contexts is demonstrated by the consistency with which employers report an unmet demand for these skills (Cunningham & Villasenor, 2016). However, there is mixed evidence as to whether soft skills training programs are successful (Campos et al., 2017; Gielnik et al., 2015; Chioda et al., 2021; Alibhai et al., 2019; Ubfal et al., 2022). Soft skills may be difficult to teach or easily forgettable. They may also be practically unusable: some groups may face backlash for the use of certain skills, like negotiation (Bowles et al., 2007), or there may not be sufficient job or capital opportunities to leverage these skills. Finally, learned skills may be used to target goals that are not consistent with economic empowerment. Which skills matter most, and for whom, is likely to differ widely by occupation, local social norms, economic context, and demographics such as gender (Ajayi et al., 2022).

Unfortunately, there are several pervasive roadblocks to developing soft skill measures to further policy development. First, it is difficult to consolidate evidence without a common basis for combining results across contexts and populations. The field is fractured and innovation is slow, because researchers in different fields (e.g. psychology and economics) or in different contexts (e.g. North America and Sub-Saharan Africa) do not use common terminology or conceptual frameworks. While economists often have a focus on labor outcomes and rigorously examine programming in developing countries, they rarely communicate with psychologists who have sophisticated mechanisms for examining validity of measures and a key understanding of soft skills, language, and individual perception of measures.

A second barrier to building knowledge is that most existing literature in economics tends to focus on a narrow set of commonly used measures that address only a fraction of the vast array of soft skills, such as measures of self-efficacy and locus of control. The tools in most common use tend to be the ones that are best known but are not always ones that hold up well to rigorous scrutiny (Laajaj and Macours 2018; Arias et al. 2019). Even if we can agree on measures, evidence is only beginning to emerge on how to implement the measures effectively (Chen et al. 2020).

Finally, researchers and practitioners find it difficult to take advantage of existing measures. Measures are often proprietary, and many dozens of overlapping and duplicative measures have emerged in recent years (Galloway et al. 2017, Anvari et al., 2024) without clear criteria for selecting the most appropriate ones to use. Measurement experts have tried to publish data on these instruments, but the focus has been on measure *reliability*, which refers to the stability of a measure and is easy to calculate using conventional statistics from cross-sectional data, and less on measure *validity*, which refers to the ability of the measure to capture the concept it describes. Furthermore, predictive validity – the ability of a measure to predict some future outcome of interest – requires longitudinal data, which is more difficult to compile. Other types of validity analysis may include an examination of the underlying latent factor(s) for a set of observed items, whether the measure is theoretically representative of all aspects of the skill, how the measure is understood by respondents, how a measure relates to other measures of the same skill, how the measure relates to similar and differentiated concepts, and whether the measure relates to real world outcomes of interest.

### B. Predictive validity and causality

A simple theoretical framework helps illustrate the ideas behind (soft) skill accumulation and how we measure it to predict labor market outcomes. We start by considering observable skills and non-skill determinants of outcomes as three sets of variables, which can be vectors $X$, $Y$, and $Z$, and the unobservable determinants $u$.

$X$ = Family environment and other background characteristics that may relate to both skills and labor market outcomes

$Y$ = Labor market outcomes such as employment, earnings, and job performance

$Z$ = Skills, which includes imperfect measures of unobservable "true" skills and traits $Z^*$ such as "conscientiousness"

u = unobservable determinants of labor market outcomes

$$(1)\ Y = f(X, Z, u)$$

Defining a true Z* separately from Z acknowledges that measurement error could play a role in empirical estimates of the returns to soft skills. These could be subscripted by individual and time period, but to simplify, we only use periods 0 and 1 for before and after entering the labor market. We want to know if $Z_0$, soft skill measured at time zero, is a good proxy for $Y_1$. A useful empirical specification for estimating the relationship between Y and Z would look like the linear approximation in equation 2:

$$(2)\ Y_1 = \beta X_0 + \gamma Z_0 + u$$

$$(3)\ Z_0 = Z_0^* + v$$

This supposes that skill determination occurs completely in time zero. It also assumes that we know which skills are indeed skills (changeable over time through training) or immutable traits. Furthermore, we have proposed classical measurement error in equation (3), which may oversimplify the properties of current measures of soft skills but acknowledges the distinction between soft skills and soft skill indicators like indices derived from self-report questionnaires. In practice, most developers of soft skill measures report the indicator's psychometric properties such as reliability, measured as internal consistency of multiple items or test-retest consistency of the scale. These can in theory be used to directly estimate measurement error.

Obtaining a significant relationship between $Z_0$ and $Y_1$ requires the persistence of the skill level, the persistence of the effect of the skill, or potentially reverse causality. An additional key consideration in examining predictive validity is selecting a time period that matches the planned policy application. A short timeline would be required when seeking to use measures for hiring, selection, or work-ready training, whereas a longer timeline may be required when examining school-based training.

This review of the literature includes papers that report on empirical estimates of the coefficients $\gamma$ or the correlation ($\gamma$, scaled by the ratio of standard deviations of Z and Y) used to represent the predictive validity of Z. The challenge is that they differ widely in specification of equation (2). Nearly all ignore the measurement error in Z (equation 3) so long as the reported psychometric properties are acceptable by some standard. More importantly they rarely address the endogeneity of Z resulting from the correlation with unobservable $u$. Thus, the estimated regression coefficients and correlations we summarize do not capture causal relationships, but associations that may or may not be replicated in future applications.

We note this as a limitation but acknowledge that even causal estimation of the returns to soft skills would be context-specific. As noted in a recent review of skills and human capital in the labor market (Deming and Silliman 2024), there is no universal relationship between skills and earnings or productivity. The relationship is a result of the equilibrium between supply and demand for skills, which can differ from one labor market to another across place, time, industry, and occupation.

It is possible to derive causal estimates, even if this type of study is not found in the review we conducted. Researchers can induce exogenous shifts in soft skills Z through randomized trials of interventions that are likely to have a positive impact on these skills. Thus, assignment to a treatment T that has a non-zero impact on $Z_0$ can be used as an instrument for endogenous skills as long as there is a strong enough impact and a long enough follow-up period to observe labor market outcomes. We return to this concept in the conclusions where we discuss future research.

Though not as rigorous as a randomized trial, research would also benefit from utilizing a dynamic specification that examines whether the change in the soft skill measure is associated with changes in labor outcomes. This exercise would be essential for testing whether measures can be used for pre-post assessments of soft skills interventions.

## II. Methods and Data

### A. Scope

The goal of this research synthesis was to identify and review published studies that assessed the ability of soft skill measures to predict labor market outcomes, with the skill measure collected prior to measuring the outcome. We included literature that addressed a range of labor market outcomes, including employment, income/salary and job performance.

### B. Search strategy and selection criteria

We identified papers eligible for this review if they met the following criteria: (1) available in the English language; (2) abstracts included the term “validity” and at least one keyword related to methodology, soft skills, and labor market outcomes. Implementers and psychologists often debate the mutability of certain terms associated with soft skills. In order to adopt an inclusive approach and remain agnostic, we included several concepts specific to socio-emotional skills, as well as other soft skills terms, such as personality traits, beliefs, preferences, and attitudes. Search terms were pilot-tested and details of the keyword search can be found in Box 1.

**Box 1. Keyword Search**

We classified search terms into four categories, which were combined with Boolean operators to search both platforms utilizing the 'abstract' field:

- **Analytical Methods & Research Methodology:** validity AND (predictive OR longitudinal OR follow-up) AND
- **Labor Market Terms:** (employment OR income OR job OR labor OR profit OR earnings) AND
- **Socio-emotional Skills:** "socio-emotional skills" OR noncognitive OR "soft skills" OR "life skills" OR "social skills" OR (perseverance OR grit OR resilience OR adaptability OR empathy OR listening OR initiative OR proactive OR "self-control" OR attention OR "self-awareness" OR "self-efficacy" OR "self-concept" OR communication OR "problem solving" OR creativity OR emotional OR initiative OR decision OR network OR teamwork OR intrapersonal OR interpersonal OR awareness OR leadership OR regulation OR "cognitive reappraisal" OR relationship OR negotiation OR "conflict resolution" OR collaboration OR cooperation OR persuasion OR trust OR relatedness) OR
- **Other Soft Skills:** personality OR belief OR mindset OR attitude OR character

The search also excluded each of the following words: injury, rehab, "return-to-work"

In EBSCO, additional limitations were placed on the search:

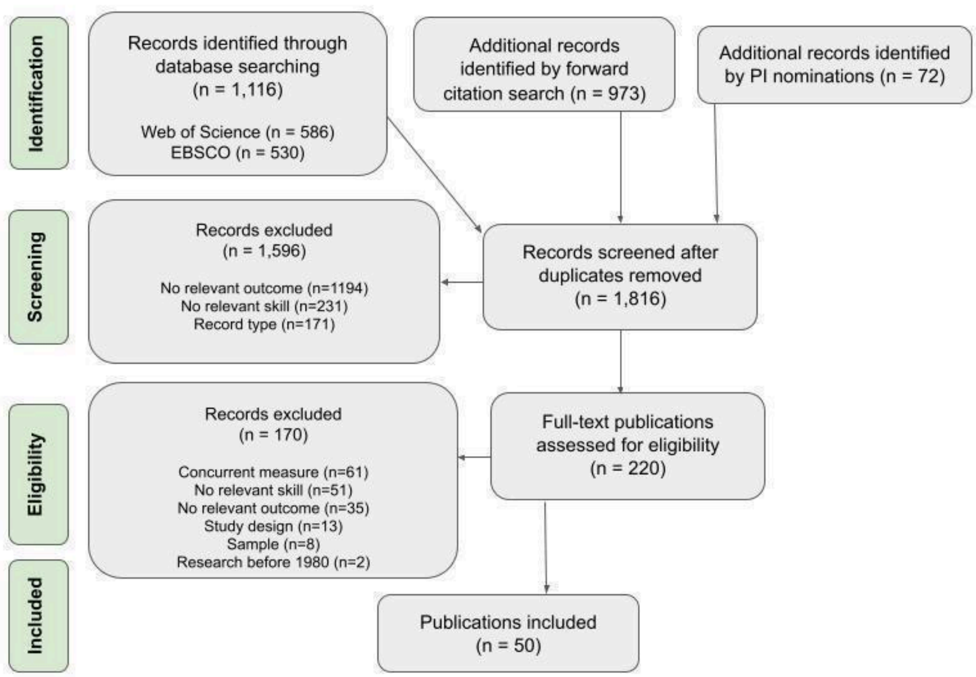
- Databases: Academic Search Elite, APA PsycArticles, Business Source Complete, Business Source Elite, Econlit, ERIC. Source types: academic journals, working papers, dissertations, reports, electronic resources
- Publication year: 1990 - 2022
- English language

The process included a primary search of electronic platforms and a complementary forward citation search using an online tool that tracks citation patterns. The primary search of platforms included EBSCO and Web of Science (WoS). Additional search parameters were applied in EBSCO, which are detailed in Box 1. The forward citation search, a feature in WoS, identified papers that cited a review (Roberts et al., 2007) deemed relevant due to its focus on the predictive validity of personality traits and use of prospective longitudinal studies for evidence.

Second, we assembled a group of experts and solicited recommendations for papers that were likely to meet the listed criteria. The experts were researchers who successfully responded to a competitive open call for grant funding, described below, to conduct a new round of this type of research.

The primary search process produced 1,116 records in EBSCO and Web of Science. An additional 72 papers were nominated by experts and the forward citation search produced 973 records. After removing duplicates, 1,816 records were identified in total. As illustrated in Figure 1, the selection procedure included two primary steps. We performed a preliminary screening of abstracts to exclude any records unrelated to the predictive validity of soft skills on labor outcomes, as well as records that did not include full-text, such as editorials, conference abstracts, theses, and meeting papers. In doing so, we excluded 1,596 records.

**Figure 1. Flow Diagram for Selection of Eligible Papers for Review**




Next, we reviewed the full text of the remaining 220 papers to screen more carefully for eligibility. The following information was systematically extracted from all full-text publications using a standardized form created by the authors: skill, outcome, target population, sample size, year/s of data collection, number of times skill/s measured, number of times outcomes measured, time span between the measurements, measure type, study type, sample selection and results. Exclusion criteria included irrelevant skill and/or outcome, sample size less than 50, final round of data collection before 1980, and if the skill and outcome were measured concurrently. A single-time measurement is adequate for concurrent validity, but not predictive validity which is the focus of this research synthesis. We excluded studies where the papers did not report enough detail to know the timing of the soft skill measurement relative to the outcome measurement.

Systematic reviews and meta-analyses were evaluated using the same criteria, and if excluded, their content was searched for eligible primary studies.

After comparing the set of studies that survived this search process with papers that were nominated by experts, we concluded that these search terms were too restrictive, potentially excluding eligible studies from the economics literature. Several papers published in economics journals are relevant for predictive validity but do not use such terminology in keywords or abstracts. To address this, we conducted a forward and backward citation search based on citations to and from Heckman (2006). Screening these additional sources for eligibility left us with 50 papers for this review, which were then coded by the authors. The full list of the included studies is available in the Appendix.

A quantitative meta-analysis was not possible because studies did not report predictive validity with sufficient detail to produce a common metric. Studies differed in whether they reported correlation coefficients, regression coefficients (standardized or not standardized, with rich covariates or none), changes in R-squared, or other metrics. They also differed on model specification in important ways that can affect the interpretation such as the use of control variables, specifying the key explanatory variable as the skill measure, the change in measure, or the independent effects of the measure on outcome levels or growth in outcomes if measured more than once. Instead, this study presents a narrative review, with vote-counting among subgroups of studies that focus on a common soft skill or measure type. Unless otherwise noted, regression coefficients significant at the 5% level or higher are reported in the text.

## C. Included studies

Applying specific criteria identified 50 studies eligible for this review. They are listed in the references section in bold typeface. There was diversity in sampled populations, including broadly representative samples and specialized professions. One study was conducted in a lower-middle-income country, while the rest took place in high-income countries, primarily in Europe and North America. Sample sizes range from 66 to 16,780.

The Big Five personality traits were the most prevalent skills measured in this literature. Table 1 shows the breakdown of skills that were identified in the studies we included. Some studies measured multiple skills across multiple samples resulting in more than 50 entries in the tables below.

**Table 1. Skill Types Included in Reviewed Studies**

| Soft Skill Included | Number | Percentage |
|---|---|---|
| Big Five | 21 | 28 |
| Positive Self Concept, Aspirations & Motivations | 15 | 20 |
| Interpersonal | 14 | 19 |
| Responsibility | 7 | 9 |

| | | |
|---|---|---|
| Stress/Anxiety | 6 | 8 |
| Grit/Task Persistence | 5 | 7 |
| Higher Order Thinking | 4 | 5 |
| Other/Emotionality | 3 | 4 |
| **All** | **75** | **100** |

Note: Table counts all skills, not studies. Some studies contained multiple skills.

These skills were most often (71%) measured using self-reports (Table 2) and 61% of the papers were found in psychology journals, although several fields and disciplines are represented (Figure 2). Among the labor market outcomes we selected for, job performance and income were the most common (39%), followed by employment (22%). See Table 3. Job performance was measured in a variety of ways across studies, often dependent on the job itself.

**Table 2: Measurement Approaches**

| Measurement Approach | Number | Percentage |
|---|---|---|
| Self-report | 61 | 71 |
| Observer-report | 16 | 18 |
| Alternative: Performance/task-based, vignettes | 5 | 6 |
| Artifact coding and interviews | 4 | 5 |
| **Total** | **86** | **100** |

**Figure 2. Distribution of Included Studies by Journal Type**

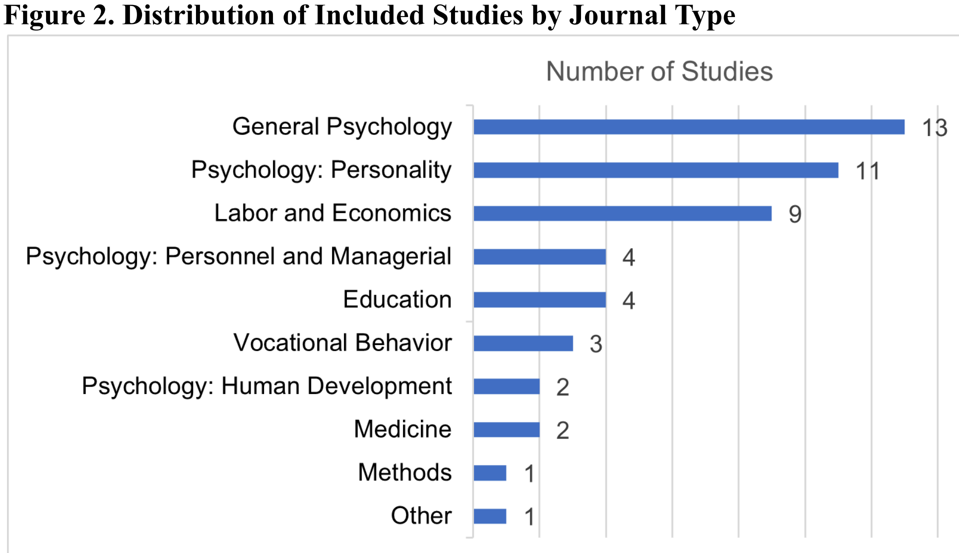


**Table 3: Outcomes Included in Reviewed Studies**

| Outcome | Number | Percentage |
|---|---|---|
| Job performance | 37 | 39 |
| Income/wages | 37 | 39 |
| Employment | 21 | 22 |
| **All** | **95** | **100** |

## D. Emerging research

In addition to reviewing published literature, we also incorporate findings from a set of new research projects that we commissioned. In 2021, we issued a competitive call for proposals to support researchers to integrate analysis of predictive validity into existing studies of soft skill training interventions or studies that measured both soft skills and economic outcomes. The call was issued to researchers working in low- and middle-income countries. We received 19 proposals and assessed them in relation to a set of predetermined criteria, such as the use of innovative soft skills measures, the potential usefulness and feasibility of these measures, and inclusion of relevant labor market outcomes. We selected five projects to support.

We identified two studies in addition to the five competitively-sourced emerging studies that were poised to generate similar results and included them in this set of emerging research. One is a 9-year follow-up of an RCT evaluation of the Skills for Effective Entrepreneurship Development (SEED) program in Uganda. The other is an initiative which assessed the impact of soft skills training initiatives in Tanzania, and accordingly presented a unique opportunity to assess predictive validity within an experimental design. The ability to evaluate a causal link between soft skills and labor market outcomes also introduces some methodological challenges and choices. Insight from these two studies and the commissioned studies are presented separately from the narrative literature review because these are not yet published or peer reviewed. Also, the emerging evidence we cite is based in low- and middle-income countries, in contrast with the published literature in this review, which is predominantly from high-income countries.

## E. Analysis: How is predictive validity measured?

There is no single way to determine whether a measure predicts future outcomes. The simplest and most common approach is the bivariate correlation between the predictor and outcome. Another is to produce a multivariate regression coefficient with the predictor and other conditioning variables as covariates. For example, one might regress age-30 earnings, or its logarithm, on adolescent soft skills at baseline with controls for parents' education and income at baseline or even controls for baseline scores on math and reading tests.

The distinction between pairwise correlation and partial (regression-adjusted) correlation is important for interpretation. Heckman et al. (2006) model the relationship between soft skills (referred to as non-cognitive skills) with labor market outcomes as operating through educational attainment and occupational choice. A simple bivariate correlation might capture the more complete association, but it should not be interpreted as causal for reasons noted in Section 1. Many researchers are interested in the additional variation that can be explained by measures of soft skills, above and beyond what is predicted by academic performance or educational attainment, even if the soft skill is generating the increase in these precursors to successful job-related outcomes or if mutual determinants of soft skills and labor market outcomes are the cause of stronger earnings, employment, or job performance.

As noted in Section 1, there is ideally exogenous variation in soft skills that allows us to treat the correlation or regression coefficient as the causal relationship between a unit change in the soft skill measure and a unit change in the outcome. We can achieve this through a randomized controlled trial (RCT) where study subjects are randomly assigned to an intervention that is able to shift those skills. In the absence of exogenous variation, examination of the relationship between changes in skill measures and changes in outcomes may provide insights most relevant for program design.

Another approach that we see in some of the older literature is to report the percentage of variance in the outcome that is explained by adding soft skill measures to a regression model or an analysis of variance model with other predictors included. That is, researchers may report the change in the R-squared value in a stepwise regression. Unfortunately, the different ways of reporting this kind of evidence are not standardized.

There are two challenges for our synthesis. The first is to compare across different types of statistics reported. For example, a correlation of 0.20, which is bounded by -1 and 1, is not equivalent to an unbounded standardized regression coefficient of 0.20, which means that a one-standard deviation change in the measure is associated with a 0.20 standard deviation change in the outcome. These metrics capture different things.

The other challenge is to interpret the magnitude of the effects. How large a correlation is considered adequately "predictive" of outcomes? How much change in the explanatory power should we expect? We do not currently have reliable benchmarks for judging whether a prediction is sufficiently good to use for policy applications. Given that interest in soft skills and their potential association with labor market outcomes is often framed in relation to traditional academic skills such as literacy and numeracy, one option would be to compare the magnitude of a correlation or coefficient for soft skills to that of these "hard skills," but this literature is also complex and nuanced without a clear standard for a meaningful association or how to account for the correlations between different skills and between skills and schooling completion.

## III. Findings

### A. Findings by measure approach

One of the most basic questions related to soft skill measurement is whether to use survey-based self-reports, observer reports, or task-based performance measures. These measure types are summarized in Table 4. For adults and children old enough to complete questionnaires, survey-based self-reports are the most widely used, likely because they are easiest to administer and standardize. The Big Five personality questionnaires, discussed in more detail below, are among the oldest and most common of these. The main criticisms of self-reports are their vulnerability to reference bias and social desirability bias. Reference bias is the error introduced by different respondents evaluating their ability levels relative to those around them, often valuing the endpoints of a subjective scale differently. Social desirability bias comes from the tendency of individuals to answer questions in a way that will be viewed favorably by others.

Observer reports are often similar to self-report surveys, but they ask another person such as the focal individual's teacher, parent, co-worker, or supervisor to answer questions about their skills, personality, or behavior. These observer reports can also suffer from reference bias and they vary in the amount of exposure the observer has to the individual being assessed.

The third measure type we consider, is performance and/or task-based, which includes scenario-based assessments, or more objective tasks where the respondent must demonstrate a skill by performing it or identifying an appropriate response to a hypothetical scenario. Scenario-based measures, such as situational judgement tests and vignette-anchored tasks, often consider specific behaviors, and allow for the inclusion of the conceptual breadth of a soft skill while limiting the context in which the skill is being evaluated. Thus, they improve content validity and limit reference bias but can be difficult to design. Performance-based tasks assess actual behaviors in response to a game or activity. A well-known performance-based assessment is the "marshmallow test" of delayed gratification, in which a child is offered a marshmallow and told they can eat it right away or wait for a period of time to receive two marshmallows (Mischel, 2014).

There is a fourth measure type that we did not explicitly contrast with these three, which is artifact coding. This type of measure requires review of documents, such as resumes, course transcripts, or work products. Trained coders would apply a rubric for translating aspects of the artifacts into scores on some trait measure. The metric can combine multiple coded artifacts as well. For example, one of the studies in our review constructed a measure of grit from reviews of resumes indicating persistence in jobs as well as patterns of course taking on a transcript that demonstrate grit through willingness to take on especially challenging courses. Studies using artifact coding as a measurement type were too rare to be able to compare with the other measure types.

**Table 4. Measure Approaches**

| | **Measure Approaches** | | | |
|---|---|---|---|---|
| | **Self-report (*Forced-choice*)** | **Observer report** | **Alternative: Performance/task-based, vignettes** | **Artifact coding** |
| Examples (instruments) | Big Five Inventory survey of 44 questions; *Multidimensional forced-choice scale* | Parent, teacher, co-worker, or supervisor ratings | Mirror Tracing Measure of Persistence; Situational judgment test (Tacit Knowledge Inventory for Managers); Vignette-anchored task | Resumes, interviews, work portfolios, job performance records |
| Examples (questions/ items) | Conscientiousness: "I see myself as someone who does things efficiently" *"Which of the following best describes your personality? (Select one)* | Agreeableness:"I see this person as someone who is helpful and unselfish with others" | "Trace the diagram of a star while looking at your hand only as a reflection in a mirror" "You realize that your profits are lower than last month. Select the most effective and least effective course of action" | "Evidence of at least two instances of multiple-year involvement in an activity" "Frequency of workplace collaborations recorded in team logs" |
| Strengths | Easy to standardize and administer *Reduces response bias, harder to game* | Easy to standardize and often easy to administer, less subject to biases of self-report | Face validity, harder to game Context specific | Objective, avoids self-report bias, reflects real world performance |
| Weaknesses | Response bias (intentional or unintentional): social desirability, acquiescence, reference, response styles, overconfidence; Easy to game *Respondents may find it harder to answer, more complex to analyze* | Requires finding respondent population, biased by beliefs and discrimination, limited opportunities to observe behavior, some skills less observable, response bias | Difficult to generalize across contexts, often measure narrow domain of a given skill, high fidelity measures often have intensive time and technical requirements | Limited availability of relevant artifacts, may not capture effort or intention, can be context-dependent |
| Prevalence | High *Low* | Somewhat low | Low | Low |

Source: Authors' elaboration.
* These are primarily measured using a likert scale.

Because the estimates of predictive validity depend on so many contextual factors and design and analytic choices made by study authors, we focus on within-study comparisons of measure approaches here, which narrows the field considerably. Our search criteria identified four studies that allowed for some type of comparison across measure approaches.

The four published studies are summarized in Table 5. All four include a self-report. Two used job performance as an outcome measure and the other two used income as an outcome measure.

Two studies are based on schoolchildren followed into adulthood, one is military, and one is with call center employees.

The evidence on measure type is, unfortunately, mixed. Observer reports demonstrated slightly greater predictive validity than self-reports. A self-report and a forced-choice measure were significantly related to job performance.The authors of the source studies did not specifically test whether the magnitude of the self-reports were larger or smaller than those of the alternatives (task-based or observer-reported measures), so we report in Table 5 which were reported to be statistically significant, with p=0.05 being the typical cutoff. Given that the evidence is inconclusive with respect to measure type, we look to the specific skill types.

**Table 5. Studies comparing more than one measurement approach**

| Study | Population | Measure approaches | Conclusion |
|---|---|---|---|
| Connelly et al. | 422 South Korean military cadets | Self-report (Big Five) vs. Peer ratings (Big Five) | 1 of 5 self-rated and 2 of 5 observer-rated Big 5 traits significantly related to job performance. |
| Luan | 186 German children followed into adulthood | Self-report (Big Five) vs. Parent- and Peer-report (Big Five) | None of the self-reported Big-Five traits measured at age 12 or 17 was significantly related to income. Parent- and friend-reported Big 5 trait measures were significant for 1 of 5 at age 12 (friend report) and 1 of 5 at age 17 (father report) |
| Le | 442 rural Iowa children followed into adulthood | Self-report (MPQ-BF) vs. Parent-report (MPQ) | 2 of 4 self-reported traits and 1 of 4 of the same traits reported by parents were significantly related to income. |
| Goffin | 68 call center employees | Self-report vs. forced-choice scale to measure dependability | Both self-report and forced-choice scale significantly correlated with job performance. |

Source: Authors' elaboration.

## B. Findings by skill type

Due to the heterogeneity in the data and outcomes reported, statistical pooling was not possible. Instead, we took a narrative approach to synthesizing the included studies, grouping studies by the skills that they examined.

### *1. Big Five personality traits*

The Big Five Framework has long been the most well-validated and widely used model of personality traits (Costa & McCrae, 1992), and several papers have examined its predictive validity. This model represents personality at the broadest level of abstraction, with each of the five traits summarizing additional distinct personality characteristics (John & Srivastava, 1999). The five personality traits are most often labeled as: (i) Openness, which is associated with being imaginative, creative, curious and unconventional; (ii) Conscientiousness, which is associated with being systematic, goal-oriented and self-disciplined; (iii) Extraversion, which is associated with sociability and being active; (iv) Agreeableness, which is associated with altruistic, sympathetic and trusting tendencies; and (v) Neuroticism, which is associated with anxiety, worrying and emotional instability. Neuroticism is sometimes coded on a reversed scale and labeled as Emotional Stability (Alderotti et al., 2021). Instruments frequently used to measure the Big Five include the 240-item Revised NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992), 60-item NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992), 100-item trait descriptive adjectives (TDA; Goldberg, 1992)  and 44-item Big-Five Inventory (BFI; (John et al., 1991), where NEO stands for Neuroticism, Extraversion, Openness. The NEO PI-R is the most comprehensive instrument, measuring the Big Five domains and six specific facets within each dimension, and takes roughly 45 minutes to complete. The BFI, NEO-FFI, and TDA are abbreviated instruments and take approximately 5, 15 and 15 minutes to complete, respectively (John & Srivastava, 1999). A number of other instruments have been developed and used, often for specific research purposes (John & Srivastava, 1999).

A large body of evidence shows that the Big Five personality traits are correlated with labor outcomes such as earnings, employment, and performance (Almlund et al, 2011; Borghan et al, 2008; Oh, Wang & Mount, 2011; Ones, Dilchert, Viswesvaran & Judge, 2007; Palczyńska & Swist, 2018). In particular, conscientiousness and emotional stability (i.e. neuroticism reverse coded) are consistently strong predictors of job performance and wages across jobs and criteria (Barrick & Mount 1991; Barrick, Mount & Judge, 2001; He, Donnellan & Mendoza, 2019; Hoff et al, 2021; and Hurtz & Donovan, 2000; Salgado, 1997 as cited in Oh et al., 2011). Other personality traits, such as openness, agreeableness, and extraversion, show some correlations with specific performance measures and occupational outcomes, but these findings are often inconsistent and not easily generalizable (He et al., 2019; Oh et al., 2011; Wilmot et al., 2019). Notably, the majority of evidence examining these relationships rarely addresses the timeline of measurement, specifically whether the skill measure preceded the outcome measure.

Our search process identified 21 studies that assessed the predictive validity of Big Five measures on labor force outcomes. One of these studies assessed facets of the Big Five - aspects of the broader five traits - on labor force outcomes, using the NEO-PI-R instrument. Six studies used the NEO-FFI instrument; three studies used NEO-PI-R; two studies used BFI and two studies used the BFI short (BFI-S) instrument. Six studies used other instruments, most of which were developed for the study. All but two studies relied on self-reports alone. Connelly et al., (2021) used the Trait–Reputation–Identity (TRI) Model (McAbee & Connelly, 2016) to assess the effects of self and observer ratings in understanding how personality characteristics predict performance. The trait factor accounts for both the self and observer reports of personality; the identify factor draws solely on the self-report and the identity factor reflects multiple observer reports, independent of the self-report.  In addition to self-reports, parents and a friend rated the participant’s personality in Luan et al., (2019).

To synthesize the included studies, we group the results according to the sign and level of significance of effects between each of the Big Five traits and pooled labor outcomes, which includes job performance, employment and income. The results for emotional stability were reverse coded and included with neuroticism. To allow for proper pooling, the results for negative labor outcomes, including unemployment and counterproductive work behavior, were reverse coded. From this procedure we do not see a consistent correlation between any of the Big Five traits and labor outcomes. However, we do see some patterns in the directionality of the relationships. Conscientiousness and extraversion have a positive relationship with labor outcomes, though these are often not statistically significant. Neuroticism and agreeableness generally have a negative relationship with labor outcomes, and openness is mixed in terms of direction.

**Table 6. Relationship of Big Five with labor outcomes**

| **Trait** | **Distribution of studies by sign and significance level** | | | |
|---|---|---|---|---|
| | **Positive, Significant** | Positive, Insignificant | Negative, Insignificant | **Negative, Significant** |
| Conscientiousness (n=17) | **26%** | 68% | 3% | **3%** |
| Openness (n=16) | **6%** | 43% | 51% | - |
| Neuroticism (n=14) | - | 27% | 50% | **23%** |
| Agreeableness (n=14) | - | 38% | 50% | **25%** |
| Extraversion (n=15) | **12%** | 52% | 36% | - |

Note: Regression coefficients, correlation coefficients, odds ratios and average marginal effect are included in this table.
Source: Authors' elaboration.

Given the volume of studies assessing Big Five, we also grouped results with a focus on the time span between skill measurement and outcome measurement to see if any correlation exists. Our time span groups include 3 -11 months, 1- 3 years and 8 - 50 years. As with the previous analysis, we do not see a consistent correlation between any of the Big Five traits and labor outcomes (pooled) based on the time span between skill and outcome measurement, with the exception of Neuroticism in 3-11 months. In general, the predictive power of these measures appears to be stronger for longer term outcomes, in which case conscientiousness has a mostly positive relationship with labor outcomes, and neuroticism has a primarily negative relationship with labor outcomes. Evidence on other traits is decidedly mixed.

**Table 7. Time span between Big Five and labor outcomes measurement: Vote counting**

| **Time Span** | **Trait** | **Distribution of studies by sign and significance level** | | | |
|---|---|---|---|---|---|
| | | **Positive, Significant** | Positive, Insignificant | Negative, Insignificant | **Negative, Significant** |
| 3-11 months (n=5) | Conscientiousness | **17%** | 67% | - | **17%** |
| | Openness | - | 17% | 83% | - |
| | Neuroticism | - | 40% | - | **60%** |

| | | | | | |
|---|---|---|---|---|---|
| | Agreeableness | - | 40% | 60% | - |
| | Extraversion | - | 40% | 60% | - |
| 1-3 years (n=8) | Conscientiousness | **31%** | 62% | 8% | - |
| | Openness | **15%** | 31% | 54% | - |
| | Neuroticism | - | 20% | 50% | **30%** |
| | Agreeableness | - | 50% | 17% | **33%** |
| | Extraversion | - | 50% | 50% | - |
| 8-50 years (n=7) | Conscientiousness | **27%** | 67% | 7% | - |
| | Openness | - | 56% | 44% | - |
| | Neuroticism | - | 27% | 60% | **13%** |
| | Agreeableness | - | 14% | 57% | **29%** |
| | Extraversion | **38%** | 63% | - | - |

Note: Regression coefficients, correlation coefficients, odds ratios and average marginal effect are included in this table.
Source: Authors' elaboration.

There were a variety of samples within the included studies, including specialized professions that may not reflect the general population. We grouped results between the Big Five traits and pooled labor outcomes for each of the unique population groups, which include students (undergraduate and medical), employees (sales representatives, managerial employees and transit operators), job seekers (females re-entering the labor force, job applicants and unemployed), military and/or armed forces and the general population. We do not see a consistent pattern based on the population though most population groups are only supported by two to three studies. However, it is notable that results are not significant for several studies focused on job seekers and employees.

Each of the Big Five traits has its own set of facets, which offer a more nuanced understanding of the broader traits. A single study in French-Quebec examined the relationship between Big Five facets and job performance amongst two occupation samples (Denis et al., 2010). As such, the results were not included in the previous vote counting exercises and are instead summarized here. Amongst specialized workers in technical trades, two facets positively and significantly predicted task performance: self-consciousness, a facet of Neuroticism (0.41), and competence, a facet of Conscientiousness (0.34).[1] Excitement-seeking, a facet of Extraversion (-0.26), negatively predicted performance. Amongst professional workers, straight-forwardness, a facet of Agreeableness (0.34) positively predicted task performance, while anxiety, a facet of Neuroticism (-0.27), negatively predicted it. These results suggest that the predictive validity of

[1] This study reported correlation coefficients.

measures of Big Five facets varies with job type. Viinikainen et al. (2010) began a longitudinal study in the 1960s before the Big Five framework was available. The measures of child personality include extraversion, which assessed how outgoing and energetic a child was, and constructiveness, which refers to the ability and inclination to positively contribute to situations. Given the similarities with the Big Five traits we have included them here. Both constructiveness (0.22) and extraversion (0.15) had positive and significant correlations with income at age 43.[2]

### 2. *Interpersonal skills*

While *interpersonal skills* include a vast array of soft skills, they are not commonly measured and were present in only 14 of the studies included in this review. Among these, 10 included samples with significant positive relationships with labor outcomes. However, there was no observable pattern related to the occupations of the populations tested. A measure of charisma based on the NEO PI-R facets was significantly predictive of several labor outcomes for college alumni 15 years later: income (0.25), number of subordinates (0.32), management level jobs (0.12), contextual performance (0.28), adaptive performance (0.42), but not task performance (Vergauwe et al., 2017). Three additional studies examined the predictive validity of interpersonal skill measures for income. A measure of communal positive emotionality (an orientation to interpersonal relationships and the tendency to experience positive affect) self-rated or by parents in rural Iowa had no predictive validity for income 9 years later (Le et al., 2014).[3] Among senior managers at a Fortune 50 firm, measures of customer and external relations, staffing, and communication and climate setting were found to predict trends in income, sales, profits, and job performance (Russell, 2001). Carneiro et al (2007) found a one standard deviation increase in social skills measured at age 11 by teachers to predict a 3 percentage point increase in the likelihood of being employed employment status and a 3 percentage point increase in hourly wages at age 42.[4]

For individuals in sales, several measures of interpersonal skills have been found to predict short term job performance: impression management (Ispas et al., 2014), socialization as measured by the California Psychological Inventory (0.25), and interpersonal orientation as measured by the Telemarketing Applicant Inventory (0.23), and communicator competence as measured by the same inventory (0.27) (Hakstian et al., 1997).[5] However, among primarily male insurance salesmen in the UK, a similar measure of socialization based on the Californian Psychology Inventory was indicative of effort (0.17) but negatively correlated with sales (-0.18), though effect sizes were small to moderate (Corr & Gray, 1995).

Using the Chinese Personality Assessment Inventory (CPAI-2), Ion et al. (2016) examined the predictive validity of personality dimensions deemed culturally-specific, such as graciousness and relationship-orientation, over universal predictive traits, such as cognitive ability and conscientiousness on job performance. Among 142 Chinese and 218 Romanian workers in a Romanian textile production company, measures of graciousness (0.28) and relationship

[2] This study reported correlation coefficients.
[3] This study reported correlation coefficients.
[4] This study reported marginal effects.
[5] Both studies reported correlation coefficients.

orientation (0.26) were significantly predictive for the Chinese sample, but not the Romanian sample one year later. However, they added only marginal predictive power over universal predictive traits.

A study of 2010 police candidates in Spain (Forero et al., 2009) examined their job performance (as measured by supervisor ratings, performance ratings, commendations, and sanctions) one to six years after training.[6] Here, the researchers discovered a small relationship between performance and self-reported measures of interpersonal relations, leadership, and sociopathy (-0.04 to 0.10). However, observational ratings of interpersonal skills were highly correlated. Observational measures examined adaptation to norms, group integration, social skills and tolerance/flexibility (0.40 to 0.51). In one study of teachers, an interview-based rating of leadership had no relationship with teacher effectiveness or retention (Robertson-Kraft and Duckworth, 2014).[7] In a study of undergraduate business students, assessment center exercises to assess relationship management, interpersonal skill competencies, problem-solving and decision-making skills were predictive of salary (0.24) and promotions (0.29) (Waldman and Korbar, 2004).

Three studies focused on the medical field. Interpersonal video-based SJTs were moderately predictive (0.15) of performance as a general practitioner in Belgium, but results were not robust to various corrections (Lievens et al., 2013).[8] For medical students, self-reported measures of social confidence and tolerance were not found to robustly predict the academic or clinical performance in Canada (McLarnon et al., 2017). Among junior doctors, a situational judgement test targeting empathy, integrity and coping with pressure was predictive of selection center results (0.58), supervisor reviews one year into practical GP training (0.56), scores on an applied knowledge test (0.69), and performance during clinical simulations (0.57) (Patterson et al., 2013).

### *3. Grit or task persistence*

The personality trait referred to as "grit" is a relatively new construct, but it has received a lot of interest and attention, particularly in the field of education. Grit is defined as "perseverance and passion for long-term goals" (Duckworth, Peterson, Matthews, & Kelly, 2007, p. 1087). It is typically conceptualized as a higher order construct consisting of two lower-level facets: "perseverance of effort" and "consistency of interest," which are reflected in the two subscales of the most popular self-report inventories used to measure grit: the Grit Scale and the Short Grit Scale (Credé, Tynan & Harms, 2017; Duckworth et al., 2007, Duckworth & Quinn, 2009). Proponents of grit argue that its measures are highly predictive of success, possibly even more so than cognitive ability, and that grit provides unique information distinct from conscientiousness, the Big Five factor with which it is often highly correlated (such as 0.64 and 0.74 for the Consistency of interest and Perseverance of effort subscales, respectively (Duckworth and Quinn 2009)). Critics argue that the proposed conceptual structure of grit, its meaningful distinction

[6] This study reported correlation coefficients.
[7] This study reported bivariate correlations.
[8] This study reported correlation coefficients.

from conscientiousness, and its predictive power should all be questioned (Credé, Tynan & Harms, 2017).

While there has been a significant body of research on grit in relation to academic performance and retention (a 2017 meta-analysis included data from 73 studies representing 88 unique samples (Credé, Tynan & Harms, 2017)), there appears to be far less research on the predictive validity of grit in relation to labor market outcomes.

Our search process identified only four studies. Three of these studies used the Short Grit Scale, two as a total score and one as the scale's two separate subscales, perseverance of effort and consistency of interest. This latter study, conducted among Polish adults, found positive associations between grit and education attainment, as well as other life outcomes such as trust, while controlling for Big Five personality traits and numeracy; it found no relationship between grit and economic outcomes, measured as labor force participation, employability, or wages (Palczyńska & Świst, 2018). The study that used the Short Grit Scale as a total score looked at retention in relation to staying in a job (sales representatives) or finishing a course (Army special operations forces course participants) and across both populations found that individuals with higher scores on the scale were more likely to stay in their jobs and finish the course (Eskreis-Winkler et al., 2014).[9]

The final two studies identified in our search focused on teachers. The first used raters to review teachers' resumes and rate grit from information on college activities and work experience using a 7-point rubric (Robertson-Kraft & Duckworth, 2014).[10] The study investigated whether these scores predicted first time teachers' retention (if they would remain in their jobs through the end of the year or resign midway) and effectiveness (if their students mastered at least 70% of content on the standardized achievement test). Across two populations of teachers they found that grit predicted teacher effectiveness, and for one population, retention as well. Teachers who were one standard deviation higher on the grit rating were more than twice as likely to complete the school year than teachers with lower grit rating. (There was not sufficient variation in retention in one population to draw any conclusions, as in that sample 99% of teachers completed the school year). The second study assessed novice teachers' grit before the start of the school year and evaluated their effectiveness based on students' academic gains at the year's end (Duckworth et al., 2009). In both independent and simultaneous models, grit predicted teacher effectiveness. Teachers who were one standard deviation higher in grit were 31% more likely to outperform their peers in the independent model, and 23% more likely in the simultaneous model.

One additional study addressed related concepts. Andersson, Loven & Bergman (2014) investigated "task persistence," using teachers' ratings of 13 year old students. The highest level of task persistence described students as, "They have a marked ability to concentrate on a task and persevere with it. They never allow themselves to be distracted and do not give up as long as a task suits their level of intelligence." On its own task persistence did not predict occupation and income 30 years later, but did as a composite with other skills such as educational aspirations (Andersson, Loven, & Bergman, 2014).

[9] This study reported odds ratios.
[10] This study reported bivariate correlations.

### 4. *Positive self-concept, motivations and aspiration*

Although certain conceptualizations of skills do not include beliefs, values, or attitudes, we took an inclusive approach to our review for three reasons. First, these concepts are often related to or predictive of other skills. Psychological literature, for example, links self-esteem to the ability to work productively in groups and perseverance in the face of adversity (Murnane et al., 2001). Understanding the predictive validity of beliefs, values, and attitudes might accordingly be a key step towards understanding the definition and predictive validity of other skills. Second, these concepts have been found to be malleable and responsive to intervention and are therefore potentially useful for a range of actors and intervention design and testing. And third, these concepts comprise an important part of existing research and practice, recognizing what is currently known and promoting areas for future research will be useful for the field.

Our search identified 15 articles across the interrelated areas of aspirations, motivations, and self-concept. A first conclusion is that much of the literature concerns very specific, sub-concepts rather than more broad or global views, for instance, the job aspiration to become a scientist rather than more general aspirations and innovative esteem ("extent to which individuals feel pride and worthiness in their incremental and/or radical innovative capabilities" Anwar et al., 2020) rather than more global views of self-esteem. All measures were based on self-reported measures. When more general concepts were identified, recognized scales and measures were used; in the case of unique and specific concepts, novel scales were developed.

Across all but one of these different studies, positive aspirations, motivations, and self-concept were found to be positively associated with various labor market outcomes, such as employment and job performance, but these associations were only statistically significant for eight studies and were often smaller in magnitude than other predictors. Among job seekers, confidence in job search self-efficacy outcomes was significantly predictive of job offers (0.34) (Saks et al., 2015). Innovative esteem was predictive of job performance (0.26) among employees working in tech and research and development organizations (Anwar et al., 2020). Self-acceptance among telemarketers was predictive of job performance (0.38)[11] (Hakstian et al., 1997).

Several studies have identified significant relationships between soft skill measures and labor market outcomes. For instance, job aspirations at age 16 to pursue a scientific occupation were strongly linked to career attainment in adulthood (Schoon et al., 2001). Teenagers aspiring to be health professionals, scientists, or engineers were 12.6, 3.38, and 1.54 times more likely, respectively, to achieve those careers. Similarly, educational aspirations among Swedish school children predicted their adult income, with a regression coefficient of 0.25 (Andersson et al., 2014). Enterprising vocational interests assessed at the end of high school were also found to predict income (0.11) a decade later (Stoll et al., 2017). Furthermore, self-esteem in high school boys[12] was a significant predictor of their average hourly wages seven years later (Murnane et al., 2001), and Drago (2011) found a similar association between youth self-esteem and earnings eight years later, with a coefficient of 0.04.

[11] This study reported correlation coefficients.
[12] Girls were not included in this study sample.

### 5. Higher order thinking

"Higher-order thinking skills" describes a set of socio-emotional skills, in high demand by employers, involving the ability to collect, synthesize, and think critically to draw conclusions from multiple sources of information (Lippman et al., 2015). It involves many sub-skills from problem solving to creativity to decision-making, and is distinct from cognitive ability, often measured by memory, attention, and processing. Our review uncovered four studies examining the predictive validity of ten higher-order thinking skills, six of which were found to have a significant relationship with job performance measures. With the exception of Russell (2001), all studies included in this section reported correlation coefficients, which ranged from 0.06 to 0.48.

Patterson et al. (2013) examined junior doctors applying for training in general practice. A clinical problem-solving test was predictive of selection center results (0.47), supervisor reviews one year into practical GP training (0.54), scores on an applied knowledge test (0.85), and performance during clinical simulations (0.55). For the previously discussed study of police candidates in Spain (Forero et al., 2009), measures had a minimal direct effect on job performance, though effect sizes increased substantially when examining their indirect effect mediated by training. A 6-item self-report measure of intellectual efficiency (from the Law Enforcement Assessment and Development Report- LEADR) had limited ability to indirectly predict variation in performance (0.10). However, trainer ratings of practical judgment were found to have an estimated effect size of 0.48 on performance. These ratings were particularly predictive of the likelihood of receiving a disciplinary sanction.

A study of 85 Canadian part-time telemarketers found that a measure of perceptual speed and accuracy (based on the Comprehensive Ability Battery) did not significantly correlate with sales but correlated with supervisory ratings (0.20) and overall performance (0.24). A measure of brainstorming, based on the Innovative and Divergent Elaboration Aptitudes (IDEA) battery was only significantly correlated with supervisory ratings (0.18) (Hakstian et al., 1997).

Finally, among 98 general managers at a top firm, Russell (2001) examined the predictive validity of interview-based measures of strategic planning, business understanding, product planning, organizational acumen, and financial analysis, among others. Product planning and organizational acumen were found to predict bonus increases and trends in performance, profits, and sales.

### 6. Stress, anxiety management and emotional instability

Our review revealed six papers examining the predictive validity of skills focused on stress, anxiety, and negative emotions. However, it is important to note that these skills are conceptually similar to Neuroticism in the Big Five. Correlations often had low effect sizes with low levels of significance. One study of police cadets in Spain (Forero et al., 2009) found that, among self-reported measures of several psychological constructs, the Clinical Analysis Questionnaire (CAQ;Krug, 1980) measure of anxiety had one of the largest indirect effects on supervisor ratings of job performance (0.15).[13] This contrasts with 0.02 for neuroticism and 0.06 for

[13] This study reported correlation coefficients.

emotional adjustment. In a voluntary response survey among unemployed individuals in the Southwest United States, Wanberg, Zhang, & Diehn (2010) found that stress and worry was negatively correlated with person-job fit at the 10 percent level (-0.12) among re-employed individuals.[14] However, it was not predictive of unemployment insurance exhaustion or reemployment status. Similarly, McLarnon et al. (2017) found no relationship between being "Calm-relaxed" and clerkship performance among 300 medical students. Groves (2005) found that teacher-measured aggression and withdrawal at age 11 are significant and negative predictors of wage at age 33 for women in the UK. Aggressive behavior, a measure of high activity and low self-control, assessed at age eight by teachers had no association with duration of unemployment or income in adulthood (Viinikainen et al 2010).[15] Le et al. (2014) found no relationship between self and parent-measured negative emotionality, which is the susceptibility to negative emotions such as anxiety, anger and general distress, and income.[16]

### 7. *Responsibility*

Studies indicate that responsibility-related traits such as maturity, dependability, self-control, and discipline are associated with labor outcomes, including income, employment, job performance, and occupational status, though effects vary by context, population, and specific characteristics assessed. In Spengler et al. (2018), the mature personality (ability to get work done efficiently and to accept assigned responsibility) of 1,952 high school students was measured as a predictor of income, which was assessed 50 years later. Mature personality was significantly associated with income (0.02). Among actively employed white women, external locus of control - the belief that outcomes are the result of fate or luck rather than their own actions - has a significant and negative influence on earnings (-0.023) over twenty years later (Groves, 2005).

Dependability, which was measured by responsibility and negatively-keyed risk-taking scales, was found to be negatively predictive of counterproductive work behavior whether measured conventionally or with a forced-choice scale (Goffin et al., 2011).[17] Among 142 Chinese and 218 Romanian workers in a Romanian textile production company, responsibility and discipline were significantly predictive of job performance for the Chinese sample (0.23 responsibility; 0.23 discipline), and the Romanian sample (0.34 responsibility; 0.19 discipline) one year later (Ion et al. 2016). Evidence from two nationally representative British cohorts (n=16,780) found that on average, a 1-SD increase in childhood self-control, as rated by teachers, was significantly associated with a 1.4 percentage point reduction in the probability of unemployment after adjustment for intelligence, social class, and gender (Daly et al., 2015). The predictive strength of childhood self-control was equal to or greater than that of intelligence.

A study that followed a Luxembourgish sample from childhood to middle adulthood found that students' responsibility defined as being industrious and achievement-striving, was predictive of higher occupational status but not of higher income in adulthood when self-reported or measured by teachers (Spengler et al., 2015). The characteristic 'rule breaking and defiance of parental authority' (0.12) was the best non-cognitive predictor of higher income after controlling for the

[14] This study reported correlation coefficients.
[15] This study reported correlation coefficients.
[16] This study reported correlation coefficients.
[17] This study reported correlation coefficients.

influence of IQ, parental SES and educational attainment (Spengler et al., 2015). This characteristic encompasses a low level of rule orientation, which may play a positive role in some contexts. For instance, individuals who scored high on this scale may be more willing to stand up for their own interests and aims when negotiating salaries or raises.

Le et al. (2014) examined the predictive validity of self and parent-measured constraint, which is associated with self-control and endorsement of traditional values, on income nine years later. While parent-measured constraint was not associated with income, self-reported constraint significantly predicted income (-0.11).[18]

### *8. Other/Emotionality*

We found predictive validity evidence for other skills that could not be grouped into the categories used above.

**Emotional reasoning.** A study of 84 working adults assessed the predictive validity of emotional reasoning skills on job performance (Blickle et al., 2009). Emotional reasoning skills refer to the ability to employ emotional knowledge to understand and analyze emotions. Respondents are given 12 situations in which they must select the feelings experienced by different target persons in each situation. Results demonstrated a significant negative relationship (-0.21) between emotional reasoning skills and job performance when performance was rated by a mix of superiors, peers, subordinates and others. There was no significant relationship when performance was measured by only superiors (-0.18) or only peers (-0.21). Emotional reasoning skills also explained an additional four percent of variance in job performance ratings beyond general mental ability and personality traits.

**Emotionality.** Le et al. (2014) examined the predictive validity of self and parent-measured personality attributes on income nine years later. Agentic Positive Emotionality, an energetic orientation to master achievement-related contexts and the tendency to experience positive affect in such settings, measured by self (0.13) and by parents (0.15) was significantly associated with income. This study reported correlation coefficients.

**Honesty-humility and emotionality.** Anglim et al. (2018) explored six personality factors as predictors of counterproductive work behavior (CWB), a measure of job performance, among job applicants and non-applicants. Four of those factors are covered in the section on Big Five, while Honesty-Humility and Emotionality are included here given they are composed of unique traits and characteristics. Honestly-Humility is associated with sincerity, fairness, greed avoidance and modesty. Emotionality is associated with fearfulness, anxiety, dependence and sentimentality. Neither factor was predictive of CWB for applicants and non-applicants.

## C. Findings from emerging evidence

As noted in the Methods section, to complement and advance the published literature where evidence largely comes from high income countries, we issued a call for proposal to identify on-going research projects that could be used to assess the predictive validity of measures of soft

[18] This study reported correlation coefficients.

skills in LMIC contexts. Of the seven studies we supported, we report on four here. (Two of the studies had not yet gathered follow-up data on labor market outcomes and one found that such outcomes showed no variation, because at the two month follow-up, none of the youth were employed.) We present this emerging evidence separately from the literature we reviewed above, because the studies are not yet published or peer reviewed and in some cases, the data relevant for predictive validity will not be included in the ultimate publication.

The four studies whose results are discussed here were conducted in Algeria, Uganda, Tanzania, and South Africa. They all included studies of training interventions meant to improve soft skills for young people. Table 8 describes the studies and the skills that each set of study authors measured.

Each study measured several different skills, but the skills were not necessarily the same or described using the same language. We categorized the skills into common groups in Table 8, where it shows that most skill types were measured in two or more studies. Two, the Uganda and South Africa studies, measured the Big Five personality traits. As noted above, the Big Five inventory is a widely used, well-known instrument with considerable agreement on how to describe and measure them. There was less consensus on how to characterize intrapersonal or interpersonal skills. For example, one study (Algeria) had a generic social skills scale, while the Tanzania study produced separate measures of seven specific social skills and the South Africa study just reported social awareness.

In order to compare results across studies, we extracted the common metric, which was a pairwise correlation between the skill measure and subsequent labor market outcomes. Several outcomes were examined in each study, but for comparability we focus on some version of earnings. In the case of the Uganda study of entrepreneurs, we used business profits. The authors of each study produced many more estimates of the relationship between soft skills and labor market outcomes, including use of other outcomes, sub-scales of the soft skill measures, change in soft skill measures, and subgroups of study participants, but we focus on just the pairwise correlation results in each study that can be compared across studies.

**Table 8. Characteristics of contributing studies to the emerging evidence on soft skill measures**

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| **Country** | **Algeria** | **Uganda** | **Tanzania** | **South Africa** |
| **Lead institution** | World Learning | University of California Berkeley | World Bank Africa Gender Innovation Lab | World Bank |
| **Lead Investigator** | Catherine Honeyman | Laura Chioda | Clara Delavallade | Samantha DeMartino |
| **Intervention** | Job skills training for youth | Entrepreneurship training for adolescents | Soft skill training for youth | Skillcraft |
| **Skills measured** | | | | |
| **Big Five personality traits** | (not measured) | Big Five | (not measured) | Big Five |
| **Intrapersonal, awareness** | (not measured) | (not measured) | Self awareness, Emotional awareness | Self awareness |
| **Self efficacy** | (not measured) | Self efficacy | Self efficacy | (not measured) |

| | | | | |
|---|---|---|---|---|
| **Intrapersonal, management** | Managing emotions<br>Reliability | Stress | Emotional regulation<br>Self control<br>Perseverance<br>Personal Initiative | Self management<br>Persistence |
| **Interpersonal** | Social skills | (not measured) | Listening, Empathy, Expressiveness, Relatedness, Influence, Negotiation, Collaboration | Social awareness |
| **Executive function** | Thinking and planning, Goal-setting | (not measured) | Problem solving/decision making | Executive function |
| **Mindset** | (not measured) | (not measured) | (not measured) | Growth mindset |
| **Other** | "Soft skills", Employability, Civic engagement, Connectedness, Cultural identity | (not measured) | (not measured) | Working memory, Nonverbal fluid intelligence, Attention, Cognitive processing speed |

Source: Authors' elaboration.

The results show small correlations between soft skill measures and labor market outcomes. Most of the correlations (93 percent of the 74 estimates) are less than 0.08. All of those that are greater than 0.08 are from the World Learning Algeria study's direct instructor observation measure (see Table 9).

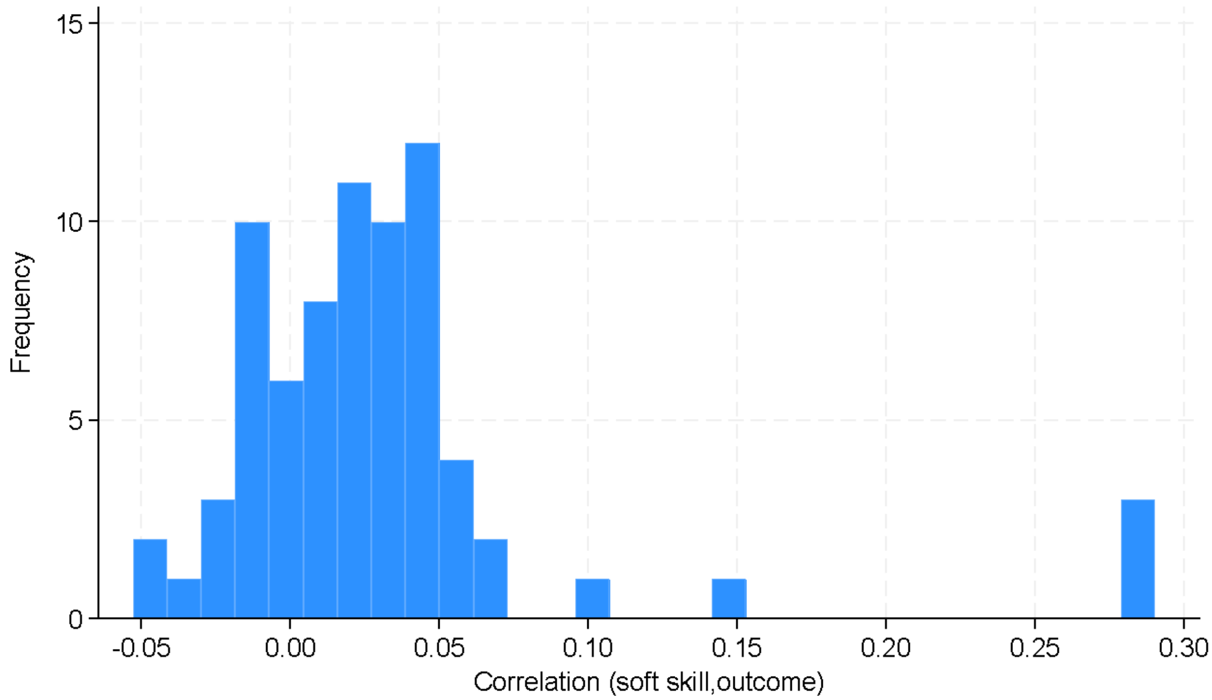
**Table 9. Correlations between soft skills and earnings, five largest values**

| Study | Skill name | Measure type | Correlation with earnings |
|---|---|---|---|
| Algeria | Managing emotions | Observation | 0.10 |
| Algeria | Social skills | Observation | 0.15 |
| Algeria | Reliability | Observation | 0.29** |
| Algeria | Thinking and planning | Observation | 0.29** |
| Algeria | Goal-setting | Observation | 0.29** |

* $p < 0.05$, ** $p<0.01$

The distribution of estimates illustrates how these are outliers. See Figure 3. The average correlation for each of the three measure types, and by study are shown in Table 10.

**Figure 3. Distribution of correlations between soft skill measures and labor market outcomes**




Note: N = 74 estimates

**Table 10. Predictive validity by measure type and study**

| | Average correlation with labor market outcomes | Number of correlations reported |
|---|---|---|
| **Measure type** | | |
| Self-report | 0.02 | 40 |
| Task-based | 0.01 | 29 |
| Observation | 0.22 | 5 |
| **Study** | | |
| Tanzania | 0.02 | 36 |
| Uganda | 0.02 | 7 |
| South Africa | 0.02 | 20 |
| Algeria | 0.10 | 11 |
| **All** | **0.03** | **74** |

## IV. Conclusions and Limitations

There are several limitations in this review. First, it is possible that some relevant research was not included in this review, even though we undertook an extensive search that included multiple databases, numerous search terms and hand searching reference lists of published papers. The criteria were strict and we had to exclude many papers because of insufficient reporting. For example, a number of papers failed to report when the skill and the outcome were measured and were thus excluded. Also, the keywords used in the search may not have captured the diverse ways that economists in particular refer to soft skills and thus excluded potentially important evidence. The language of predictive validity for measurement scales is common among experimental psychologists but not as common for economists. To address this, we intentionally sought expert nominations from economists or experts who are familiar with the literature on skill formation and skill measurement and did use these nominated papers and their forward and backward references to capture a more comprehensive set of studies in the review.

A second major limitation is that we were not able to conduct a quantitative analysis of the point estimates provided in the source studies. There was insufficient basis for putting the estimates of predictive validity on a common scale or common metric so we often summarized at the conclusion level or reported on simple pairwise correlations. As such, it was not possible to discuss the magnitudes of the predictive validity estimates in absolute terms.

Finally, we were unable to find sufficient convincing evidence in the literature nor was there evidence that was comparable across the commissioned studies that would allow us to make causal statements about the relationship between soft skills and outcomes.

Summarizing the published and emerging evidence, we showed that no single measure type or skill has been consistently shown to predict labor market outcomes. This includes self-reports, observer reports, and task-based measures and it includes measures of both widely used instruments like the Big Five inventory and its variants as well as highly customized instruments. The correlational evidence in early findings from a set of commissioned studies suggested that observer-based measures might be better predictors than self-reports and task-based measures.

These findings may be discouraging for practitioners who might be seeking guidance on which measures to use and whether they can be relied upon to serve as short-term proxies for longer-term outcomes in the context of impact evaluations or monitoring and evaluation efforts. They suggest that we cannot point to published or even recently generated evidence to justify these choices. Despite the growing interest in these measures, there is more work to do to prove which, if any, deserve the attention of providers of training services to young people or that of employers. In addition to the issue of measurement, there is need for more investigation into whether there is a causal association between soft skills and labor market outcomes. In particular, we outlined a straightforward approach that can be used to derive causal estimates of predictive validity from RCTs that generate exogenous variation in soft skills.

Based on the study limitation that we noted above about over-excluding, and also that many research studies do not focus on issues of measurement and accordingly do not report all relevant data, there is likely vastly more data than would meet the inclusion criteria for a formal review, but it is necessary to take extra steps to make those data usable. Therefore, a productive direction for the field would be to analyze the data from all the longitudinal studies for which such data can be obtained on soft skills and labor market outcomes specifically in relation to predictive validity and try to harmonize the data and analysis to a greater degree than was possible in this review. By doing so, and publishing results, it would provide a template that researchers, particularly developers and implementers of soft skill measures,  can use going forward to incorporate predictive validity analysis into their data collection and analysis. In that way, predictive validity analysis could become almost as routine as reporting point-in-time-measured psychometric properties such as reliability. Such work is required if we are to be guided by evidence rather than faith in the adoption of soft skills as an intermediate outcome of job and entrepreneurship training.

## References

*References in bold are included in synthesis*


Acevedo, P., Cruces, G., Gertler, P., & Martinez, S. (2017). Living up to expectations: How job training made women better off and men worse off (No. w23264). National Bureau of Economic Research. doi:10.3386/w23264

Ajayi, K., Das, S., Delavallade, C., Ketema, T., & Rouanet, L. (2022). Gender Differences in Socio-Emotional Skills and Economic Outcomes. World Bank Policy Research Working Paper, 10197.

Alderotti, G., Rapallini, C., & Traverso, S. (2023). The Big Five personality traits and earnings: A meta-analysis. Journal of Economic Psychology, 94, 102570. https://doi.org/10.1016/j.joep.2022.102570

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In Handbook of the Economics of Education (Vol. 4, pp. 1-181). Elsevier. doi: 10.3386/w16822

**Andersson, H., Lovén, J., & Bergman, L. R. (2014). The importance of high competence in adolescence for career outcomes in midlife. *Research in Human Development*, *11*(3), 204–216. https://doi.org/10.1080/15427609.2014.936180**

**Anglim, J., Lievens, F., Everton, L., Grant, S. L., & Marty, A. (2018). HEXACO personality predicts counterproductive work behavior and organizational citizenship behavior in low-stakes and job applicant contexts. *Journal of Research in Personality*, *77*, 11–20. https://doi.org/10.1016/j.jrp.2018.09.003**

Anvari, F., Alsalti, T., Oehler, L., Hussey, I., Elson, M., & Arslan, R. C. (2024, March 22). A fragmented field: Construct and measure proliferation in psychology. https://doi.org/10.31234/osf.io/b4muj

**Anwar, Ch. M. (2020). Innovative Esteem: Antecedents and Relationship with Job Performance. *E+M Ekonomie a Management*, *23*(4), 137–151. https://doi.org/10.15240/tul/001/2020-4-009**

Arias, O., Evans, D. K., & Santos, I. (2019). The skills balancing act in Sub-Saharan Africa: Investing in skills for productivity, inclusivity, and adaptability. World Bank Publications.

**Averett, S. L., Bansak, C., & Smith, J. K. (2021). Behind Every High Earning Man is a Conscientious Woman: The Impact of Spousal Personality on Earnings and Marriage. *Journal of Family and Economic Issues*, *42*(1), 29–46. https://doi.org/10.1007/s10834-020-09692-x**

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance. Personnel Psychology, 44, 1–26. doi:10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). The FFM personality dimensions and job performance: Meta-analysis of meta-analyses. International Journal of Selection and Assessment, 9, 9–30. doi:10.1111/1468-2389.00160

**Barron, L. G., Carretta, T. R., & Rose, M. R. (2016). Aptitude and Trait Predictors of Manned and Unmanned Aircraft Pilot Job Performance. *Military Psychology*, *28*(2), 65–77. https://doi.org/10.1037/mil0000109**

**Blickle, G., Momm, T. S., Kramer, J., Mierke, J., Liu, Y., & Ferris, G. R. (2009). Construct and Criterion-Related Validation of a Measure of Emotional Reasoning Skills: A two-study investigation. *International Journal of Selection and Assessment*, *17*(1), 101–118. https://doi.org/10.1111/j.1468-2389.2009.00455.x**

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. Journal of human Resources, 43(4), 972-1059. DOI: https://doi.org/10.3368/jhr.43.4.972

Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. Organizational Behavior and human decision Processes, 103(1), 84-103. https://doi:10.1016/j.obhdp.2006.09.001

Campos, F., Frese, M., Goldstein, M., Iacovone, L., Johnson, H. C., McKenzie, D., & Mensmann, M. (2017). Teaching personal initiative beats traditional training in boosting small business in West Africa. Science, 357(6357), 1287-1290. DOI: 10.1126/science.aan5329

**Carneiro, P., Crawford, C. and A. Goodman (2007). The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes, CEE Discussion Paper 0092.**

Chen, Y., Feng, S., Heckman, J. J., & Kautz, T. (2020). Sensitivity of self-reported noncognitive skills to survey administration conditions. Proceedings of the National Academy of Sciences, 117(2), 931-935. https://doi.org/10.1073/pnas.1910731117

Chioda, L., Contreras-Loya, D., Gertler, P., & Carney, D. (2021). Making entrepreneurs: Returns to training youth in hard versus soft business skills (No. w28845). National Bureau of Economic Research. DOI: 10.3386/w28845

**Connelly, B. S., McAbee, S. T., Oh, I.-S., Jung, Y., & Jung, C.-W. (2021). A multirater perspective on personality and performance: An empirical examination of the trait–reputation–identity model. *Journal of Applied Psychology*, *107*(8), 1352. https://doi.org/10.1037/apl0000732**

Corcoran, R. P., Cheung, A. C., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. Educational Research Review, 25, 56-72. https://doi.org/10.1016/j.edurev.2017.12.001

**Corr, P. J., & Gray, J. A. (1995). Attributional style, socialization and cognitive ability as predictors of sales success: A predictive validity study. *Personality and Individual Differences*, *18*(2), 241–252. https://doi.org/10.1016/0191-8869(94)00153-J**

Costa Jr., P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In *The SAGE handbook of personality theory and assessment, Vol 2: Personality measurement and testing* (pp. 179–198). Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n9

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. Journal of Personality and Social Psychology, 113(3), 492–511. https://doi.org/10.1037/pspp0000102

Cunningham, W. V., & Villaseñor, P. (2016). Employer voices, employer demands, and implications for public skills development policy connecting the labor and education sectors. The World Bank Research Observer, 31(1), 102-134.

**Daly, M., Delaney, L., Egan, M., & Baumeister, R. F. (2015). Childhood Self-Control and Unemployment Throughout the Life Span: Evidence From Two British Cohort Studies. *Psychological Science*, *26*(6), 709–723. https://doi.org/10.1177/0956797615569001**

**Denis, P. L., Morin, D., & Guindon, C. (2010). Exploring the Capacity of NEO PI-R Facets to Predict Job Performance in Two French-Canadian Samples: Capacity of NEO PI-R Facets to Predict Job Performance. *International Journal of Selection and Assessment*, *18*(2), 201–207. https://doi.org/10.1111/j.1468-2389.2010.00501.x**

**Drago, Francesco. "Self-esteem and earnings." Journal of Economic Psychology 32, no. 3 (2011): 480-488. https://doi.org/10.1016/j.joep.2011.03.015**

**Drasch, K. (2019). Behavioral Intentions, Actual Behavior and the Role of Personality Traits. Evidence from a Factorial Survey Among Female Labor Market Re-Entrants. *Methods, Data, Analyses*, *13*(2), Article 2. https://doi.org/10.12758/mda.2017.14**

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). Journal of Personality Assessment, 91(2), 166–174. https://doi.org/10.1080/00223890802634290

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. Journal of personality and social psychology, 92(6), 1087. DOI: 10.1037/0022-3514.92.6.1087

**Duckworth, Angela Lee, Patrick D. Quinn, and Martin EP Seligman. "Positive predictors of teacher effectiveness." The Journal of Positive Psychology 4, no. 6 (2009): 540-547. https://doi.org/10.1080/17439760903157232**

**Duncan, Greg J. and Dunifon, Rachel (1998). "Soft-Skills and Long-Run Labor Market Success." in Research in Labor Economics. S. W. Polachek, ed. Stamford, CT, JAI Press. 17: 123-149. DOI:10.1108/S0147-9121(2012)0000035036**

**Egan, M., Daly, M., Delaney, L., Boyce, C. J., & Wood, A. M. (2017). Adolescent conscientiousness predicts lower lifetime unemployment. *Journal of Applied Psychology*, *102*(4), 700–709. https://doi.org/10.1037/apl0000167**

**Eskreis-Winkler, L., Duckworth, A., Shulman, E., & Beal, S. (2014). The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00036**

**Forero, C. G., Gallardo-Pujol, D., Maydeu-Olivares, A., & Andrés-Pueyo, A. (2009). A Longitudinal Model for Predicting Performance of Police Officers Using Personality and Behavioral Data. *Criminal Justice and Behavior*, *36*(6), 591–606. https://doi.org/10.1177/0093854809333406**

Galloway, T., Lippman, L., Burke, H., Diener, O., & and Gates, S. (2017). Measuring soft skills and life skills in international youth development programs: A review and inventory of tools. Washington, DC: USAID YouthPower: Implementation.

Gielnik, M., Frese, M., Kahara-Kawuki, A., Katono, I., Kyejjusa, S., Ngoma, M., Munene, J., NamatovuDawa, R., Nansubuga, F., Orobia, L., Oyugi, J., Sejjaaka, S., Sserwanga, A., Walter, T., Bischoff, K., and Dlugosch, T. (2015). Action and Action-Regulation in Entrepreneurship: Evaluating a Student Training for Promoting Entrepreneurship. Academy of Management Learning & Education, 14(1): 69–94. https://doi.org/10.5465/amle.2012.0107

Glewwe, P., Huang, Q., & Park, A. (2017). Cognitive skills, noncognitive skills, and school-to-work transitions in rural China. Journal of Economic Behavior & Organization, 134, 141-164. https://doi.org/10.1016/j.jebo.2016.12.009

**Glewwe, P., Song, Y., & Zou, X. (2022). Labor market outcomes, cognitive skills, and noncognitive skills in rural China. Journal of Economic Behavior & Organization, 193, 294-311. https://doi.org/10.1016/j.jebo.2021.11.021**

**Goffin, R. D., Jang, I., & Skinner, E. (2011). Forced-choice and conventional personality assessment: Each may have unique value in pre-employment testing. *Personality and Individual Differences*, *51*(7), 840–844. https://doi.org/10.1016/j.paid.2011.07.012**

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*, 26–42. https://doi.org/10.1037/1040-3590.4.1.26

**Groves, Melissa Osborne (2005) 'How Important Is Your Personality? Labor Market Returns to Personality for Women in the US and UK', Journal of Economic Psychology 26: 827–841.**

**Hakstian, A. R., Scratchley, L. S., MacLeod, A. A., Tweed, R. G., & Siddarth, S. (1997). Selection of telemarketing employees by standardized assessment procedures. *Psychology &***

***Marketing*, *14*(7), 703–726. https://doi.org/10.1002/(SICI)1520-6793(199710)14:7<703::AID-MAR5>3.0.CO;2-K**

**Hausdorf, P. A., & Risavy, S. D. (2015). Predicting Training and Job Performance for Transit Operators: Predictive Validity of Transit Operator Performance. *International Journal of Selection and Assessment*, *23*(2), 191–195. https://doi.org/10.1111/ijsa.12107**

**Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor economics, 24(3), 411-482. https://doi.org/10.1086/504455**

**Heineck, G. (2011) "Does it pay to be nice? Personality and earnings in the United Kingdom," Industrial and Labor Relations Review, Vol. 64, No. 5, pp. 1020–1038. https://doi.org/10.1177/001979391106400509**

**Hoff, K. A., Einarsdóttir, S., Chu, C., Briley, D. A., & Rounds, J. (2021). Personality Changes Predict Early Career Outcomes: Discovery and Replication in 12-Year Longitudinal Studies. *Psychological Science*, *32*(1), 64–79. https://doi.org/10.1177/0956797620957998**

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. Journal of Applied Psychology, 85, 869–879.doi:10.1037/0021-9010.85.6.869

**Ion, A., Iliescu, D., Ilie, A., & Ispas, D. (2016). The emic–etic approach to personality measurement in personnel selection. *Personality and Individual Differences*, *97*, 55–60. https://doi.org/10.1016/j.paid.2016.02.082**

**Ispas, D., Iliescu, D., Ilie, A., Sulea, C., Askew, K., Rohlfs, J. T., & Whalen, K. (2014). Revisiting the relationship between impression management and job performance. *Journal of Research in Personality*, *51*, 47–53. https://doi.org/10.1016/j.jrp.2014.04.010**

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory-Versions 4a and 54. *Berkeley, CA: Berkeley Institute of Personality and Social Research*. https://doi.org/10.1037/t07550-000

John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of personality: Theory and research, 2nd ed* (pp. 102–138). Guilford Press.

Judge, Timothy A., Christine L. Jackson, John C. Shaw, Brent A. Scott, and Bruce L. Rich. "Self-efficacy and work-related performance: the integral role of individual differences." *Journal of applied psychology* 92, no. 1 (2007): 107. DOI: 10.1037/0021-9010.92.1.107

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. DOI: 10.3386/w20749

Kelcey, B., Shen, Z., & Spybrook, J. (2016). Intraclass correlation coefficients for designing cluster-randomized trials in sub-Saharan Africa education. Evaluation Review, 40(6), 500-525. DOI: 10.1177/0193841X16660246

Laajaj, R., & Macours, K. (2021). Measuring skills in developing countries. Journal of Human resources, 56(4), 1254-1295. DOI: https://doi.org/10.3368/jhr.56.4.1018-9805R1

Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., & Vakis, R. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. Science advances, 5(7), eaaw5226. DOI: 10.1126/sciadv.aaw5226

**Le, K., Donnellan, M. B., & Conger, R. (2014). Personality Development at Work: Workplace Conditions, Personality Changes, and the Corresponsive Principle. *Journal of Personality*, *82*(1), 44–56. https://doi.org/10.1111/jopy.12032**

**Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests: Interpersonal skills and medical school admission. *Medical Education*, *47*(2), 182–189. https://doi.org/10.1111/medu.12089**

Lippman, L.H., Ryberg, R., Carney, R. and Moore, K.A. (2015). Key "Soft Skills" that Foster Youth Workforce Success: Toward a Consensus Across Fields. Washington, DC: USAID, FHI 360, Child Trends. Published through the Workforce Connections project managed by FHI 360 and funded by USAID.

**Luan, Z., Poorthuis, A. M. G., Hutteman, R., Denissen, J. J. A., Asendorpf, J. B., & van Aken, M. A. G. (2019). Unique predictive power of other-rated personality: An 18-year longitudinal study. *Journal of Personality*, *87*(3), 532–545. https://doi.org/10.1111/jopy.12413**

McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. Psychological Review, 123(5), 569–591. https://doi.org/10.1037/rev0000035

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: a clarification of the literature. Journal of Applied Psychology, 86(4), 730. DOI: 10.1037/0021-9010.86.4.730

**McLarnon, M. J. W., Rothstein, M. G., Goffin, R. D., Rieder, M. J., Poole, A., Krajewski, H. T., Powell, D. M., Jelley, R. B., & Mestdagh, T. (2017). How important is personality in the selection of medical school students? *Personality and Individual Differences*, *104*, 442–447. https://doi.org/10.1016/j.paid.2016.09.006**

**Meyers, R., & Houssemand, C. (2010). Socioprofessional and psychological variables that predict job finding. *European Review of Applied Psychology*, *60*(3), 201–219. https://doi.org/10.1016/j.erap.2009.11.004**

Mischel, W. (2014). The Marshmallow Test: Mastering self-control. New York, NY: Little, Brown.

**Murnane, R. J., Willett, J. B., Braatz, M. J., & Duhaldeborde, Y. (2001). Do different dimensions of male high school students' skills predict labor market success a decade later? Evidence from the NLSY. Economics of Education Review, 20(4), 311-320. https://doi.org/10.1016/S0272-7757(00)00056-X**

Oh, I.-S., & Berry, C. (2009). The Five-Factor Model of Personality and Managerial Performance: Validity Gains Through the Use of 360 Degree Performance Ratings. *The Journal of Applied Psychology*, *94*, 1498–1513. https://doi.org/10.1037/a0017221

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: a meta-analysis. Journal of applied psychology, 96(4), 762. DOI: 10.1037/a0021832

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*(4), 995–1027. https://doi.org/10.1111/j.1744-6570.2007.00099.x

**Palczyńska, M., & Świst, K. (2018). Personality, cognitive skills and life outcomes: Evidence from the Polish follow-up study to PIAAC. *Large-Scale Assessments in Education*, *6*(1), 2. https://doi.org/10.1186/s40536-018-0056-z**

**Patterson, F., Lievens, F., Kerrin, M., Munro, N., & Irish, B. (2013). The predictive validity of selection for entry into postgraduate training in general practice: Evidence from three longitudinal studies. *British Journal of General Practice*, *63*(616), e734–e741. https://doi.org/10.3399/bjgp13X674413**

**Pinquart, M., Juang, L. P., & Silbereisen, R. K. (2003). Self-efficacy and successful school-to-work transition: A longitudinal study. *Journal of Vocational Behavior*, *63*(3), 329–346. https://doi.org/10.1016/S0001-8791(02)00031-3**

Puerta, M. L. S., Valerio, A., & Bernal, M. G. (2016). Taking Stock of Programs to Develop Socioemotional Skills. A Systematic Review of Program Evidence. Washington: The World Bank.

**Ramo, L., Saris, W., & Boyatzis, R. (2009). The impact of social and emotional competencies on effectiveness of Spanish executives. *Journal of Management Development*, *28*, 771–793. https://doi.org/10.1108/02621710910987656**

Roberts, Brent W., Nathan R. Kuncel, Rebecca Shiner, Avshalom Caspi, and Lewis R. Goldberg. "The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes." *Perspectives on Psychological science* 2, no. 4 (2007): 313-345. DOI: 10.1111/j.1745-6916.2007.00047.x

**Robertson-Kraft, C., & Duckworth, A. L. (2014). True Grit: Trait-level Perseverance and Passion for Long-term Goals Predicts Effectiveness and Retention among Novice Teachers. *Teachers College Record (1970)*, *116*(3), http://www.tcrecord.org/Content.asp?ContentId=17352.https://doi.org/10.1177/0161468114116003**

**Russell, C. (2001). A longitudinal study of top-level executive performance. *The Journal of Applied Psychology*, *86*, 560–573. https://doi.org/10.1037/0021-9010.86.4.560**

**Saks, A. M., Zikic, J., & Koen, J. (2015). Job search self-efficacy: Reconceptualizing the construct and its measurement. *Journal of Vocational Behavior*, *86*, 104–114. https://doi.org/10.1016/j.jvb.2014.11.007**

Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. Journal of Applied Psychology, 82,30–43. doi:10.1037/0021-9010.82.1.30

**Schoon, I. (2001). Teenage Job Aspirations and Career Attainment in Adulthood: A 17-Year Follow-Up Study of Teenagers Who Aspired to Become Scientists, Health Professionals, or Engineers. *International Journal of Behavioral Development*, *25*. https://doi.org/10.1080/01650250042000186**

**Schwaba, T., Robins, R. W., Grijalva, E., & Bleidorn, W. (2019). Does Openness to Experience matter in love and work? Domain, facet, and developmental evidence from a 24-year longitudinal study. *Journal of Personality*, *87*(5), 1074–1092. https://doi.org/10.1111/jopy.12458**

**Spengler, M., Brunner, M., Damian, R. I., Lüdtke, O., Martin, R., & Roberts, B. W. (2015). Student characteristics and behaviors at age 12 predict occupational success 40 years later over and above childhood IQ and parental socioeconomic status. *Developmental Psychology*, *51*, 1329–1340. https://doi.org/10.1037/dev0000025**

**Spengler, M., Damian, R. I., & Roberts, B. W. (2018). How you behave in school predicts life success above and beyond family background, broad traits, and cognitive ability. *Journal of Personality and Social Psychology*, *114*(4), 620–636. https://doi.org/10.1037/pspp0000185**

Spychala, A., & Sonnentag, S. (2011). The dark and the bright sides of proactive work behaviour and situational constraints: Longitudinal relationships with task conflicts. *European Journal of Work and Organizational Psychology*, *20*(5), 654–680. https://doi.org/10.1080/1359432X.2010.487646

**Stoll, G., Rieger, S., Lüdtke, O., Nagengast, B., Trautwein, U., & Roberts, B. (2017). Vocational Interests Assessed at the End of High School Predict Life Outcomes Assessed 10 Years Later Over and Above IQ and Big Five Personality Traits. *Journal of Personality and Social Psychology*, *113*. https://doi.org/10.1037/pspp0000117**

**Vergauwe, J., Wille, B., Hofmans, J., & De Fruyt, F. (2017). Development of a Five-Factor Model charisma compound and its relations to career outcomes. *Journal of Vocational Behavior*, *99*, 24–39. https://doi.org/10.1016/j.jvb.2016.12.005**

**Viinikainen, Jutta, Katja Kokko, Lea Pulkkinen, and Jaakko Pehkonen (2010) "Personality and labour market income: Evidence from longitudinal data," LABOUR, Vol. 24, No. 2, pp. 201–220. https://doi.org/10.1111/j.1467-9914.2010.00477.x**

**Waldman, D. A., & Korbar, T. (2004). Student Assessment Center Performance in the Prediction of Early Career Success. *Academy of Management Learning & Education*, *3*(2), 151–167. https://doi.org/10.5465/amle.2004.13500529**

**Wanberg, C., Zhang, Z., & DIEHN, E. (2010). Development of the «Getting Ready for your Next Job» inventory for unemployed individuals. *Personnel Psychology*, *63*, 439–478. https://doi.org/10.1111/j.1744-6570.2010.01177.x**

Williams, E. R. (1997). Work personality profile: Validation within the supported employment environment. *Journal of Rehabilitation*, *63*(2), 26–30.