



**Noncognitive Skills in Economics:  
Models, Measurement, and Empirical Evidence**

Hendrik Thiel • Stephan L. Thomsen

**FEMM Working Paper No. 37, November 2009**

***F E M M***

*Faculty of Economics and Management Magdeburg*

**Working Paper Series**

# NONCOGNITIVE SKILLS IN ECONOMICS: MODELS, MEASUREMENT, AND EMPIRICAL EVIDENCE\*

Hendrik Thiel<sup>†</sup>

*NIW, Hannover*

Stephan L. Thomsen<sup>‡</sup>

*Leibniz University Hannover & NIW, Hannover & ZEW, Mannheim*

First version: June 2009

Revised version: November 24, 2011

## Abstract

There is an increasing economic literature considering personality traits as a source of individual differences in labor market productivity and other outcomes. This paper provides an overview on the role of these skills regarding three main aspects: measurement, development over the life course, and outcomes. Based on the relevant literature from different disciplines, the common psychometric measures used to assess personality are discussed and critical assumptions for their application are highlighted. We sketch current research that aims at incorporating personality traits into economic models of decision making. A recently proposed production function of human capital which takes personality into account is reviewed in light of the findings about life cycle dynamics in other disciplines. Based on these foundations, the main results of the empirical literature regarding noncognitive skills are briefly summarized. Moreover, we discuss common econometric pitfalls that evolve in empirical analysis of personality traits and possible solutions.

**Keywords:** noncognitive skills, personality, human capital formation, psychometric measures

**JEL Classification:** I20, I28, J12, J24, J31

---

\*We thank Katja Coneus, Friedhelm Pfeiffer and the participants of the University of Magdeburg Mentoring Seminar 2009, Potsdam, for valuable comments. Financial support from the Stifterverband für die Deutsche Wissenschaft (Claussen-Simon-Stiftung) is gratefully acknowledged. The usual disclaimer applies.

<sup>†</sup>Hendrik Thiel is Research Assistant at the NIW, Hannover. Address: Königstraße 53, D-30175 Hannover, e-mail: thiel@niw.de, phone: +49 511 12331635, fax: +49 511 12331655.

<sup>‡</sup>Stephan L. Thomsen is Professor of Applied Economic Policy at Leibniz University Hannover, Director of the Lower Saxony Institute of Economic Research (NIW), and Research Associate at the Centre for European Economic Research (ZEW) Mannheim. Address: Königstraße 53, D-30175 Hannover, e-mail: thomsen@niw.de, phone: +49 511 12331632, fax: +49 511 12331655.

# 1 Introduction

There is a long literature in economics that investigates the sources and mechanisms of individual differences in labor market productivity. Starting with the seminal work by Becker (1964), numerous approaches modeling the relationship between innate ability, acquired skills, educational investment, and economic outcomes in terms of educational or labor market success have been established.<sup>1</sup> Unfortunately, empirical analysis in this field has been always burdened with a lack of observability of individual differences in these determinants. For several decades, measures of cognitive ability, mostly IQ or achievement tests, have been used for approximation.<sup>2</sup> In psychology, the personality as another source of individual differences in achievements has been a core topic for a long time (see, e.g., Roberts et al., 2007, for an overview). Although these determinants have been implicitly addressed in human capital theory, empirical economists have just started to take these findings into account and to emphasize the crucial role of personality traits in explaining economic outcomes (see, e.g., Heckman and Rubinstein, 2001). For traits related to these outcomes economists use the term *noncognitive skills*. The practical implementation of measurement is attained by means of psychometric constructs.<sup>3</sup> The consideration of trait measures in empirical analysis contributes to a better understanding of the genesis and the evolution of skills other than those indicated by formal education and labor market experience.

However, the objectives in applying these measures in economics and psychology fundamentally differ. Economists are interested in establishing traits as productivity enhancing skills for rather specific settings. Personality psychologists, on the contrary, try to explain an individual's complete spread of behavior and thoughts. Delving into this literature is cumbersome for researchers from the economic discipline. Nonetheless, using the set of psychometric measures in an ill-advised manner due to a lack of knowledge of the broad notions in psychology can lead to very contradictory results.

The overview at hand gives an extensive treatise of central questions and findings regarding definitions, measurement, development, theoretical modeling, outcomes, and problems arising with empirical analysis of personality traits. It complements other surveys on the topic in that it focuses on notational and methodological specifics of the psychological literature and links it to the field of economics. It addresses some critical assumptions necessary to establish existence of a set of noncognitive skills in the sense of the human capital theory. We also address econometric challenges inherent to the analysis of personality traits and provide an overview on the methodological literature that intends to overcome these problems. The survey is completed by a cursory summary of studies

---

<sup>1</sup> Another landmark contribution to the literature on human capital development is due to Ben-Porath (1967). See also Becker and Tomes (1986), Aiyagari et al. (2003), and Coleman and DeLeire (2003).

<sup>2</sup> See, for example, Hause (1972), Leibowitz (1974), Bound et al. (1986), and Blackburn and Neumark (1992). See also Griliches (1977) for an overview.

<sup>3</sup> Psychometrics is the field of psychology that deals with measurement of psychological constructs, including personality traits.

relating to outcomes and formation of personality traits.

The topics are discussed in the following order: we will introduce crucial definitions and elicit how the notion of noncognitive skills is embedded in the psychologic literature in the next section. Afterwards, a selection of psychometric measures for personality traits will be presented and evaluated with respect to their virtues and drawbacks. In addition, we give an introductory overview on validity and reliability measures commonly used in psychometrics, and what should be considered when applying them for construct choice. Section 4 embeds the psychologic and sociologic literature on personality development into a formal framework of human capital formation suggested by Cunha and Heckman (2007). Section 5 outlines general notions and first evidence on how to map personality traits into economic preference parameters. In Section 6 we review a number of studies establishing causal inference for noncognitive skills on several outcomes. Econometric approaches that account for the fact that measures based on test scores only imperfectly represent latent personality traits are introduced in section 7. It intends to provide a valuable means to researchers unfamiliar with the topic. Section 8 concludes and gives an outlook.

## 2 Noncognitive Skills: Some Notational Clarifications

The term *noncognitive skills* originates from the economic literature starting to emerge in course of the work by Heckman and Rubinstein (2001). It comprises the notion of personality traits which are, besides pure intelligence, particularly relevant for several human capital outcomes, such as educational or labor market achievements. These traits constitute, along with other determinants, an individual's personality. In economic terms, personality is a kind of response function to various tasks (see Almlund et al., 2011). There are several approaches in the psychologic literature that target modeling personality in light of environmental entities. A good point of departure is the model suggested by Roberts et al. (2006) and therefore we use it as a reference framework for the remainder of this article. It designates four core factors of personality: personality traits, abilities (cognitive), motives, and narratives. Together with social roles and cultural determinants, these core factors produce the identity (consciously available self-image about the four factors, including self-reports) and reputation (others' perspectives) of an individual. The model also accounts for the possibility of feedback processes, that is, the possibility of environment activating the core factors and vice versa. In its original definition, personality traits are relatively persistent attributes of behaviors, feelings and thoughts, i.e., they are largely non-situational (see Allport, 1937). However, the prevalence of consistency (or at least a certain degree) across situations is not without controversy in the literature. We will elaborate on this in a later section.

Prominent examples for personality traits are self-discipline, self-control, agreeableness, self-

esteem, or conscientiousness, just to mention a few.<sup>4</sup> As the aforementioned model by Roberts et al. (2006) suggests, few issues of personality are devoid of cognition. Sometimes it is even hard to conceptually (not empirically, we will elaborate on this in section 3) distinguish cognitive and personality traits. For instance, emotional intelligence (see Salovey et al., 2004), which describes the processing ability to anticipate the consequences of feelings and the resulting behavior, is a marginal case in terms of cognitive and personality traits.<sup>5</sup> Hence, the denotation *noncognitive* is rather imprecise. Nonetheless, in accordance with most of the economic literature, we use the terms noncognitive skills and personality traits interchangeably in the following.

In order to relate the notions from psychology to economic terms, one can think of cognitive and noncognitive skills as an acquired and inherited stock of human capital. However, to attain definitional plainness we need to resolve the dichotomy of skills and abilities prevailing in traditional human capital literature. Becker (1964) claims a binary stratification where abilities are innate and genetically predetermined, whereas skills are acquired over the life cycle. According to this view, skills and abilities are two distinct determinants of potential outcomes.<sup>6</sup> Contemporary literature, enriched by constructs of personality and intelligence as an interdisciplinary means of measuring human capital, emphasizes and empirically proves that innate abilities provide the initial input in the process of skill formation (see Cunha and Heckman, 2006, Cunha et al., 2010). These findings suggest to waive the distinction between skills and abilities.

Put together, it is obvious that human personality is a highly complex construct that goes beyond the concept of personality traits and requires consideration of multiple factors combined in an interactional pattern. As the remainder of this paper will show, personality and its impact on various outcomes are of particular interest for the field of economics. To make theory and assessment from psychology a powerful toolbox for empirical research in economics, however, one has to presume a sufficient degree of stability and make certain simplifications. Fortunately, an economist's objective is rarely to model and map virtually all facets of personality, but to identify relatively stable and conveniently assessable determinants of the outcomes of interest. This circumstance in association with a general tendency to reconcile different views about cross-situational stability of the personality in the psychological field (Roberts, 2009) will ease some of the discussions rendered subsequently.

---

<sup>4</sup> For example, Allport and Odbert (1936) obtained about 18,000 attributes describing individual differences in the English language.

<sup>5</sup> Borghans et al. (2008a) discuss further examples like cognitive style, typical intellectual engagement, and practical intelligence.

<sup>6</sup> To that effect, Becker (1964) figures out that acquired skills possess higher explanatory power for future earnings than innate abilities do.

### 3 Measuring Noncognitive Skills

There is no uniform assertion about the adequate assessment of personality and the underlying personality models prevalent in psychology. Hence a brief overview on the relevant psychologic literature is a sensible first step. The crucial issue in terms of postulating persistent skills for economic analysis is to ensure the stability of personality traits across situations. If consistency across situations is only fragile or at worst non-existent, and situational and contextual determinants drive observed measures instead, it would be meaningless to evaluate its relevance and to construe it as a persistent set of skills which constitutes human capital.

#### 3.1 Personality and Situations

The existence of persistent traits has been subject to vigorous discussion in the literature in the last decades. The common reading of the influential work by Mischel (1968) is that all patterns of behavior, feelings, and thoughts are manifestations of specific situations, not of personality traits. Mischel and his proponents constitute this as a misinterpretation in the aftermath. Mischel (1973) has endeavored to incorporate situation into whatever drives stable characteristics (e.g. personality traits), dubbing it the *if-then* signature of personality. This signature characterizes an individual by stable patterns of variability across situations. Orom and Cervone (2009) therefore claim that Mischel's initial point was that cross-situational consistency in personality assessment is low when limiting to global, nomothetic trait constructs.<sup>7</sup> Things become more clear-cut when dropping this rigid assumption and allowing for other factors to affect measured personality. The ensuing discourse in the literature led to alternative notions of personality.

The *social-cognitive approach* established and advocated by Mischel (1973) and Bandura (1986) provides such an alternative structure of personality. It mainly focuses on explaining the cognitive processing underlying thoughts and behaviors. Accordingly, people differ in terms of cognitive abilities relevant for the implementation of certain behaviors. The awareness of these abilities in conjunction with expectations about self-efficacy, goals, and valuation standards constitute the personality. All four subsystems of personality are interactional in nature and therefore not separately assessable.<sup>8</sup> Strictly speaking, the evaluator has to account for the situation as perceived by the observed individual and analyze consistent patterns in this situational context. One of Bandura's contributions is the concept of *reciprocal determinism*. It essentially states that there is no actual source of behavior as asserted by trait theorist or behaviorists. Instead there is a triangular feedback system composed of personal characteristics, behavior, and environmental factors. An interior processing approach underlying this system is the *cognitive-affective processing system* by Mischel and Shoda (1995). It

---

<sup>7</sup> This does not imply that cross-situational consistency is non-existent.

<sup>8</sup> A social-cognitive theorist would not assign a certain score to one of these systems.

interrelates the abovementioned subsystems (abilities, expectations, goals, and valuation standards) by means of cognition and affects. Individual differences therefore arise from differences in activation levels of cognitions and affects. The accessibility of activation levels differs over various situations.

A contrary view to the whole situation debate is held by the proponents of the *global dispositional approach*. It is exemplified best by concepts like the Five Factor Model of Goldberg (1971) and related.<sup>9</sup> Proponents of the Five Factor Model constantly highlight the consistency of personality across time and situation.

The most widely held approach in contemporary personality psychology is to combine the assumption of a certain stability in traits with social-cognitive units, such as goals expectations, and assign them to different levels of analysis.<sup>10</sup> The Roberts model for personality that we use as a baseline accounts for these units and their interaction with environmental factors.<sup>11</sup> Moreover, it maintains the notion of personality traits irrespective of the underlying cognitive processing function and refers to other units of analysis in the system of personality when modeling contextual patterns.

All observed (or otherwise assessed) measures of latent personality traits can also be manifestations of the other units of analysis addressed in the Roberts model. For instance, fulfilling a certain social role at the time the assessment takes place is a contextualizing variable one has to control for (see Wood, 2007).<sup>12</sup> This proceeding gives rise to a certain stability of personality traits across situations and therefore paves the way for application of the kind of personality tests empirical economists are most interested in.

### 3.2 Types of Assessment-Tools

How do trait theorists or social-cognitive theorists assess the entities in their respective models and what are the pros and cons of the respective methods with regard to different assessment situations? There are three main dimensions an evaluator has to decide on: (1) the type of assessment, (2) the person to be assessed, and (3) the dimension.

(1) Proponents of the social-cognitive approach usually rely on qualitative assessment methods conducted by experts who passively observe or interview. This method involves variations of situational stimuli and substitutions of the assessed person until systematic evidence for the underlying processing is revealed. For applications within the scope of observational (and even experimental) data prevailing in empirical economics this type of assessment is rather cumbersome and costly. For large scale

<sup>9</sup> The most widely applied version is the Big Five inventory of Costa and McCrae (2008).

<sup>10</sup> Even very early definitions of personality traits implicitly account for situational variance in behavior (Allport, 1961, p. 347)

<sup>11</sup> Roberts (2009, p. 138) vividly summarizes this reconciling approach in the following manner: “The trait psychologists can continue to focus on factor structure and test retest stability. The social cognitive psychologists can study goals, motives, beliefs, and affect - things that putatively change.”

<sup>12</sup> As section 4 will briefly address, permanently fulfilling certain social roles does not solely affect measures of personality, but induces changes of states of personality traits as well.

investigations quantitative assessment methods are undeniably more appealing and, therefore, in our focus henceforth. In general, their aim is to provide scores for respective dimensions of the construct. These scores are directly used for analyses or employed to derive an underlying latent construct.

(2) Moreover, the evaluator has to choose between self-reports and observer reports. Self-reported measures are convenient due to their simple implementation but implicitly assume that respondents are capable to consciously perceive their personality or at least their actions manifesting it. This prerequisite does not generally apply. For instance, infants and children are not capable of doing so and, thus, are usually assessed by observers from the social environment (parents or teachers). Distortions of self-reports or observer ratings can also be more generic. For traits related to typical social settings, like meeting a stranger or having a discussion, observer ratings tend to predict behavior better than self-reports since the potential for disorder in self-perception is high. For instance, what the narrator of a joke believes to be funny is not perceived by others in the same manner. Vice versa, self-reported personality ratings are more strongly related to assessments of emotional issues driven by interior processes and less shared with others. An illustrative example for that is a person annoyed by depressions who would usually try to conceal his or her problems from others. On the other hand, they lack the virtue of easy implementation of test scores administered as a questionnaire. The choice of the person to be assessed therefore strongly depends on the trait of interest.

(3) The remaining question is to which extent a measure captures the personality. Besides various low-dimensional scores for assessing the magnitude of specific traits, there is a large number of taxonomies mapping human personality as a whole. Proponents of these personality inventories or high-order constructs advocate the global dispositional approach we discussed above and therefore construe these models as comprehensive personality inventories.

**Higher-Order Constructs:** Most mappings of personality impute a hierarchical structure which is based on common (exploratory) factor analytic approaches exploiting lexical-linguistic information. These inventories build on measurement systems comprising a multitude of respective questions (items). They follow the notion of the setups usually applying to the segmentation of general IQ.<sup>13</sup> In case of personality the level of abstraction is lower. Despite early efforts to identify a general factor for personality (see Webb, 1915), personality inventories from the prosperity period of the global dispositional approach usually assume at least three major factors. Table 1 provides an overview on the commonly used concepts in the literature.

< Include Table 1 about here >

---

<sup>13</sup> A version of Cattell (1971) includes fluid intelligence, that is, the ability to solve novel problems, and crystallized intelligence, comprising knowledge and developed skills.

A widely accepted taxonomy is the Five Factor Model established by Goldberg (1971) and the related Big Five by Costa and McCrae (1992). The identification of five high-order factors is not uncontested in the literature. Some factor analytic results suggest a lower number of factors, whereas others claim a higher number. Eysenck (1991), for example, provides a model with just three factors; Digman (1997) curtails the Big Five distinction to only two higher-order factors.<sup>14</sup> In contrast to that, Hough (1992) proposes a more stratified version of the Big Five taxonomy, the so-called Big Nine. Due to their factor analytic genesis virtually all the aforementioned concepts lack a theoretical foundation and, therefore, are largely inconsistent with the type of personality models discussed above. Only for exceptional cases neurological support for the constructs is available (see, e.g., Canli, 2006, pertaining to the Big Five).

As a consequence, low predictive power of a high-order factor does not necessarily imply that all of the lower-order factors in Table 1 exert no influence on an outcome of interest. Using lower-order constructs or even uni-dimensional factors often entails a gain in explanatory power, but at the potential cost of not covering all relevant personality facets.

**Lower-Order Constructs:** There is also a large number of lower-order constructs. Prominent examples in the context of educational outcomes are self-control (see Wolfe and Johnson, 1995) and the related self-discipline (see Duckworth and Seligman, 2005). The Brief Self-Control Scale (Tangney et al., 2004) is a commonly used means of assessing self-control. It includes 13 items which sum up to a score increasing with self-control. The Internal-External Locus of Control by Rotter (1966) is often perceived as a related measure, but merely assesses an individual’s attitude on how self-directed (internal) or how coincidental attainments in her or his life are. The original Locus of Control (Rotter, 1966) comprises 60 items. Usually, longitudinal datasets apply abbreviated versions.<sup>15</sup> A similar scale for Locus of Control is the Internal Control Index (Duttweiler, 1984), a 28-item scale that scores in the internal direction. Self-esteem provides a further important determinant of educational and labor market outcomes (see Heckman et al., 2006). It is often quantified by the Rosenberg Self-Esteem Scale (Rosenberg, 1965), a 10-item scale. For the assessment of children’s personality the use of observation from third persons prevail. A corresponding scale based on observational report is the Self-Control Rating Scale (Kendall and Wilcox, 1979), a 33-item scale indicating the ability of inhibiting impulsiveness.

This leads to the related field of temperamental studies prevailing in developmental psychology.<sup>16</sup> Constructs to assess temperament rather refer to behavioral tendencies than pure behavioral

<sup>14</sup> The factors are not presented in Table 1 since they are simply denoted *metatraits* without further specification.

<sup>15</sup> The German Socio Economic Panel (SOEP), for instance, comprises a 10 item version, whereas the National Longitudinal Survey of Youth uses a 23 item version.

<sup>16</sup> Developmental psychology deals with all kind of psychological changes over the life course, not only personality. The major focus, however, is on infancy and childhood.

acts. An influential model has been suggested by Thomas et al. (1968). It stratifies temperament to nine categories each grouped into three types of intensity. There are further established concepts of temperament, e.g., Buss and Plomin (1975) and Rothbart (1981), but even more recent literature is still involved in this topic (see, e.g., Rothbart and Bates, 2006).<sup>17</sup> Meanwhile, some interrelations between concepts of personality psychology and developmental psychology have been established. For instance, Caspi (2000) reveals links between the extent of temperamental facets at age 3 and personality at adulthood. Temperament at infancy and early childhood designates later personality but is remittently affecting behavior as the individual matures. According to Thomas and Chess (1977), purely temperamental expressions at later age are only likely in case of being faced with a new environmental setting. However, the inferences from studies linking temperament and personality are far from being conclusive (see Rothbart et al., 2000, Shiner and Caspi, 2003, Caspi et al., 2005, for a review of the literature).

### 3.3 Reliability

Reliability refers to the consistency of answers to a psychometric task over time or across observations.<sup>18</sup> It originates from methods of classical test theory, one of the very first fields analyzing issues of measurement error. What should be considered when picking tests in order to ensure a high degree of reliability? The consistency of a task to measure a trait is mainly imperiled if other units of analysis in the (Roberts) model are captured by it. Separately assessing these units is difficult due to the fact that they are not isolated but instead influence each other simultaneously. For instance, when measuring a certain personality trait by means of a questionnaire, it is important not to prompt the respondent to project his thoughts into a particular situation in order to reply to an item. In this case the score can be a manifestation of the trait of interest, but also of motivation, past experiences, or narratives and abilities helpful for comprehension of the task.

Though proponents of the global dispositional approach claim that most of the variables discussed in the Roberts model can be mapped into the dimensions of their personality inventories (see Costa and McCrae, 1992, for empirical evidence), this result is dissatisfying when it comes to identification of persistent traits and subsequent anchoring to economic outcomes. Methods of exploratory

<sup>17</sup> See Goldsmith et al. (1987) for an overview on temperamental measures.

<sup>18</sup> Reliability could be most directly checked by means of test-retest correlations over time. Generally, each test item  $i$  can be expressed as

$$x_i = \alpha_i \tau_i + \varepsilon_i,$$

where  $x_i$  is the attained score,  $\tau_i$  is the true score with  $\alpha_i$  as the corresponding scaling parameter, and  $\varepsilon_i$  is an error term. Since test-retest settings are rarely at hand, other coefficients prevail. A standard measure to quantify reliability across several items is Cronbach's alpha (see Cronbach, 1951) which can be determined as follows:

$$\rho_\alpha = \left( \frac{l}{l-1} \right) \left( 1 - \frac{\sum_{i=1}^l \text{Var}(x_i)}{\text{Var}(\sum_{i=1}^l x_i)} \right),$$

where  $l$  is the number of items used to measure the true score. It relates item variance to the variance of the total score and therefore increases with rising inner consistency of the construct.

factor analysis are a widely used tool for development and construction of global personality mappings like the Big Five.<sup>19</sup> In case of low-order constructs or uni-dimensional factors, exploratory factor analysis is primarily used for verification of the assumed structure. In either instance, neglecting the influences of accompanying determinants can be harmful for the resulting measures of traits.<sup>20</sup> By construction, exploratory factor analysis cannot disentangle the effects of immediate common factors and indirect pathways. Therefore, the identification problem inhering a lack of contextualization frequently causes some variation to be attributed to measurement error or spurious pattern of the trait under study. The former may occur if the item formulation unsystematically induces the measured scale to include framing effects due to motivation or social roles. The latter is likely to result from more systematic distortions.

In order to elude these drawbacks it is necessary to contextualize the measurement, that is to control for situational determinants that potentially affect expression of abilities, motivation and the like. When using questionnaires as an assessment tool, the evaluator should avoid to mentally force the respondent into specific situations to answer an item. Intuitively, low-dimensional or uni-dimensional constructs are less susceptible to these phenomena since they usually rely on a higher number of items to examine a certain trait and are easier to validate by means of other constructs or outcomes (see next section).

Contextualization comprises all interactions between entities of the personality that may distort identification of traits. If this bias, however, is solely conscious, the researcher has to deal with faking.<sup>21</sup> The potential for faking is higher for measures of personality traits than for cognitive abilities. The background of the assessment can urge the respondent to understate and/or overstate. As an example consider a test administered for making a hiring decision. The faking behavior in tests is also a projection of other personality traits or cognitive capabilities. Borghans et al. (2008b) provide evidence for an interrelation between personality and incentive responsiveness. Fortunately, Morgeson et al. (2007) conclude that correcting for intentional faking does not improve the validity of measures.<sup>22</sup>

### 3.4 Validity

After a construct has been developed by means of data reduction (exploratory factor analysis) or detailed theoretical knowledge, validity is concerned with whether a scale measures what it is supposed to measure. It should be tested when developing a scale, but should also be considered whenever a construct is otherwise applied. In the psychometric literature, three types of validity are distinguished (see, e.g., Cervone et al., 2005).

---

<sup>19</sup> A comprehensive introduction into the methods of exploratory factor analysis is provided by Mulaik (2010).

<sup>20</sup> A vivid impression of construct development is given in Tangney et al. (2004).

<sup>21</sup> If one assumes intentional faking to be part of the influences accounted for by contextualization.

<sup>22</sup> The next section discusses drawbacks of psychometric validity measures.

**Content Validity:** Content Validity is a qualitative type of validity and requires sound theoretical foundation in order to evaluate whether the whole theoretical domain is captured by the data. For instance, if a construct justified by theory comprises three different dimensions, that is, three latent factors, one needs measures for all of them. Otherwise, content validity is questionable. A lack of theoretical consensus is the major weakness of this kind of validity.

**Criterion Validity:** To test for criterion validity one needs a variable that constitutes a standard measure to which to compare the used measures. It can be a concurrent measure from the same measurement system or a predictive measure provided by a future outcome. The magnitude is usually represented by means of correlations between measurement and criterion variables. It can be shown (see Bollen, 1989, for a detailed discussion) that the magnitude is largely sensitive to unsystematic error variance in both, measurement and criterion variable, and depends on the choice of criteria.<sup>23</sup> Moreover, validity measures based on simple correlation are not necessarily capturing a causal relationship.

**Construct Validity:** For many constructs in psychometrics it is difficult to find measures that establish criterion validity. Instead one has to rely on construct validity. It assesses to what extent a construct relates to other constructs in a fashion that is in line with underlying theory. The resulting coefficient is again a correlation. By arguments similar to those invoked for criterion validity, other driving forces than the quality of the proxy, like factor correlation and reliability of the measure, can contaminate the validity coefficient.<sup>24</sup> Moreover, the choice of comparison constructs is arbitrary. A more systematic approach of establishing construct validity is the multitrait-multimethod design

<sup>23</sup> Consider both variables in an additive separable factor representation

$$\begin{aligned}x &= \lambda_1\theta_1 + \varepsilon_1 \\C &= \lambda_2\theta_1 + \varepsilon_2,\end{aligned}$$

where  $x$  is the applied measure,  $C$  is the criterion measure,  $\theta$  is the latent factor constituting both with respective factor loadings  $\lambda_1$  and  $\lambda_2$  (i.e. the scale), and  $\varepsilon$  is the measurement error. The correlation between  $x$  and  $C$  (which is the validity coefficient by Lord and Novick, 1968) is

$$\rho_{x,C} = \frac{\lambda_1\lambda_2\phi_{11}}{\sqrt{\text{Var}(x)\text{Var}(C)}}.$$

Even if all measures are standardized (usually this assumption extends to the latent factors, which makes  $\phi_{11}$  a correlation matrix) and the denominator therefore vanishes, the coefficient depends on more than the quality of  $x$  as a proxy for  $\theta_1$  (quantified by  $\lambda_1$  and  $\varepsilon_1$ ).

<sup>24</sup> A formalization is less straight forward than in the previous case but can be sketched as follows. Consider two measures  $x_1$  and  $x_2$  for two latent traits  $\theta_1$  and  $\theta_2$  with different loadings  $\lambda_{11}$  and  $\lambda_{22}$ .

$$\begin{aligned}x_1 &= \lambda_{11}\theta_1 + \varepsilon_1 \\x_2 &= \lambda_{22}\theta_2 + \varepsilon_2.\end{aligned}$$

It can be shown that the construct validity depends on more than the relation of the latent factors:

$$\rho_{x_1x_2} = \sqrt{\rho_{x_1x_1}\rho_{x_2x_2}}\rho_{\theta_1\theta_2},$$

where  $\rho_{x_1x_1}$  represents reliability.

suggested by Campbell and Fiske (1959). It requires that two or more traits are measured by two or more constructs (i.e. methods). If the correlations for the same trait across measures are significantly large, there is evidence for convergent validity. Discriminant validity arises if convergent validity is higher than the correlation between measures which neither share trait nor method and higher than the correlation between different traits measured with the same method. Again, the magnitude of convergent validity can be sensitive for other reasons than closeness of the measure, like latent factor correlation and reliability.

As the foregoing discussion suggests, there is a twofold circularity to solve in order to obtain reasonable validity measures. The first is circularity in a statistical sense, that is, a simultaneous causality between measures and the latent traits. The second is a circularity in justification of genuine measures and resultant measures of validation. This is what Almlund et al. (2011) denote an intrinsic identification problem rather than a parameter identification problem. Loosely speaking, one should always be aware of the “chicken and egg problem” of choosing a construct and validating it by means of constructs that were established in the same manner. In order to resolve the former problem and to ensure causality one has to rely on structural equation approaches.

To deal with the latter, at least one *dedicated* measurement equation per trait is required (following the term by Carneiro et al., 2003), that is, a measure that exclusively depends on a particular trait. At the same time the use of dedicated measures warrants parameter identification. Consider a psychometric task. Even if one controls for situational determinants, identification is restricted to tuples of traits without dedicated measures. In case of low-order constructs and respective real world outcomes this reasoning is easier to achieve. For more general dimensions it requires a more profound justification in choosing measurement and validation constructs.

## 4 Determinants and Dynamics of Personality Traits

Arguably, only those components that are sufficiently stable across situations, that is personality traits and cognitive abilities, can be construed as skills in the sense of the human capital literature.

With this subtle notion about the complexity of personality at hand, the remaining questions with particular relevance for policy recommendation are (1) what determines the formation of the personality traits and (2) to what extent are they influenced by the environment. In the following we will present a theoretical approach known as the Technology of Skill Formation (see Cunha and Heckman, 2007) along with a brief overview on the underlying empirical literature.

Similar to the Roberts model of thoughts and behavior in an environmental context, the interactional pattern between personality traits and IQ has to be considered for the formation process.<sup>25</sup>

---

<sup>25</sup> As mentioned in the previous section, cognitive capabilities can have an impact on faking behavior in responding to a personality test. Vice versa, IQ tests never exactly measure pure cognitive intelligence. The results also can

Cunha et al. (2006) refer to a range of intervention studies which capture different periods of childhood and adolescence. The respective results are summarized in Table 2. Most of the data sets used in the empirical studies cover childhood and adolescence retrospectively and only provide measures of cognitive abilities and scholastic achievement. Fortunately, there is a strong consensus in the literature that IQ largely stabilizes before schooling age. If scholastic achievements are an outcome of intelligence and some other abilities and a certain treatment results in a permanent shift in achievements but not in IQ, then there are other, presumably noncognitive, skills that are affected by the treatment.

< Include Table 2 about here >

Although the evaluation of interventions providing such kind of treatment does only provide implicit evidence for the formation of noncognitive skills, Cunha et al. (2006) reveal a clear formation pattern involving two important features: *self-productivity* and *dynamic complementarity*. Self-productivity postulates that skills acquired at one stage enhance the formation of skills at later stages. Dynamic complementarity captures that a higher level of skills at an earlier stage enhances the productivity of investments in the ensuing stages and that early investments should be followed by later ones. As a consequence, the early childhood constitutes a bottleneck period for investments in the formation process.

Evidence for this pattern is provided by research from various disciplines. In neurobiology the existence of such critical periods is attributed to a superior susceptibility of neural circuits and brain architecture in early lifetime (see Knudsen, 2004, Knudsen et al., 2006). Studies from clinical psychology draw the same inference. For instance O'Connor et al. (2000) assess cognitive abilities among a group of Romanian orphans who were adopted into UK families between 1990 and 1992 and compare them at ages four and six to adopted children from within the UK. As opposed to the Romanian orphans, the UK orphans were all placed into their new families before the age of six months. Their findings suggest that early deprived children never catch up. In case of personality traits the time period of malleability is longer. Intervention studies aiming at children in school age usually report gains in behavioral measures. As the findings of Table 2 illustrate, even interventions at primary school age boost scholastic performance in a lasting manner without permanently raising IQ. By the arguments above, these findings provide implicit evidence on the susceptibility of personality beyond early childhood. This is in line with the literature in pediatric psychiatry (see, e.g., Dahl, 2004) highlighting the role of the prefrontal cortex in governing emotion and self-regulation and its malleability up into the early twenties of life. There is evidence for an even more extensive period of plasticity. For instance, Roberts and Delvecchio (2000) in a meta-analysis show that the rank-order of the Big Five factors stabilizes beyond adolescence, but there are still moderate changes until age 50.

---

reflect motivational and thus aspects of personality traits.

Roberts et al. (2006) show the highest mean-level change to be concentrated on young adulthood. The authors suggest that these changes are induced by persistent shifts in social roles and role expectations common to most individuals. Given the discussion on the accuracy of the Big Five to measure pure traits in the previous section, these findings seem reasonable. A social role is a situational factor that determines measured traits. As long as there are changes in social roles over the life course, it is tempting to interpret them as instability in actual traits.<sup>26</sup>

The assumption of complementarity across stages is in line with the following empirical picture. Table 2 shows that early interventions which involve a long-term follow-up are most successful. However, most of the gains fade out if no further efforts are made. Vice versa, sole remediation attempts in adolescence exhibit only weak effects.<sup>27</sup> However, the efficiency of interventions in adolescence is definitely lower compared to early intervention programs. As established by Cunha and Heckman (2007), a CES production function provides enough flexibility to account for complementarity and self-productivity of investments. It allows for different elasticities of substitution between inputs at different stages and for different skills. This yields the following functional form for successive periods  $t \in \{1, \dots, T\}$ :

$$\theta_{t+1}^j = \left[ \gamma_{1,t}^j (I_t)^{\rho_t^j} + \gamma_{2,t}^j (\theta_t^C)^{\rho_t^j} + (1 - \gamma_{1,t}^j - \gamma_{2,t}^j) (\theta_t^N)^{\rho_t^j} \right]^{\frac{1}{\rho_t^j}}, \quad (1)$$

where  $\theta$  with  $j \in \{C, N\}$  denotes the latent cognitive and noncognitive skills. Moreover,  $\rho_t^j$ ,  $\gamma_{1,t}^j$  and  $\gamma_{2,t}^j$  are the respective complementarity and multiplier parameters. The notation in equation (1) therefore accounts for cross-productivity between cognitive and noncognitive skills. The functional specification also allows for the explicit incorporation of additional determinants like parental characteristics, prenatal environment, or children's health capabilities (see, e.g., Coneus and Pfeiffer, 2007; Cunha and Heckman, 2009). Gene endowment could be regarded as the initial input into the skill formation process and not as an additive component. Even before birth, crucial modules for future skill formation are established by environmental influences (see Shonkoff and Phillips, 2000). These environmental and genetic components are not simply additive. A large literature from behavioral genetics deals with this question. For instance, Fraga et al. (2005) reveal that monozygotic twins exerted to different stimuli throughout early childhood can exhibit significantly different gene expressions due to differences in DNA methylation. This is in line with twin and adoption studies from social science. For example Turkheimer et al. (2003) show that a simple additive model structure is inappropriate

<sup>26</sup> According to Almlund et al. (2011) there can be a kind of feedback between traits and situation since many situations are a consequence of trait endowment earlier in life.

<sup>27</sup> There are a number of studies evaluating adolescent mentoring programs, like the Big Brothers/Big Sisters (BB/BS) and the Philadelphia Futures Sponsor-A-Scholar (SAS) program. The BB/BS assigns educated volunteers to youths from single parent households for the purpose of providing surrogate parenthood or at least an adult friend. Grossman and Tierney (1998) stress that meeting with mentors decreases the probability of initial drug and alcohol abuse, exertion of violence, and absence from school. Moreover, the participants had higher grade points and felt more competent in their school activities. SAS targets at public high school students and supports them in making it to college by academic and financial support. Johnson (1996) reveals a significant increase in grade point average and college attendance.

to capture the complexity of the IQ generating process and that there are substantial interactions of genes and environment. The fact that most empirical results from adoption studies promote genetical factors as the main driving force of skill formation is due to the low share of low income families from adverse environments in these samples. Personality and behavioral patterns also have a genetic and an environmental component (see Bouchard and Loehlin, 2001) and the same pattern applies. For instance, Caspi et al. (2002) reveal this relationship for psycho-pathologic phenomena like antisocial behavior.<sup>28</sup>

When adulthood is attained in period  $T + 1$ , the disposable stock of human capital can be regarded as the outcome of the acquired cognitive and noncognitive skills developed up to  $T$  in a specification as in equation (1). Cunha and Heckman (2006) present estimates of the parameters of equation (1) and directly quantify the degrees of self-productivity and complementarity. The data they use comprise measures of cognitive ability, temperament, motor and social development, behavioral problems, and self-confidence of the children and of their home environment. The results yield strong evidence for self-productivity within the production of the respective skill types.<sup>29</sup> The cross-effects are weaker. Complementarity is evident for both, cognitive and noncognitive stocks, but somewhat higher in case of the former. The average parameter estimate is slightly below zero which indicates that the production technology could be approximated well by a Cobb-Douglas function. Slightly altered estimation strategies yielding similar results are provided by Cunha and Heckman (2008) and Cunha et al. (2010).

The estimation approaches used to quantify the parameter values in equation (1) yield factor loadings that represent the roles played by different environmental resources in the skill formation process. According to these results, indicators that relate to cultural and educational involvement, like having special lessons or going to the theater, are of particular importance. Family income however is less important. As Currie (2009) suggests, parents obtaining higher labor market returns may invest less time in children and are only partially able to compensate this neglect by provision of substituting goods. The properties of the skill formation discussed above, suggest that schooling, in particular post-primary schooling, is a minor determinant compared to investments outside school. The major foundation is already set in preschool age. Additional data constraints even bolster these effects. As discussed by Todd and Wolpin (2007) in context of education production functions it is generally difficult to find data that combine rich information on schooling and home resources. Moreover, there is always less variation in more aggregated indicators for schooling resources. This could lead to additional attenuation of the estimated effects.

<sup>28</sup> A further discussion including additional empirical evidence is given in Heckman (2008) and Cunha and Heckman (2009).

<sup>29</sup> The identification strategy is in spirit of the factor structure models discussed in section 7. It allows for endogenous choice variables and measurement error in indicators.

## 5 Personality Traits and Economic Preference Parameters

It is clearly intuitive to assume a relationship between the expressions of cognitive and noncognitive skills and economic preference parameters. For instance, the patience of an individual is arguably related to his or her time preference. As Borghans et al. (2008a) summarize, from an economic point of view it is meaningful to relate personality concepts to common parameters like time-preference, risk-preference, and leisure-preference, but also to the more recently studied concepts of social preferences like altruism and reciprocity (Fehr and Gächter, 2000). Relating traits and preference parameters in a causal way requires a notion of the underlying theory. Due to the complexity of human thoughts and behavior, such a theory is difficult to establish. Even without consideration of traits, finding a functional form for a utility function that accounts for all facets of social preferences observed in the lab is tedious (see Fehr and Schmidt, 2006, for a discussion). Following Almlund et al. (2011) and their various model suggestions, from an economic point of view personality traits can be construed as preferences as well as constraints. Formally, a utility maximizing agent under uncertainty could be characterized by the following implicit representation:

$$E[U(x, P_{\theta,e}, e|\psi_{\theta})|\mathcal{I}_{\theta}] \text{ s.t. } I + r'P_{\theta,e} = x'w \text{ and } \sum e = \bar{e} \quad (2)$$

All variables with  $\theta$  as a subscript constitute a possible pathway of influence of traits on the economic representation of an agent's response function.  $E[U(\cdot)|\mathcal{I}_{\theta}]$  is the expected utility for the arguments  $x$ ,  $P(\cdot)$ , and  $e$  given the information set  $\mathcal{I}$  which in turn depends on traits. All arguments are vectors:  $x$  is a vector of consumption goods and  $e$  is the vector of effort devoted to all possible tasks and the sum of its elements cannot exceed  $\bar{e}$ .<sup>30</sup> Since effort can cause a kind of "good feeling" after endeavor, it also enters the utility function directly. In addition, it is a complement for the vector of available traits  $\theta$  in the vector function for productivity  $P(\theta, e)$ , which maps  $\theta$  and  $e$  into outcomes for all possible tasks.  $P(\cdot)$  is the intangible pathway of productivity into utility, whereas the tangible one is through consumption goods. The goods with price vector  $w$  are funded by income not depending on productivity for tasks  $I$  and the income from performing tasks for task-specific rewards  $r$ . Traits can also be a constraint in another sense than in equation (2). Dohmen et al. (2007) discuss the potential for confounding due to observational equivalence of differences in actual preferences and differences in capabilities required to perform the task that is used to measure the preferences. In terms of equation (2) this means it is difficult to disentangle  $\psi$  and  $\mathcal{I}$ . As an example, consider the degree of numeracy that affects the comprehension of an investment decision used to assess time preference.

Borghans et al. (2008b) examine potential links between noncognitive traits and responsiveness for incentives in answering cognitive tests using primary data for a sample of Dutch students.

<sup>30</sup> Think of effort as a representation of the situational parameters discussed in the psychologic literature above.

The responsiveness is captured by common economic preference parameters. They find a negative correlation between the Internal Locus of Control and the personal discount-rate and similarly a correlation between emotional stability and risk-preference. Both appear intuitively plausible. Dohmen et al. (2008) use data from the German Socioeconomic Panel (SOEP) to reveal possible relationships between Big Five personality traits, measures of reciprocity, and trust. All Big Five factors exert significant positive influence on positive reciprocity, especially conscientiousness and agreeableness. Moreover, neuroticism promotes trust and negative reciprocity.

Given the complexity described above, studies which rely on correlations between skills and economic preference parameters provide only vague and sometimes inconclusive evidence. Generally, questionnaire assessments of preference parameters are likely to suffer from a number of potential problems. The observed preferences are either simply stated, i.e., on hypothetical items, or if revealed, only within a non-market setting. Yet, it is ambiguous whether preferences for artificial and real market settings are identical (see Kirby, 1997, and Madden et al., 2003, for two opposing views). If an experimental assessment embodies real rewards, choosing the respective payoffs binds the participant to maintain his or her choice. In a real life setting, however, the individual also has to withstand other opportunities, and there may be a higher degree of uncertainty for future payoffs. It proves difficult to partial out time preference from risk-aversion (see Borghans et al., 2008a, and the literature they refer to). Moreover, measures of time preference may be subject to framing effects. Non-linearities with respect to the payoffs are also likely and limit the external validity of experimental findings. There are numerous other inconsistencies which indicate the premature status of this research field.<sup>31</sup>

## 6 Direct and Indirect Outcomes of Personality Traits

Until recently, noncognitive skills have not played an important role in explaining labor market outcomes. Bowles et al. (2001) argue that early explanations of wage differentials like disequilibrium rents (Schumpeter, 1934) and incentive effects (Coase, 1937) can be explained in light of personality traits. In the sense of these models, the respective traits are construed as not being productivity enhancing. They stress that it is important to distinguish different scopes of the labor market. Two illustrative examples demonstrate this: in a working environment where monitoring is difficult, behavioral traits like truth telling may be higher rewarded than in other cases. Considering a low-skill labor market, docility, dependability, and persistence may be highly rewarded, whereas self-direction may generate higher earnings for someone who is a white collar worker. Besides different rewards in different occupation segments of the labor market, people also opt in these occupations owing to personality (see Antecol and Cobb-Clark, 2010). Heinicke and Thomsen (2011) show that returns to noncognitive skills

<sup>31</sup> For further discussion see Almlund et al. (2011).

within occupational groups provide a mixed signal due to group-specific returns and self-selection.

It is difficult to determine if certain traits increase wages by affecting occupational choice, productivity, or if market mechanisms additionally induce wage premiums for certain traits. On a more general level Borghans et al. (2008c) show that supply and demand for workers more or less endowed with directness relative to caring create a wage premium for directness. Another explanation is that the society solidifies certain expectations about appropriate traits and behavior, and rewards or punishes individuals who deviate from them in either direction. This interpretation is fostered by the results of Mueller and Plug (2006) for the gender wage gap in the US. They show that particularly men obtain a wage penalty for Big Five agreeableness, a trait stronger associated with women.

Our aim in what follows is to give a very short review of empirical studies dealing with predictive power of noncognitive skills. Tables 3 and 4 provide characteristics of a selection of studies but are far from comprehensive. Borghans et al. (2008a) and Almlund et al. (2011) give a more widespread overview of empirical evidence, including the literature from other disciplines.

< Include Tables 3 and 4 about here >

Irrespective of how the traits are valued in the market, noncognitive skills explain differences in the earnings structure well. Heckman et al. (2006) provide empirical evidence on the effects of noncognitive skills constituted by self-control and self-esteem on log hourly wages. Especially for the lower deciles of the distribution of latent skills a strong influence is revealed. Flossmann et al. (2007) reproduce these results for German data.

< Include Figure 1 about here >

Figure 1 compares the net effects of an increase in noncognitive abilities on log wages obtained in the two studies. Particularly for the upper and lower deciles of the distribution the marginal effect of an increase in noncognitive skills is higher. Both results provide an important indication on how personality traits affect earnings.

Noncognitive (and cognitive) abilities do not solely affect wages, but educational outcomes as well. Presumably, the major effects of abilities on wages are mediated through the endogenous schooling choice (see Piatek and Pinger, 2010). The structural approach pursued by Heckman et al. (2006) and Flossmann et al. (2006) accounts for this issue. Besides wages in general Heckman et al. (2006) also assess the effects of cognitive and noncognitive abilities on wages given certain levels of schooling and on the probability of graduating at certain levels. For instance, for males, noncognitive skills hardly affect the probability of being a regular high school dropout but rather promote the probabilities of being a GED<sup>32</sup> participant, of graduating from high school, of graduating from a two year, and

<sup>32</sup> GED stands for General Educational Development and is a test that certifies college eligibility of US high school graduates.

from a four year college.

Hence, it is of particular interest to identify which traits affect educational performance and along with it schooling choices. Duckworth and Seligman (2005) show that self-discipline even exceeds the explanatory power of IQ in predicting performance at school. They define self-discipline as a hybrid of impulsiveness and self-control. Highly self-disciplined adolescents outperform their peers on all inquired outcomes including average grades, achievement-test scores, and school attendance.

The choice of self-discipline as the noncognitive skill of interest is related to the findings by Wolfe and Johnson (1995). They assess which measure is most eligible for predicting grade point averages (GPA) in a sample of 201 psychology students. The outstanding GPA predictors are measures displaying the level of control and items closely related, like self-discipline. Thus, besides cognitive skills, noncognitive skills play an equally important role in affecting schooling choices or years of schooling, respectively.

Since personality is malleable throughout adolescence and IQ is fairly set earlier in life, the inverse causation also applies. This induces the aforementioned simultaneity. Hansen et al. (2003) determine causal effects of schooling on achievement tests. They reveal that an additional year of schooling increases the Armed Forces Qualification Test (AFQT) score by 3 to 4 points. Achievement tests provide a mixed signal constituted of IQ and personality traits (see Borghans et al., 2011), where IQ is relatively stable from school age on.

Noncognitive abilities also exhibit an intense influence on social outcomes. Closely related to the previously discussed wage achievements are employment status and mean work experience which are likewise substantially affected by the personality.<sup>33</sup> Further outcomes like the probabilities of daily smoking, of incarceration, and of drug abuse are examined and are significantly determined by noncognitive skills, albeit to different extents.

## 7 A Brief Guide to Empirical Analysis

This section intends to give a brief overview on the eligibility of different estimation strategies to deal with the specifics of personality test scores. The previous sections on formation of personality traits and measurement suggest various sources for simultaneity and measurement error. Therefore, one has to scrutinize the data generating process very carefully before using measures obtained from test scores for empirical analysis. There are occasions when measured traits are employed as an outcome variable, usually program evaluation or longitudinal settings, which aim at examining the role of environmental influences on the formation processes. As discussed by Cunha and Heckman (2008), the multiplicity and endogeneity of investments that foster the development of personality

---

<sup>33</sup> See also Heckman et al. (2006) for a detailed discussion and magnitudes.

traits causes a lack of instruments. To overcome the resultant econometric problems one needs proper randomization of investments or structural approaches in the spirit of those employed to generate the results referred to in section 4. Most research questions dealing with noncognitive skills and their relation to human capital outcomes, however, incorporate them as explanatory variables. In this case the threat to parameter consistency of standard regression approaches is more severe. Again, instrumental variable methods are a self-evident response to various kinds of endogeneity problems, but as with the endogeneity of investments in the formation process, it is difficult to find a sufficient number.<sup>34</sup> As an alternative eluding both problems, one can try to correct standard estimates and avoid settings with obvious simultaneity, or one can use latent variable approaches imposing some additional structure.<sup>35</sup> Below, we will elaborate on both approaches.

## 7.1 Adjusted Regression

The virtue of measurement error correction, as opposed to all approaches presented in what follows, is its simplicity. When it comes to more complex structures, including simultaneity, factor approaches are more due. The simplicity of estimation comes at another cost: relatively precise information about the magnitude of measurement error, i.e. reliability, is required. Such a source of information can be reliability measures from classical test theory, which, however, impose very strong assumptions on the relation between true and measured score. For instance, Cronbach's alpha requires scaling parameters (or slope parameters in regression terms) between measured and true score to be equal across items in order to yield a consistent reliability estimate (see Bollen, 1989, for discussion). For most measures this assumption does not hold and reliability is therefore underestimated. Nonetheless, given a decent estimate of the share of measurement error in overall variation of the item sum, it is straightforward to adjust least square estimates by weighting the variation in the erroneous explanatory variables.<sup>36</sup> Unfortunately, accounting for simultaneity still requires proper instruments in this set up. To resolve this problem structural approaches with latent variables are more common.

<sup>34</sup> In conjecture with the problem of attaining the just-identified case, the existence of weak instruments entails further complications.

<sup>35</sup> We follow the definition of Aigner et al. (1984) for latent variables. According to their definition, as opposed to unobserved variables, latent variables cannot be represented as a linear combination of observed variables.

<sup>36</sup> In the univariate case one would simply use

$$\hat{\beta}_A = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 - nVar(\text{error})}.$$

Using an arbitrary coefficient of reliability  $\rho$  this expression can also be written as

$$\hat{\beta}_A = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\rho \sum_{i=1}^n (x_i - \bar{x})^2}.$$

The multivariate case is derived by Schneeweiss (1976) among others.

## 7.2 Methods based on Factor Analysis

Latent variable models are a generalization of error-in-measurement (EIV) models in that in either case the observed personality score is a manifestation of the latent true score (see Aigner et al., 1984). However, the aim of the EIV literature is somewhat different. It primarily intends to obtain consistent estimates when some explanatory variables are erroneous. In contrast, latent variable approaches also aim at estimation of parameters that represent the relationship between latent factors and observed response variables.<sup>37</sup> Most common in estimation are different kinds of maximum likelihood approaches. Parameters which designate the model are referred to as structural parameters, whereas those parameters which vary over observation (like individual latent skills) are denoted incidental parameters.

In EIV models it is common to maximize the conditional likelihoods iteratively or to integrate out the latent factor to obtain a closed form expression. In case of the former, simultaneous ML estimation of structural and incidental parameters can cause severe consistency problems (see Neyman and Scott, 1948).<sup>38</sup> In case of maximizing the marginal likelihood one has to impose mostly overly restrictive distributional assumptions on the latent factors (see Heckman et al., 2006). However, both approaches provide no inference about the latent factors.

In order to overcome this, traditional factor analysis uses other estimation strategies. Given some identification restrictions on the latent factors and their factor loadings (see, e.g. Jöreskog, 1977, and Aigner et al., 1984, for discussion) latent factor structure models can identify both, structural and incidental parameters.<sup>39</sup> Moreover, depending on the number of latent factors, a sufficient number of measurement equations and some knowledge about the structure between the latent factors and other observables is required.<sup>40</sup> The LISREL approach by Jöreskog (1977) estimates the parameters of the complete model by minimizing the discrepancy between the sample correlation matrix and the correlation matrix imputed by the model. This can be attained by different estimation techniques such as maximum likelihood or least squares (see Bollen, 1989, for a comparison of the different approaches with regard to consistency and efficiency). Generally, the scale for the loadings is set by standardizing all observables in the system. Hence, the estimates are only interpretable to a

---

<sup>37</sup> To illustrate the close relation, consider the following notations for an erroneous explanatory variable and a simple factor model with  $x$  representing the observed manifestation of a latent factor  $\theta$  and an unexplained residual  $\varepsilon$ .

$$\begin{aligned}x &= \lambda\theta + \varepsilon \\x &= \theta + \varepsilon\end{aligned}$$

The obvious difference is that factor analysis is interested in identification of both, factor loadings  $\lambda$  and latent factors  $\theta$ .

<sup>38</sup> Baker and Kim (2004) discuss assumptions for iterative estimation to resolve this problem.

<sup>39</sup> It is common in standard factor analysis to assume uncorrelated factors. However, there are infinite combinations of factors and loadings that are in a sense observationally equivalent. The standard approach to warrant identification is to fix the variances of the common factors to unity and to impose sign restrictions on the factors.

<sup>40</sup> This number can be reduced if the underlying theory justifies to fix some parameters or to introduce identities among parameters.

limited extent. One possible way to overcome this limitation is to simulate the model with different specifications of standardized observables and latent factors (see Piatek and Pinger, 2010). A more severe problem in classical factor structure models, however, is the strong distributional dependence of the results. Tractability of the discrepancy function even requires normality assumption for the observables. If these assumptions are not valid for the population, the estimates are inconsistent.<sup>41</sup> Another drawback is that classical factor structure models require linear responses in measurements and outcome equations. The critical assumption for this functional form to apply is that the responses are linear in latent traits, which is very restrictive. The restrictiveness increases as the item scales get smaller.

Carneiro et al. (2003) discuss identification assumptions of more general types of factor structure models accounting for multiple factors and different types of link functions for the response variables. They show how identification of the factor loadings can be established by exploiting covariance structures in conjecture with some additional restrictions. Identification is eased by using dedicated measures, that is, a response variable per latent factor that is only determined by this particular factor. This step even allows for dropping the independence assumption between factors.<sup>42</sup> Carneiro et al. (2003) solve the problem of switching signs causing observationally equivalent structures of factors and loadings by normalizing a particular factor loading in the measurement system to unity. Given a subtle choice of this normalization one can anchor the estimated parameters into appropriate real world outcomes (see, e.g., Cunha and Heckman, 2008) and thus assign an interpretable metric to them. These, along with some further independence and support conditions discussed by Carneiro et al. (2003), finally establish identification of the factor model. When choosing a measurement construct one should always consider that increasing the dimensionality dramatically raises the required number of measurement equations. Therefore, some careful exploratory analysis should be conducted in advance.

Extensions of discrepancy function estimation described above for ordered discrete response variables are available (see Jöreskog and Moustaki, 2001, for an overview). However, these methods become more intricate as the number of factors and, therefore, the number of discrete response equations increases. The same holds for most of the suitable frequentist approaches like Expectation Maximization (EM, Dempster et al., 1977) or Maximum Simulated Likelihood (MSL, Gouriéroux and Monfort, 1991). More flexible approaches drawing on Bayesian techniques, particularly Markov Chain Monte Carlo methods (MCMC), have evolved in the last years due to progress in computational speed.<sup>43</sup> The most convenient properties for estimation of factor structure models for latent abilities

<sup>41</sup> As shown by Heckman et al. (2006), the distributions of latent cognitive and noncognitive skills are non-normal.

<sup>42</sup> In common factor analysis this factor rotation problem prohibits identification without additional assumptions.

<sup>43</sup> Bayesian estimation builds on the notion to enhance the imposed assumptions on the data generation process (which is the likelihood in common reading) by prior beliefs about the parameter distribution. Applying Bayes' Theorem yields a posterior distribution that unifies the assumptions made on the data generating process and the

has the Gibbs sampler (Geman and Geman, 1984) and its extension due to Tanner and Wong (1987), the so-called Data Augmentation (see also van Dyk and Meng, 2001). It uses a simple contrivance for ease of computation. Instead of relying on the posterior of the parameters conditional on observable responses, latent responses, and latent factors for estimation, it explicitly models the posterior for parameters and latent elements conditional on observed data. This is achieved by integrating over the product of the latent factors, latent responses and the initial posterior conditional on both. Since it is only a reformulation, the marginal distributions implied by the augmented distribution have to coincide with the original posterior. The procedure is a two step version of the Gibbs sampler. First, the latent components are drawn from their distributions conditional on data and parameters within the so-called imputation step and then draws for the parameters conditional on those from the imputation step are conducted in a posterior step. The algorithm cycles between imputation and posterior step until convergence.<sup>44</sup> A very precise summary for practitioners along with some refinements is provided by Piatek (2010).

Given covariance structures, only the first two moments of the distributions of latent factors can be identified and therefore is not sufficient for relaxation of normality assumptions. Carneiro et al. (2003) show conditions under which the complete distribution of a latent factor is nonparametrically identified.<sup>45</sup> The imputation step can then be sampled from a mixture of normals which provides enough flexibility to approximate any other distribution (see Diebolt and Robert, 1984).<sup>46</sup> Another virtue of the Data Augmentation Algorithm is that the sampled results for the individual latent factors can be stored after convergence. For a better illustration and interpretation of the results it is common to use them to simulate the outcomes of interest for different percentile ranges of the latent factors and, for instance, the mean of the observable variables (see Piatek and Pinger, 2010). This yields a graphical representation of the results as shown in Figure 1.

---

unconditional parameter distribution (a comprehensive introduction to the topic is provided by Geweke, 2005). Formally, this implies

$$p(\Theta|data) = \frac{f(data|\Theta)f(\Theta)}{f(data)} \propto L(data|\Theta)f(\Theta),$$

where  $\Theta$  is a set of parameters,  $p(\Theta|data)$  is the posterior,  $L(data|\Theta)$  is the data generating process or the Likelihood, and  $f(\Theta)$  is the imposed prior distribution of the parameters. The outstanding virtue of this approach is its ability to deal with lacking closed forms. Estimates from the posterior distribution can be obtained in different manners. One possibility is to compute marginal distributions for the parameters of interest by means of numerical or Monte Carlo integration and obtaining the respective moments of the marginal as a spinoff. Another way is to directly simulate draws from the posterior. Both methods become increasingly complex as the dimensionality of the posterior rises. This is where MCMC algorithms come into play (Gilks et al., 1996, provide a detailed treatise of the topic). MCMC algorithms sample from chains of conditional distributions more simple in nature than the posterior. The simulation finally converges to draws from the joint posterior. The resulting chains fulfil the Markov chain condition in that transition probabilities between states only depend on the current state and that after a sufficient number of iterations the probability of being in particular state is independent of the initial state (stationarity).

<sup>44</sup> In that sense, Data Augmentation is the Bayesian equivalent to EM algorithm.

<sup>45</sup> Particularly, the identification requires combinations of continuous and discrete response variables.

<sup>46</sup> The implementation is also discussed in Piatek (2010).

### 7.3 Item Response Theory

Another strand of the psychometric literature which arose from classical test theory is item response theory (IRT, see Sijtsma and Junker, 2006, for a historical classification). It traces back to the work of Lazarsfeld (1950) and Lord (1952). The breakthrough contributions in the psychometric field are by Lord and Novick (1968) and Samejima (1952). The basic notion that IRT exploits is the probabilistic relation between latent ability and categorical responses by analyzing the pattern across observations and items. The most convenient way to illustrate IRT is to consider dichotomous items. If abilities are positively related to the response, the response probability is usually modeled by means of a normal or logistic cumulative density function, the so-called item characteristic curve (ICC).<sup>47</sup> In terms of the normal density the mean provides the scale location, which in some sense is the difficulty, whereas the variance determines the discriminatory power of the item. The extension to polytomous responses is analogous. The ICC for the lowest response on the scale has an inverted shape, whereas the ICC for the highest response has the usual CDF form. The scale realizations in between exhibit a decreasing response probability towards both bounds of the scale and therefore have a bell shape with probability mass at different locations. A more consistent functional form for estimation is obtained by treating different permutations as dichotomous and cumulate the ICCs accordingly. For  $m$  response categories this yields  $m - 1$  non-intersecting and monotonic characteristic curves of the usual shape.

The aim is then to estimate the parameters determining the shape of the ICC, i.e., the structural parameters, and the incidental parameters representing the latent abilities of the assessed individuals. Given some independence assumptions traditional estimation approaches use an iterative maximum likelihood procedure cycling between conditional likelihoods for incidental and structural parameters (see Birnbaum, 1968). As pointed out by Neyman and Scott (1948), simultaneous ML estimation of structural and incidental parameters can lead to severe inconsistency. Eluding these problem requires assumptions only valid for models with a discrimination parameter fixed at unity, so-called Rasch Models (see Andersen, 1972). Enhancements in computational speed have led to more robust strategies like the Bock and Aitken solution (Bock and Aitken, 1981) which uses EM over numerically integrated marginal likelihoods.<sup>48</sup>

The previous techniques, however, impose several parametric assumptions to enforce properties like monotonicity and non-intersection. There are more flexible contributions from the econometric literature which rely on semiparametric estimation approaches.<sup>49</sup> Conditions for this kind of model to be feasible are established by Spady (2006) and comprise monotonicity, stochastic dominance and local

---

<sup>47</sup> For the opposite relation the survival function can be used.

<sup>48</sup> See Baker and Kim (2004) for a discussion of further methods. Common Item Response Models can be formulated as Generalized Linear Latent and Mixed Models (GLLAMM) and estimated accordingly (see, Rijmen et al., 2003, Rabe-Hesketh et al., 2007, for a detailed treatise).

<sup>49</sup> Following the notation of Chen (2007) the applied methods are semi-nonparametric. See also Härdle et al. (2004) and Horowitz (2009) for a comprehensive overview of such techniques.

independence assumptions (i.e. independence of item responses conditional on the latent attitude). Suggestions for the implementation of potential estimation procedures are provided by Spady (2007) and Weiss (2010).<sup>50</sup> The basic notion is to exploit the joint empirical distribution of item responses to obtain the item characteristic curves.<sup>51</sup> As opposed to the fully parametric IRT approaches discussed above, the semiparametric version only imposes a specific distribution for the latent trait conditional on background information and approximates the response functions nonparametrically. The most direct way is to use a Sieve Maximum Likelihood strategy established by Grenander (1981).<sup>52</sup> Alternatively, the response functions can be approximated via an Exponential Tilting procedure (see Barron and Sheu, 1991).<sup>53</sup> Given a set of items the estimates can be used to estimate expected values (or other moments) of the latent ability location on a continuum like the interval  $[0, 1]$ .<sup>54</sup> If an interpretable scale is chosen the estimated latent ability can be directly employed for further regression analysis.<sup>55</sup>

## 8 Concluding Remarks

This paper has reviewed the recent influential literature that considers the role of noncognitive skills as a determinant of human capital. In addition, a selection of the empirical evidence that highlights the determinism of crucial achievements and outcomes as a result of these skills has been briefly summarized. Moreover, we have discussed the notion of noncognitive skills in light of the relevant psychological literature. In terms of measuring noncognitive skills, empirical research in economics strongly benefits from psychometric concepts, and economists should be aware of the underlying assumptions. The commonly used approaches to measure personality traits are not completely conclusive. On the one hand, overall measures tend to be too general in that they veil important variation, whereas on the

<sup>50</sup> Conditional independence also implies that background characteristics affect the response probabilities only through latent abilities.

<sup>51</sup> Formally this means that the observed pattern is assumed to be generated by the following ML setup:

$$p(r_1, r_2, \dots, r_m | X) = \int p_1(r_1 | \theta) p_2(r_2 | \theta) \dots p_m(r_m | \theta) f(\theta | X) d\theta,$$

where  $p_i$  ( $i = 1, \dots, m$ ) is the response probability for the  $i$ th item,  $\theta$  is a scalar latent skill, and  $X$  is a set of background characteristics.

<sup>52</sup> Sieve methods can be extended to other estimators than ML. Without distributional assumptions the parameter space for the response functions is infinite and maximization of the criterion function is therefore infeasible. The Method of Sieves defines a series of approximation spaces in order to reduce dimensionality of the previously infinite dimensional parameter space. For concave optimization problems with finite dimensional linear sieve spaces, this technique is also denoted series estimation (see Geman and Hwang, 1982, Barron and Sheu, 1991, and Chen, 2007 for technical overview). Appropriate base functions are orthogonal polynomials, trigonometric polynomials and shape-preserving splines, just to mention a few (see Härdle, 1994, and Chen, 2007).

<sup>53</sup> The tilting procedure for density approximation is part of the computationally feasible optimization problem of the Exponential Tilting estimator discussed by Kitamura and Stutzer (1997) and Imbens et al. (1998).

<sup>54</sup> Applying Bayes' Theorem, the distribution providing the expectations can be computed as follows:

$$f(\theta | r, X) = \frac{f(r, \theta | X)}{p(r | X)} = \frac{p_1(r_1 | \theta) \dots p_m(r_m | \theta) f(\theta | X)}{\int p_1(r_1 | \theta) p_2(r_2 | \theta) \dots p_m(r_m | \theta) f(\theta | X) d\theta},$$

where the numerator uses the distribution imposed on  $\theta$  conditional on  $X$  together with the estimates for the item response functions and the denominator is its integral obtained by numerical integration.

<sup>55</sup> See Spady (2007) for simultaneous estimation strategies.

other hand, measures of specific personality traits may put the researcher to a hard choice regarding their adequacy. As shown, psychometric coefficients that assess the eligibility of constructs all have their own limitations.

With regard to formation and stratification of skills, the role and the timing of educational and parental investments have been proven to be crucial in the empirical literature. Regardless of the particular effects, virtually all empirical studies suggest a joint conclusions: early investments are most crucial, but nonetheless, should be complemented later on. Early neglect, on the other hand, cannot be compensated in later stages of life without prohibitively high costs. Hence, in terms of support for low skilled or disadvantaged individuals the focus should be on early preschool age. Given this pattern for the intertemporal allocation of resources, the role of schooling investments is rather subordinate. The Cunha-Heckman model formalizes this process by means of a dynamic production function and also provides parameter estimates. Though the estimation approach accounts for measurement error, the insights on pattern and transmission of parental investments are far from definite. Attributing parental traits to preferences like altruism that allow to model the investment behavior of parents into their children is a complex but desirable aim.

As has been briefly discussed, noncognitive skills are important determinants of several outcomes, like educational achievement and labor market success. The revealed pattern for different personality traits are relatively unequivocal across studies. However, yet it is mostly unclear in how far the compound of productivity enhancement, occupational sorting, wage premia due to social desirability, and self-selection affects the results.

Personality measures applied within an econometric framework tend to suffer from measurement error, simultaneity bias, and spurious influences by other unobservables. Due to these issues, the relation between personality traits and economic preference parameters is still patchwork and leaves many unanswered questions. Drawing inference on correlations between traits and preferences is a necessary first step but provides only cursory results. A better theoretical understanding of the pathways between both concepts is inevitable in order to obtain more conclusive insights. As a consequence, empirical analysis urges adequate methods. The literature reviewed in Section 7 has given an introductory glance on appropriate factor analytic methods and those based on Item Response Theory.

The summarized findings of this very recent literature enrich the traditional view on human capital in economics by considering noncognitive skills as an additional determinant of lifetime labor market and social outcomes. Moreover, the essential role of infancy and early childhood in producing these outcomes is accentuated. This provides new policy implications. First, good parenting is (and will remain) the major source of educational success; this is only indirectly driven by family income. Therefore, intervention policies should be adopted already at preschool age and should primarily focus

on home environment. Second, the time interval for sufficient governmental influence is more limited in case of cognitive skills than for noncognitive skills. The malleability of personality throughout adolescence and beyond provides a powerful and instantaneous policy tool. Nonetheless, this is just a crude guidance originating from an evolving literature. Both, the optimal timing and intensity for reducing upcoming and existent inequalities remain still to be determined.

## References

- AIGNER, D. J., C. HSIAO, A. KAPTEYN, AND T. WANSBEEK (1984): “Latent Variables Models in Econometrics,” in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2, chap. 23, pp. 1321–1393. Elsevier B.V., Amsterdam, North Holland.
- AIYAGARI, R., J. GREENWOOD, AND A. SESHADRI (2002): “Efficient Investment in Children,” *Journal of Economic Theory*, 102(2), 290–321.
- ALLPORT, G. W. (1937): *Personality: A Psychological Interpretation*. Holt, Rhinehart and Winston, New York.
- (1961): *Pattern and Growth in Personality*. Holt, Rinehart and Winston, New York, NY.
- ALLPORT, G. W., AND H. ODBERT (1936): “Trait-Names, a Psycholexical Study,” *Psychological Monographs*, 47(211).
- ALMLUND, M., A. L. DUCKWORTH, J. HECKMAN, AND T. KAUTZ (2011): “Personality Psychology and Economics,” IZA Discussion Papers 5500, Institute for the Study of Labor (IZA), Bonn.
- ANDERSEN, E. B. (1972): “The Numerical Solution of a Set of Conditional Estimation Equations,” *Journal of the Royal Statistical Society, Series B*, 34(1), 42–54.
- ANDREWS, S. R., J. B. BLUMENTHAL, D. L. JOHNSON, C. J. FERGUSON, T. M. LASATER, P. E. MALONE, AND D. B. WALLACE (1982): “The Skills of Mothering: A Study of Parent Child Development Centers,” *Monographs of the Society for Research in Child Development*, 47(6), 1–83.
- ANTECOL, H., AND D. A. COBB-CLARK (2010): “Do Non-Cognitive Skills Help Explain the Occupational Segregation of Young People?,” IZA Discussion Paper Series 5093, Institute for the Study of Labor (IZA), Bonn.
- BAKER, F. B., AND S.-H. KIM (2004): *Item Response Theory*. Dekker Media, New York.
- BANDURA, A. (1986): *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice Hall, New Jersey.

- BARRON, A. R., AND C.-H. SHEU (1991): "Approximation of Density Functions by Sequences of Exponential Families," *The Annals of Statistics*, 19(3), 1347–1369.
- BECKER, G., AND N. TOMES (1986): "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, 75(4), 1–39.
- BECKER, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, The University of Chicago Press. National Bureau of Economic Research, New York, Third Edition from 1993.
- BEN-PORATH, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *The Journal of Political Economy*, 75(4), 352–365.
- BIRNBAUM, A. (1968): "Some Latent Trait Models and their Use in Inferring an Exmaminee's Ability," in *Statistical Theories of Mental Test Scores*, ed. by F. M. Lord, and M. R. Novick, pp. 397–479. Addison-Wesley, Reading, MA.
- BLACKBURN, M., AND D. NEUMARK (1992): "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," *Quarterly Journal of Economics*, 107(4), 1421–1436.
- BLOMEYER, D., K. CONEUS, M. LAUCHT, AND F. PFEIFFER (2009): "Initial risk matrix, home resources, ability development and children's achievement," *Journal of the European Economic Association*, 7(2-3), 638–648.
- BOCK, R. R., AND M. AITKEN (1981): "Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm," *Psychometrika*, 46(4), 443–459.
- BOLLEN, K. A. (1989): *Structural Equations with Latent Variables*. John Wiley & Sons, New York, NY.
- BORGHAN, L., B. H. GOLDSTEYN, J. HECKMAN, AND J. E. HUMPHRIES (2011): "Identification Proglems in Personality Psychology," IZA Discussion Paper Series 5605, Institute for the Study of Labor (IZA), Bonn.
- BORGHANS, L., A. L. DUCKWORTH, J. HECKMAN, AND B. TER WEEL (2008a): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43(4), 972–1059.
- BORGHANS, L., H. MEIJERS, AND B. TER WEEL (2008b): "The Role of Noncognitive Skills in Explaining Cognitive Test Scores," *Economic Inquiry*, 46(1), 2–12.
- BORGHANS, L., B. TER WEEL, AND B. A. WEINBERG (2008c): "Interpersonal Styles and Labor Market Outcomes," *Journal of Human Resources*, 43(4), 815–858.

- BOUCHARD, T. J., AND J. C. LOEHLIN (2001): “Genes, Evolution, and Personality,” *Behavior Genetics*, 31(3), 243–273.
- BOUND, J., Z. GRILICHES, AND B. H. HALL (1986): “Wages, Schooling and IQ of Brothers and Sisters: Do the Family Factors Differ?,” *International Economic Review*, 27(1), 77–105.
- BOWLES, S., H. GINTIS, AND M. OSBORNE (2001): “The Determinants of Earnings: A Behavioral Approach,” *Journal of Economic Literature*, 38(4), 1137–1176.
- BRAAKMANN, N. (2009): “The Role of Psychological Traits for the Gender Gap in Full-Time Employment and Wages: Evidence from Germany,” DIW Discussion Papers 162, German Institute for Economic Research (DIW), Berlin.
- BUSS, A. H., AND R. PLOMIN (1975): *A Temperament Theory of Personality Development*. John Wiley & Sons, New York.
- CAMPBELL, D. T., AND D. W. FISKE (1959): “Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix,” *Psychological Bulletin*, 56(2), 81–105.
- CAMPBELL, F. A., AND C. T. RAMEY (1994): “Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-Up Study of Children from Low-Income Families,” *Child Development*, 65, 684–698.
- CANLI, T. (2006): *Biology of Personality and Individual Differences*. Guilford Press, New York.
- CARNEIRO, P., C. CRAWFORD, AND A. GOODMAN (2007): “The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes,” CEE Discussion Papers 0092, Centre for the Economics of Education at the London School of Economics (CEE), London.
- CARNEIRO, P., K. HANSEN, AND J. HECKMAN (2003): “Estimating Distributions of the Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice,” *International Economic Review*, 44(2), 361–421.
- CASPI, A. (2000): “The Child Is Father of the Man: Personality Continuities From Childhood to Adulthood,” *Journal of Child Psychology and Psychiatry*, 78(1), 158–172.
- CASPI, A., J. MCCLAY, T. E. MOFFITT, J. MILL, J. MARTIN, I. W. CRAIG, A. TAYLOR, AND R. POULTON (2002): “Role of Genotype in Cycle of Violence Maltreated Children,” *Science*, 297, 851–853.
- CASPI, A., B. W. ROBERTS, AND R. L. SHINER (2005): “Personality Development - Stability and Change,” *Annual Review of Psychology*, 56(1), 453–484.

- CATTELL, R. (1971): *Abilities: Their Structure, Growth and Action*. Houghton Mifflin, Boston.
- CHEN, X. (2007): “Large Sample Sieve Estimation of Semi-Nonparametric Models,” in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6B, chap. 76, pp. 5549–5632. Elsevier B.V., Amsterdam, North Holland.
- COASE, R. H. (1937): “The Nature of the Firm,” *Economica*, 4(16), 386–405.
- COLEMAN, M., AND T. DELEIRE (2003): “An Economic Model of Locus of Control and the Human Capital Investment Decision,” *Journal of Human Resources*, 38(3), 701–721.
- CONEUS, K., J. GERNANDT, AND M. SAAM (2009): “Noncognitive Skills, School Achievements and Educational Dropout,” ZEW Discussion Paper 19, Centre for European Economic Research (ZEW), Mannheim.
- CONEUS, K., AND M. LAUCHT (2008): “The Effect of Early Noncognitive Skills on Social Outcomes in Adolescence,” ZEW Discussion Paper 115, Centre for European Economic Research (ZEW), Mannheim.
- CONEUS, K., AND F. PFEIFFER (2007): “Self-Productivity in Early Childhood,” ZEW Discussion Paper 053, Centre for European Economic Research (ZEW), Mannheim.
- COSTA, P., AND R. MCCRAE (1992): *Revised Neo Personality Inventory (NEO PI-R) and Neo Five-Factor (NEO-FFI)*. Psychological Assessment Resources, Odessa, FL.
- COSTA, P. T., AND R. R. MCCRAE (2008): “Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI),” in *The SAGE Handbook of Personality Theory and Assessment: Personality Measurement and Testing*, ed. by G. J. Boyle, G. Matthews, and D. H. Saklofske, vol. 2, pp. 179–196. SAGE Publications Ltd., Thousand Oaks, California.
- CRONBACH, L. J. (1951): “Coefficient Alpha and the Internal Structure of Tests,” *Psychometrika*, 16(3), 297–334.
- CUNHA, F., AND J. HECKMAN (2006): “Investing in our Young People,” unpublished, available at <http://www-news.uchicago.edu/releases/06/061115.education.pdf>.
- (2007): “The Technology of Skill Formation,” *American Economic Review*, 97(2), 31–47.
- (2008): “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation,” *Journal of Human Resources*, 43(4), 738–782.
- (2009): “The Economics and Psychology of Inequality and Human Development,” *Journal of the European Economic Association*, 7(2/3), 320–364.

- CUNHA, F., J. HECKMAN, L. LOCHNER, AND D. MASTEROV (2006): “Interpreting the Evidence on Life Cycle Skill Formation,” in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, and F. Welch, vol. 1, chap. 12, pp. 697–812. Elsevier B.V., Amsterdam, North-Holland.
- CUNHA, F., J. HECKMAN, AND S. SCHENNACH (2010): “Estimating the Technology of Cognitive and Noncognitive Skill Formation,” *Econometrica*, 78(3), 883–991.
- CURRIE, J. (2009): “Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development,” *Journal of Economic Literature*, 47(1), 87–122.
- CURRIE, J., AND D. THOMAS (1995): “Does Head Start Make a Difference?,” *American Economic Review*, 85(3), 341–364.
- DAHL, R. E. (2004): “Adolescent Brain Development: A Period of Vulnerabilities and Opportunities,” in *Annals of the New York Academy of Sciences 1021*, pp. 1–22. New York Academy of Sciences.
- DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN (1977): “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.
- DIEBOLT, J., AND C. P. ROBERT (1984): “Estimation of Finite Mixture Distributions through Bayesian Sampling,” *Journal of the Royal Statistical Society, Serie B*, 56(2), 363–375.
- DIGMAN, J. M. (1997): “Higher-Order Factors of the Big Five,” *Journal of Personality and Social Psychology*, 73(6), 1246–1256.
- DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2007): “Are Risk Aversion and Impatience Related to Cognitive Ability?,” IZA Discussion Papers 2735, Institute for the Study of Labor (IZA), Bonn.
- (2008): “Representative Trust and Reciprocity: Prevalence and Determinants,” *Economic Inquiry*, 46(1), 84–90.
- DUCKWORTH, A. L., AND M. SELIGMAN (2005): “Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents,” *American Psychological Society*, 16(12), 939–944.
- DUTTWEILER, P. (1984): “The Internal Control Index: A Newly Developed Measure of Locus of Control,” *Educational and Psychological Measurement*, 44(2), 209–221.
- EYSENCK, H. J. (1991): “Dimensions of Personality: Sixteen, Five or Three? Criteria for a Taxonomic Paradigm,” *Personality and Individual Differences*, 8(12), 773–790.
- FEHR, E., AND S. GÄCHTER (2000): “Fairness and Retaliation: The Economics of Reciprocity,” *Journal of Economic Perspectives*, 14(3), 159–181.

- FEHR, E., AND K. M. SCHMIDT (2006): “The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories,” in *Handbook of the Economics of Giving, Altruism and Reciprocity*, ed. by S.-C. Kolm, and J. M. Ythier, vol. 1, pp. 615–699. Elsevier B.V., Amsterdam, North-Holland.
- FLOSSMANN, A., R. PIATEK, AND L. WICHERT (2007): “Going Beyond Returns to Education: The Role of Noncognitive Skills on Wages in Germany,” unpublished, available at <http://www.afse.fr/docs/congres2007/docs/619716flopivi.pdf>.
- FORTIN, N. M. (2008): “The Gender Wage Gap among Young Adults in the United States - The Importance of Money versus People,” *Journal of Human Resources*, 43(4), 884–918.
- FRAGA, M., E. BALLESTAR, M. PAZ, S. ROPERO, F. SETIEN, M. BALLESTAR, D. HEINE-SUNER, J. CIGUDOSA, M. URIOSTE, J. BENITEZ, M. BOIX-CHORNET, A. SANCHEZ-AGUILERA, C. LING, E. CARLSSON, P. POULSEN, A. VAAG, Z. STEPHAN, T. SPECTOR, Y.-Z. WU, C. PLASS, AND M. ESTELLER (2005): “Epigenetic Differences Arise during the Lifetime of Monozygotic Twins,” *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), 10604–10609.
- FUERST, J. S., AND D. FUERST (1993): “Chicago Experience with an Early Childhood Programme: The Special Case of the Child Parent Center Program,” *Educational Research*, 35(3), 237–253.
- GEMAN, S., AND D. GEMAN (1984): “Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), 609–628.
- GEWEKE, J. (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, New York, NY.
- GILKS, W. R., S. RICHARDSON, AND D. SPIEGELHALTER (1996): *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
- GOLDBERG, L. (1971): “A Historical Survey of Personality Scales and Inventories,” in *Advances in Psychological Assessment*, ed. by P. McReynolds, pp. 293–336. Science and Behavior Books, Palo Alto, California.
- GOLDSMITH, H. H., A. H. BUSS, R. PLOMIN, M. K. ROTHBART, A. THOMAS, S. CHESS, R. A. HINDE, AND R. B. MCCALL (1987): “Roundtable: What is temperament? Four approaches,” *Child Development*, 58, 505–529.

- GOURIÉROUX, C., AND A. MONFORT (1991): “Simulation Based Inference in Models with Heterogeneity,” *Annales d’Economie et de Statistique*, 20/21, 69–107.
- GRENANDER, U. (1981): *Abstract Inference*. John Wiley & Sons, New York, NY.
- GRILICHES, Z. (1977): “Estimating the Returns to Schooling: Some Econometric Problems,” *Econometrica*, 45(1), 1–22.
- GROSSMAN, J., AND J. TIERNEY (1998): “Does Mentoring Work? An Impact Study of the Big Brothers Big Sisters Program,” *Evaluation Review*, 22(3), 403–426.
- HANSEN, K., J. HECKMAN, AND K. MULLEN (2004): “The Effect of Schooling and Ability on Achievement Test Scores,” *Journal of Econometrics*, 121(1/2), 39–98.
- HAUSE, J. C. (1972): “Earnings Profile: Ability and Schooling,” *Journal of Political Economy*, 80(3, Part 2), 108–138.
- HECKMAN, J. (2008): “Schools, Skills, and Synapses,” *Economic Inquiry*, 46(3), 289–324.
- HECKMAN, J., AND Y. RUBINSTEIN (2001): “The Importance of Noncognitive Skills: Lessons from the GED Testing Program,” *The American Economic Review*, 91(2), 145–149.
- HECKMAN, J., J. STIXRUD, AND S. URZUA (2006): “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 24(3), 411–482.
- HEINECK, G., AND S. ANGER (2010): “The Returns to Cognitive Abilities and Personality Traits in Germany,” *Labour Economics*, 17(3), 535–546.
- HEINICKE, K., AND S. L. THOMSEN (2011): “Heterogeneous Returns to Personality - The Role of Occupational Choice,” Discussion paper, University of Magdeburg, mimeo.
- HILL, J. L., J. BROOKS-GUNN, AND J. WALDFOGEL (2002): “Sustained Effects of High Participation in Early Intervention for Low-Birth-Weight Premature Infants,” *Developmental Psychology*, 39(4), 730–744.
- HOPKINS, K., AND G. BRACHT (1975): “Ten-Year Stability of Verbal and Nonverbal IQ Scores,” *American Educational Research Journal*, 12(4), 469–477.
- HOROWITZ, J. L. (2009): *Semiparametric and Nonparametric Methods in Econometrics*. Springer, Heidelberg.
- HOUGH, L. M. (1992): “The ”Big Five” Personality Variables - Construct Confusion: Description versus Prediction,” *Human Performance*, 5(1), 139–155.

- HÄRDLE, W., M. MÜLLER, A. WERWATZ, AND S. SPERLICH (2004): *Nonparametric and Semiparametric Models*. Springer, Heidelberg.
- IMBENS, G. W., R. H. SPADY, AND P. JOHNSON (1998): “Information Theoretic Approaches to Inference in Moment Condition Models,” *Econometrica*, 66(2), 333–357.
- JACKSON, D. (1976): *Jackson Personality Inventory Test Manual*. Research Psychologists Press, Goshen, New York.
- JOHNSON, A. (1996): “An Evaluation of Long-Term Impacts of the Sponsor-a-Scholar Program on Student Performance,” *Mathematica Policy Research Inc.*
- JOHNSON, D. L., AND T. WALKER (1991): “A Follow-up Evaluation of Houston Parent-Child Development Center: School Performance,” *Journal of Early Intervention*, 15(3), 226–236.
- JOHNSON, J. S., AND E. L. NEWPORT (1989): “Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language,” *Cognitive Psychology*, 21, 60–99.
- JÖRESKOG, K. (1977): “Structural Equation Models in the Social Sciences: Specification, Estimation and Testing,” in *Applications of Statistics*, ed. by P. R. Krishnaiah, pp. 265–287. Elsevier B.V., Amsterdam, North Holland.
- JÖRESKOG, K. G., AND I. MOUSTAKI (2001): “Factor Analysis of Ordinal Variables: A Comparison of Three Approaches,” *Multivariate Behavioral Research*, 36(3), 347–387.
- KENDALL, P., AND L. WILCOX (1979): “Self-Control in Children: Development of a Rating Scale,” *Journal of Consulting and Clinical Psychology*, 47, 1020–1029.
- KIRBY, K. N. (1997): “Bidding on the Future: Evidence Against Normative Discounting of Delayed Rewards,” *Journal of Experimental Psychology. General*, 126(1), 54–70.
- KITAMURA, Y., AND M. STUTZER (1997): “An Information-Theoretic Alternative to Generalized Methods of Moments Estimation,” *Econometrica*, 65(4), 861–874.
- KNUDSEN, E. (2004): “Sensitive Periods in the Development of the Brain and Behavior,” *Journal of Cognitive Neuroscience*, 16(4), 1412–1425.
- LAZARUSFELD, P. F. (1950): “The Logical and Mathematical Foundation of Latent Structure Analysis,” in *Measurement and Prediction*, ed. by S. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Stas, and J. A. Clausen, pp. 362–412. Princeton University Press, Princeton, NJ.

- LEIBOWITZ, A. (1974): "Home Investments in Children," *Journal of Political Economy*, 82(2, Part 2), 111–131.
- LORD, F. M. (1952): *A Theory of Test Scores*. Psychometric Monograph Nr. 7, Psychometric Society.
- LORD, F. M., AND M. R. NOVICK (1968): *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA.
- MADDEN, G. J., A. M. BEGOTKA, B. R. RAIFF, AND L. L. KASTERN (2003): "Delay Discounting of Real and Hypothetical Rewards," *Experimental and Clinical Psychopharmacology*, 11(2), 139–145.
- MISCHEL, W. (1968): *Personality and Assessment*. John Wiley & Sons, New York.
- (1973): "Toward a Cognitive Social Learning Reconceptualization of Personality," *Psychological Review*, 80(4), 246–268.
- MISCHEL, W., AND Y. SHODA (1995): "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure," *Psychological Review*, 102(2), 246–268.
- MORGESON, F., M. CAMPION, R. DIPBOYE, J. HOLLENBECK, K. MURPHY, AND N. SCHMITT (2007): "Reconsidering the use of Personality Tests in Personnel Selection Contexts," *Personal Psychology*, 60, 683–729.
- MUELLER, G., AND E. PLUG (2006): "Estimating the Effect of Personality on Male and Female Earnings," *Industrial and Labor Relations Review*, 60(1), 3–22.
- MULAİK, S. A. (2010): *Foundations of Factor Analysis*. CRC Press, Boca Raton, FL.
- MURNANE, R. J., J. B. WILLETT, M. J. BRAATZ, AND Y. DUHALDEBORDE (2000): "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.
- NEWPORT, E. (1990): "Maturational Constraints on Language Learning," *Cognitive Science*, 14(1), 11–28.
- NEYMAN, J., AND E. L. SCOTT (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16(1), 1–32.
- O'CONNOR, T. G., M. RUTTER, C. BECKETT, L. KEAVENEY, J. M. KREPPNER, AND THE ENGLISH AND ROMANIAN ADOPTEES STUDY TEAM (2000): "The Effects of Global Severe Privation on Cognitive Competence: Extension and Longitudinal Follow-Up," *Child Development*, 71(2), 376–390.

- OROM, H., AND D. CERVONE (2009): “Personality Dynamics, Meaning, and Idiosyncrasy: Assessing Personality Architecture and Coherence Idiographically,” *Journal of Research in Personality*, 43(2), 228–240.
- OSBORNE-GROVES, M. (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.
- PERVIN, L., D. CERVONE, AND O. JOHN (2005): *Theories of Personality*. John Wiley & Sons, New York, NY.
- PIATEK, R. (2010): “Bayesian Inference for Factor Structure Models via Gibbs Sampling,” CMS Discussion Paper 2010-03, Center for Quantitative Methods and Survey Research (CMS), Konstanz.
- PIATEK, R., AND P. PINGER (2010): “Maintaining (Locus of) Control? Assessing the Impact of Locus of Control on Education Decisions and Wages,” IZA Discussion Paper Series 5289, Institute for the Study of Labor (IZA), Bonn.
- RABE-HESKETH, S., A. SKRONDAL, AND X. ZHENG (2007): “Multilevel Structural Equation Modeling,” in *Handbook of Computing and Statistics with Applications*, ed. by E. Kontoghiorghes, vol. 1, pp. 209–227. Elsevier B.V., Amsterdam, North Holland.
- RIJMEN, F., F. TUERLINCKX, P. D. BOECK, AND P. KUPPENS (2003): “A Nonlinear Mixed Model Framework for Item Response Theory,” *Psychological Methods*, 8(2).
- ROBERTS, B., P. HARMS, J. SMITH, D. WOOD, AND M. WEBB (2006): “Using Multiple Methods in Personality Psychology,” in *Handbook of Multimethod Measurement in Psychology*, ed. by M. Eid, and E. Diener, chap. 22, pp. 321–335. American Psychological Association, Washington, DC.
- ROBERTS, B., N. KUNCEL, R. SHINER, A. CASPI, AND L. GOLDBERG (2007): “The Power of Personality,” *Perspectives on Psychological Science*, 2(4), 313–345.
- ROBERTS, B., K. WALTON, AND W. VIECHTBAUER (2006): “Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudinal Studies,” *Psychological Bulletin*, 132, 1–25.
- ROBERTS, B. W. (2009): “Back to the Future: Personality and Assessment and Personality Development,” *Journal of Research in Psychology*, 43(2), 137–145.
- ROBERTS, B. W., AND W. F. DELVECCHIO (2000): “The rank-order consistency of personality from childhood to old age: A quantitative review of longitudinal studies,” *Psychological Bulletin*, 126, 3–25.

- ROSENBERG, M. (1965): *Society and the Adolescent Self-Image*. Princeton University Press, Princeton, New Jersey.
- ROTHBART, M. K. (1981): "Measurement of Temperament in Infancy," *Child Development*, 52, 569–578.
- ROTHBART, M. K., S. A. AHADI, AND D. E. ADAMS (2000): "Temperament and Personality: Origins and Outcomes," *Journal of Personality and Social Psychology*, 78(1), 122–135.
- ROTHBART, M. K., AND M. BATES (2006): "Temperament," in *Handbook of Child Psychology Vol. 3 - Social, Emotional and Personality Development*, ed. by W. Damon, N. Eisenberg, and R. M. Lerner, pp. 99–166. John Wiley & Sons, New York.
- ROTTER, J. (1966): *Generalized Expectancies for Internal versus External Control of Reinforcement*. American Psychological Association, Washington, DC.
- SALOVEY, P., M. BRACKETT, AND J. D. MAYER (2004): *Emotional Intelligence - Key Readings on the Mayer and Salovey Model*. National Professional Resources Inc., Port Chester, NY.
- SAMEJIMA, F. (1952): "*Estimation of Latent Ability Using a Response Pattern of Graded Scores*". Psychometric Monograph Nr. 17, Psychometric Society.
- SCHNEEWEISS, H. (1976): "Consistent Estimation of a Regression with Errors in Variables," *Metrika*, 23, 101–115.
- SCHUMPETER, J. A. (1934): *The Theory of Economics Development*. Oxford University Press, London.
- SCHWEINHART, L. J., H. V. BARNES, AND D. P. WEIKART (1993): *Significant Benefits - The High/Scope Perry Preschool Study Through Age 27*. High/Scope Educational Research Foundation, Ypsilanti, Michigan.
- SHINER, R., AND A. CASPI (2003): "Personality Differences in Childhood and Adolescence: Measurement, Development, and Consequences," *Journal of Child Psychology and Psychiatry*, 44(1), 2–32.
- SHONKOFF, J., AND D. PHILLIPS (2000): *From Neurons to Neighborhoods: The Science of Early Child Development*. National Academy Press, Washington D.C.
- SIJTSMA, K., AND B. W. JUNKER (2006): "Item Response Theory: Past Performance, Present Developments, and Future Expectations," *Behaviormetrika*, 33(1), 75–102.

- SPADY, R. H. (2006): “Identification and Estimation of Latent Attitudes and their Behavioral Implications,” CEMMAP Working Paper CWP12/06, Centre for Micro Data Methods and Practice (CEMMAP), London.
- (2007): “Semiparametric Methods for the Measurement of Latent Attitudes and the Estimation of their Behavioral Consequences,” CEMMAP Working Paper CWP26/07, Centre for Micro Data Methods and Practice (CEMMAP), London.
- TANGNEY, J., R. BAUMEISTER, AND M. BOONE (2004): “High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success,” *Journal of Personality*, 72, 271–322.
- TANNER, M. A., AND W. H. WONG (1987): “The Calculation of Posterior Distributions by Data Augmentation,” *Journal of the American Statistical Association*, 82(398), 528–540.
- TELLEGEN, A. (1985): “Structures of Mood and Personality and their Relevance to Assessing Anxiety, with an Emphasis on Self-Report,” in *Anxiety and the Anxiety Disorders*, ed. by H. Tuma, and J. Maser, pp. 681–706. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- THOMAS, A., AND S. CHESS (1977): *Temperament and Development*. Brunner/Mazel, New York.
- THOMAS, A., S. CHESS, AND H. G. BIRCH (1968): *Temperament and Behavior Disorders in Children*. New York University Press, New York.
- TODD, P., AND K. WOLPIN (2007): “The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps,” *Journal of Human Capital*, 1(1), 91–136.
- TURKHEIMER, E., A. HALEY, M. WALDRON, B. D’ONOFRIO, AND I. GOTTESMAN (2003): “Socioeconomic Status Modifies Heritability of IQ in Young Children,” *Psychological Science*, 14(6), 623–628.
- URZUA, S. (2008): “Racial Labor Market Gaps: The Role of Abilities and Schooling Choices,” *Journal of Human Resources*, 43(4), 919–971.
- VAN DYK, D. A., AND X.-L. MENG (2001): “The Art of Data Augmentation,” *Journal of Computational and Graphical Statistics*, 10(1), 1–50.
- WEBB, E. (1915): “Character and Intelligence,” *British Journal of Psychology*, 1(3), 1–99.
- WEISS, C. T. (2010): “The Effects of Cognitive and Noncognitive Abilities on Earnings: Different School Systems,” Discussion paper, European University Institute, mimeo.

WOLFE, R., AND S. JOHNSON (1995): "Personality as a Predictor of College Performance," *Educational and Psychological Measurement*, 55(2), 177–185.

WOOD, D. (2007): "Using PRISM to Compare the Explanatory Value of General and Role-Contextualized Trait Ratings," *Journal of Personality*, 76(6), 1103–1126.

**Table 1:** Personality Models and Sub-Factors

Inventory	Factors	Lower-order Factors
Big Five (Costa and McCrae, 1992) <sup>a</sup>	Openness to Experience	Fantasy, Aesthetics, Feelings, Actions, Ideas, Values
	Conscientiousness	Competence, Order, Dutifulness, Achievement Striving, Self-Control/Self-Discipline, Deliberation
	Extraversion	Warmth, Gregariousness, Assertiveness, Activity, Excitement Seeking, Positive Emotions
	Agreeableness	Trust, Straightforwardness, Altruism, Compliance, Modesty, Tender-Mindedness
	Neuroticism	Anxiety, Vulnerability, Depression, Self-Consciousness, <i>Impulsiveness</i> , Hostility
MPQ <sup>b</sup> (Tellegen, 1985)	Negative Emotionality	Stress Reaction, Alienation, Aggression
	Constraint	Control, Traditionalism, <i>Harm Avoidance</i>
	Positive Emotionality	<i>Achievement</i> , Social Closeness, Well-Being
Big Three (Eysenck, 1991)	Neuroticism	Anxious, Depressed, Guilt-Feeling, Low Self-Esteem, Tease, Irrational, Shy, Moody, Emotional
	Psychoticism	Aggressive, Cold, Egocentric, Impersonal, Anti-Social, Unempathic, Tough-Minded, Impulsive
	Extraversion	Venturesome, Active, Sociable, Carefree, Lively, Assertive, Dominant
JPI <sup>c</sup> (Jackson, 1976)	Anxiety, Breadth of Interest, Complexity, Conformity, Energy Level, Innovation, Interpersonal Warmth, Organization, Responsibility, Risk Taking, Self-Esteem, Social Adroitness, Social Participation, Tolerance, Value Orthodoxy, Infrequency	
Big Nine (Hough, 1992)	Adjustment, Agreeableness, Rugged Individualism, Dependability, Locus of Control, Achievement, Affiliation, Potency, Intelligence	

*italic:* Affiliation of facet is still in debate (see Bouchard and Loehlin, 2001).

<sup>a</sup> see also Costa and McCrae (2008)

<sup>b</sup> **M**ultidimensional **P**ersonality **Q**uestionnaire.

<sup>c</sup> **J**ackson **P**ersonality **I**nventory.

*Source:* Bouchard and Loehlin (2001) and own inquiry.

**Table 2: Overview of empirical studies**

Technology Feature	Study/Program	Data	Sample Size	Duration	Research Question	Method	Results
Sensitive/Critical Period	Hopkins and Bracht (1975)	parts of 2 high school graduating classes ( $\approx 20,000$ enrollments) from district Boulder County (Colorado)	varies between $N = 236$ and $N = 1709$	assessment (unbalanced): $\approx 10$ yrs.	stability of group verbal and group nonverbal IQ scores throughout grade 1,2,4,7,9 and 11	correlational analysis	- stability becomes evident between grades 4 and 7, i.e., correlation of adjacent measurements exceeds $r = .7$ - nonverbal IQ is less stable
	Johnson and Newport (1989), Newport (1990)	native Chinese or Korean speaking students (University of Illinois) with English as a second language and immigration ages between 3 and 39	$N = 46$	Cross-section	immigration-age dependent differences in second language proficiency in terms of syntax and morphology	correlational analysis	- correlation between age of arrival and test performance: $r = -.77^{***}$ (controlling for various background variables) - for early arrivals (age 3-15, $N = 23$ ): $r = -.87^{***}$ ; for late arrivals (age 17-39, $N = 23$ ): no significant correlation
Self-Productivity/Complementarity	O'Connor, Rutter, Beckett, Keaveney, Kreppner, and the English and Romanian Adoptees Study Team (2000)	Romanian adoptees from deprived environments placed at ages 0 to 42 months (between 1990 and 1992) and early adoptees from within the UK placed between ages of 0 and 6 months as control group	treatment group (Romanians): $N = 163$ ; control group: $N = 52$	assessment at ages 4 and 6 (except for 48 Romanian orphans placed after age of 24 months who were only assessed at age 6)	effects of early deprived environments on cognitive competence (Global Cognitive Index, GCI) and possible remediation	correlational analysis; repeated ANOVA	- at age six (whole sample): correlation between duration of deprivation and GCI, $r = -.77^{***}$ - repeated ANOVA (data from age 4 and 6): group-factor, $F(2, 150) = 14.89^{***}$ ; age-factor, $F(1, 150) = 54.90^{***}$ ; no significant interaction term, i.e., gains over time are equal across groups and early deficits are maintained
	Campbell and Ramey (1994), Carolina Abecedarian	- children from low-income families randomly assigned to treatment and control group at infancy and again before entering kindergarten (TT, TC, CT, CC) - four cohorts from 1972-1977	$N = 111$ ( $T = 57$ ; $C = 53$ )	-treatment: infancy - age 8 for TT, infancy - age 5 for TC, age 5 - 8 for CT and no treatment for CC group - assessment: at least annually up to age 8 and additionally at age 12	effects of different timing of 1. preschool program, including full day care with additional involvement and advisory for parents 2. school age program, providing home school teachers ...on Wechsler Intelligence Scale for children (WISC, longitudinal) and Woodcock Test of Academic Achievement (WTAA, at age 12)	repeated ANOVA	- longitudinal data: T-group constantly shows a significant advantage in IQ up to age of 8 - age 12 data: $T/T > TC > CT > CC$ concerning overall score (WISC and WTAA), $F(6, 170) = 2.63^{**}$
	Johnson and Houston Parent-Child Development Center (PCDC)	children from low-income Mexican-American families randomly assigned to treatment and control group	$N=216$ (initially), $T = 97$ , $C = 119$ (follow-up data only partially available)	- treatment: 2 yrs. (starting 1970) - follow-up assessment cross-sectional	effect of home visits for parent support in the first year (about age 1 to 2) and center-based programs for parents and children of aggregated 400 hrs. in year two (age 2 to 3), on grades, Iowa Test of Basic Skills (ITBS) and Classroom Behavioral Inventory (CBI) at ages 8-11	ANOVA	- at time of program completion, program children had superior IQ scores (Andrews et al., 1982) - at ages 8-11 $T = C$ for grades - ITBS: $T > C$ for vocabulary score ( $F(1, 107) = 7.40^{***}$ ), reading ( $F(1, 107) = 6.72^{**}$ ) and language ( $F(1, 107) = 5.70^{**}$ ) - CBI: $T < C$ for hostility ( $F(1, 134) = 7.62^{***}$ )

continued on next page

Technology Feature	Study/ Program	Data	Sample Size	Duration	Research Question	Method	Results
Self-Productivity/Complementarity	Schweinhart, Barnes, and Weikart (1993), High/Scope Perry Preschool Project	black children from adverse socioeconomic backgrounds from Ypsilanti (MI) randomly assigned to T and C group	$N = 123$ ( $T = 58$ , $C = 65$ )	- treatment: 2 yrs. (in five waves from 1962-1965) - assessments: scattered between age 3 to 27	- effect of daily 2 1/2 hr. classroom and weekly 1 1/2 hr. home visit on Stanford-Binet Intelligence Scale (SBI) at ages 3-9, Wechsler Intelligence Scale for Children (WISC) at age 14, California Achievement Test (CAT) and grades at ages 7-11 and 14, among various others	group-mean comparison	- SBI: $\mu_T = 91.3 > \mu_C = 86.3^{**}$ (at age 6) fades out to $\mu_T = 85$ and $\mu_C = 84.6$ (age 10) - no differences for WISC, $\mu_T = 81$ , $\mu_C = 80.7$ at age 14 - CAT score significantly higher at age 14, $\mu_T = 122.2$ , $\mu_C = 94.5^{***}$
	Fuerst and Schweinhart, Chicago Child Parent Center Program (CPC)	children from impoverished neighborhoods in Chicago assigned on a first-come-first-served basis to 6 CPCs affiliated to public schools (1965-1977)	- treatment: $N_T = 683$ - 2 controls: $C_1 = 372$ , $C_2 = 304$	- program entry at 3 to 4 yrs. of age - exit at age 8/3rd grade (in one CPC up to 6th grade) - last assessment at 8th grade	effects of daily 6 hrs. center care with special curriculums, parents-training and support from social workers, nurses and nutritionists on California Achievement Test (CAT), Test of Basic Skills (ITBS, afterwards)	group-mean comparison	- during treatment (up to third grade) $T > C$ ; at 8th grade: $T = C$ - graduation-rate: $T = 62\%$ , $C = 49\%$
	Hill, Brooks-Gunn, and Waldfogel (2002), Infant Health and Development Project (IHDP)	- low-birth-weight (lbw, < 2500g) premature infants from 8 US sites randomly assigned to treatment (1985)	$N = 1,082$ ( $T = 416$ , $C = 666$ )	- treatment: 3 yrs., from 4 weeks of age on - assessment: at age 3, 5 and 8 yrs.	- effects of weekly home visits by trained staff in the 1st year (biweekly afterwards) and daily (weekdays) center-based care (2nd and 3rd yr.) on Peabody Picture Vocabulary Test - Revised (PPVT-R) for achievement at age 3, 5 and 8, Stanford-Binet Intelligence Scale (SBI) at age 3 and Wechsler Preschool Primary Scale of Intelligence - Revised (WPPSI-R) at age 5, Wechsler Intelligence Scale for Children (WISC) at age 8 and Woodcock-Johnson-Psychological Battery (WJ, broad math and reading) at age 8	propensity score matching on intensive treatment (>400 days) using Mahalanobis match- ing within calipers (MM) and matching with replacement (MWR); regression adjusted	-age 3 PPVT-R: $TE_{MM} = 14.7$ ( $ITT = 7.0$ , $\mu_C = 82.1$ ) <sup>b</sup> ; SBI: $TE_{MM} = 17.5$ ( $ITT = 8.9$ , $\mu_C = 84.2$ ) - age 5 PPVT-R: $TE_{MM} = 9.8$ ( $ITT = 1.8$ , $\mu_C = 79.3$ ); WPPSI-R: $TE_{MM} = 7.0$ ( $ITT = -.5$ , $\mu_C = 91.1$ ) - age 8 PPVT-R: $TE_{MM} = 9.8$ ( $ITT = 1.8$ , $\mu_C = 84.3$ ); WJ Broad Reading: $TE_{MM} = 7.4$ ( $ITT = 1.0$ , $\mu_C = 96.6$ ); WJ Broad Math $TE_{MM} = 11.1$ ( $ITT = -.1$ , $\mu_C = 95.1$ )
	Currie and Thomas (1995), Head Start	children from National Longitudinal Survey's Child-Mother File (NLS-CM) and mothers' data from the NLSY79 (only households with 2 or more at least 3 year old children)	$N \approx 5000$ children	- treatment: 2-4 yrs. - survey: 1986-1990 biennially	effects of Head Start participation of disadvantaged white and black children on Peabody Picture Vocabulary Test (PPVT) percentile-points and probability of no-grade-repetition compared to no-preschool siblings	- mother/household Fixed-Effects estimation	- White: 6-percentile increase in PPVT score ( $\beta = 5.875$ ); significant increase compared to other pre-schools ( $F(\text{Head Start} = \text{other preschool}) = 7.45^{***}$ ); 47% increase in probability of never repeating a grade - Afro-American: no significant effects for both outcomes - inclusion of Head Start x Age interaction shows that insignificant effect for blacks is due to rapid fade out of PPVT score gains <sup>c</sup>

continued on next page

Technology Feature	Study/ Program	Data	Sample Size	Duration	Research Question	Method	Results
Self-Productivity	Coneus and Pfeiffer (2007)	SOEP Mother-Child Questionnaire for children 0-18 months at date of assessment and Mother-Child Questionnaire 2 for 26-42 months old children	- age 0: $N = 730$ - age 3-18 months: $N = 580$ - age 26-42 months: $N = 192$	- birth cohorts 2002-2005 - birth cohort 2002 again in 2005 when they were 26-42 months old	influence of various skill-indicators of the previous period on indicators of current stock of skills when controlling for investment and further background variables	OLS, 2SLS, 3SLS	- partial elasticities of period 1 (ln birth-weight) on period 2 (3-18 months) skill indicators: e.g. on satisfaction (.28***), cry (.43***), console (.25**), health (.64***) - elasticities for period 3 skill indicators: e.g. $t = 2$ meta skill <sup>d</sup> on social skill (.63***), $t = 1$ in birth weight on $t = 3$ everyday skill (1.04***), activity in $t = 2$ and $t = 3$ (.81**), meta skill in $t = 2$ and $t = 3$ (.32*)
	Blomeyer, Coneus, Laucht, and Pfeiffer (2009)	first-born children from the Mannheim Study of Children at Risk (MARS) born between February 1986 and February 1988	$N = 384$	assessment waves at 3 months, 2 yrs., 4.5 yrs., 8 yrs. and 11 yrs. of age	-explanatory power of IQ and persistence measures of the previous assessment wave on current ones when controlling for nine different organic-psychosocial risk-combinations and indicators of investment	OLS	partial elasticities among IQ measures ( $t - 1$ on $t$ ): - .23** (at $t = 2$ yrs.), .53** ( $t = 4.5$ yrs.), .84** (at $t = 8$ yrs.) and .89** (at $t = 11$ yrs.) noncognitive skill (perseverance at $t - 1$ on $t$ ): - -.008 (at $t = 2$ yrs.), .18** ( $t = 4.5$ yrs.), .29** (at $t = 8$ yrs.) and .31** (at $t = 11$ yrs.)

\*\*\*  $p \leq .01$ , \*\*  $p \leq .05$  and \*\*\*  $p \leq .1$

<sup>a</sup> Only results for high-intensity treatment group with more than 400 days in center-care estimated with MM are displayed.

<sup>b</sup> Benchmark: Intention-to-Treat Effect (ITT) and control group mean ( $\mu^C$ ) for no-match.

<sup>c</sup> By age 10 African-Americans have lost any gains in terms of PPVT, whereas white children maintain an advantage of 5 percentile-points at that age.

<sup>d</sup> Meta skill is the arithmetic mean of the Likert-Scale transformations of health, satisfaction, cry, console and activity.

**Table 3: Overview of Studies Analyzing the Effects of Noncognitive Skills on Various Outcomes**

Study	Data	Sample Size	Time Horizon	Research Question	Method	Results
Coneus and Laucht (2008)	Mannheim Study of Children at Risk (MARS)	$N = 384$	noncognitive skills at ages 3 months and 2 yrs., outcomes between ages 8 and 19	effects of early temperamental measures (expert ratings) on grades and social outcomes	children Fixed-Effects	particularly low attention span (benchmark: high attention span) shrinks math grades ( $\beta = .55^{***}$ ), grades in German ( $\beta = .35^{***}$ ), number of delinquencies ( $\beta = 3.11^{***}$ ), probability of smoking ( $\beta = .15^{**}$ ), and number of alcoholic drinks per month ( $\beta = 22.77^{***}$ )
Duckworth and Seligman (2005)	two consecutive cohorts of US eight-grade high school students	$N_1 = 140$ , $N_2 = 164$	2002 and 2003	effect of self-discipline (controlled of standardized self-control and impulsiveness ratings) on grade point average	OLS	a one standard deviation increase in self-discipline increases GPA by $.1^{***}$ for the 2002 and by $.08^{**}$ for the 2003 cohort
Coneus, Germandt, and Saam (2009)	German Socio-Economic Panel (SOEP) youth questionnaire, waves 2000-2005	$N = 3,650$	noncognitive skills, academic skills and background variables were assessed from 2000 on, information on school tracks up to (including) 2005	effect of internal control at age 17 on later educational dropout (up to age 21)	Correlated Random Effects	a one standard deviation increase in internal control decreases dropout probability at age 17 by 2.5% and by 6% at age 19
Wolfe and Johnson (1995)	Students from State University of New York	$N = 201$	cross section	influence of various personality inventories and distinct scales on college grade point average (GPA)	OLS	particular influence of traits related to control, i.e., organization (from JPI), $\beta = .27^{***}$ , control (from BIG3) $\beta = .32^{***}$ and conscientiousness (from Big Five) $\beta = .31^{***}$
Carneiro, Crawford, and Goodman (2007)	National Child Development Survey 1958 (NCDS58)	initial cohort size $\approx 17,000$	background variables at 1958 and 1965, skill measures at 1965 and 1969, schooling outcomes at 1974 and labor market outcomes at 2000	effect of cognitive and noncognitive skills (social adjustment) at childhood on various outcomes	OLS Probit	standardized social adjustment score effects, e.g., probability of staying on at school until age 16 ( $.038^{***}$ ), employment status at 42 ( $.026^{***}$ ) and log hourly wages at 42 ( $.033^{***}$ )
Murmane, Willett, Braatz, and Duhalderde (2000)	National Longitudinal Survey of Youth 1979 (NLSY79) subsample for males	$N = 1,448$	measures assessed in 1980, wages (age 27/28) from 1990 to 1993	effect of self-esteem (controlling for cognitive speed, scholastic achievement, ethnic group and calendar year) of 15-18 year old males on log hourly wages at age 27/28	OLS	a one point increase in Rosenberg self-esteem scale increases log hourly wage by 3,7% ( $\beta = .037^{***}$ )
Heckman, Stixrud, and Urzua (2006)	National Longitudinal Survey of Youth (NLSY79)	$N = 6,111$	annual assessment beginning 1979 on background variables, test scores only in 1979	influence of cognitive and noncognitive skills (internal control and self-esteem) assessed in adolescence (ages 14-21) on various social and labor market outcomes at age 30, controlling for schooling and background	factor structure model and Bayesian Markov Chain Monte Carlo methods	- noncognitive skills even stronger predict log wages for most educational degrees, especially at the tails of the distribution - further influence on probability of unemployment, of being a white collar worker, of graduating from college, of smoking and marijuana use and of incarceration

continued on next page

Study	Data	Sample Size	Time Horizon	Research Question	Method	Results
Flossmann, Piatak, and Wichert (2007)	German Socio Economic Panel (SOEP) wave from 1999	$N_m = 1,549$ $N_f = 695$ males/ females	cross section	effect of internal control on wages	factor structure model and Bayesian Markov Chain Monte Carlo methods	particularly strong effect at the tails of the estimated control distribution <sup>b</sup>
Osborne-Groves (2005)	-National Longitudinal Survey of Young Women 1968 (NLSYW68) for the US -National Child Development Study 1958 (NCDS58) for the UK	$N_{US} = 380$ $N_{UK} = 1,123$	- NLSYW68: earnings in 1991/1993, background variables in 1968/1969 and self-control in 1970 and 1988 - NCDS58: aggression and withdrawal in 1965 and 1969, different controls 1969-1974 and log hourly wages in 1991	effect of the respective personality measures on women's later log hourly wages (accounting for simultaneity, measurement error and sample selection)	IV	- NLSYW68: a one standard deviation increase in internal control directly lower hourly wages by 6.7%*** (when controlling for schooling) <sup>a</sup> - NCDS58: a one SD increase in aggression and withdrawal lowers hourly wages by 7.6%*** and 3.3%***
Heineck and Anger (2010)	German Socio Economic Panel (SOEP) waves 1991-2006	$N = 1,580$ (ability measures available in 2005/2006, respectively, matched on previous waves)	background variables from all panel waves, personality measures from 2005 and IQ from 2006	effect of Big Five factors, internal control and positive/negative reciprocity on log hourly wages	- OLS (controlling for simultaneity, measurement error and sample selection) - Random Effects - Fixed Effects Vector Decomposition (FEVD)	- internal control is the strongest predictor across methods for women (-.080***) and men (-.067***) <sup>a</sup>

\*\*\*  $p \leq .01$ , \*\*  $p \leq .05$  and \*\*\*  $p \leq .1$

<sup>a</sup> The self-control scale points to the internal direction.

<sup>b</sup> The MCMC results are presented graphically.

**Table 4:** Overview of Studies Analyzing the Effects of Noncognitive Skills on Wage Gaps

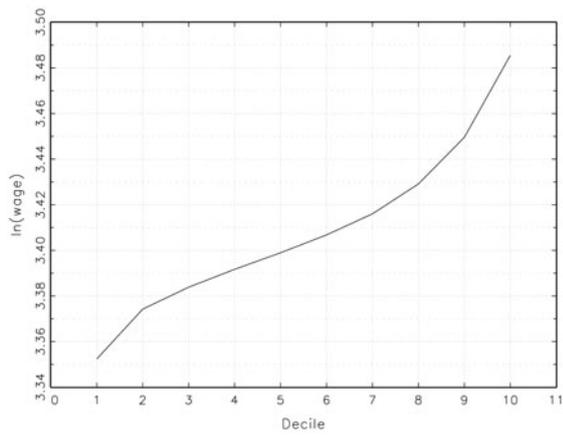
Study	Data	Sample Size	Time Horizon	Research Question	Method	Results
Fortin (2008)	- National Longitudinal Survey of High School Class of 1972 (NLS72) - 8th grade students from the National Educational Longitudinal Study (NELS88)	not reported (subsamples are used for comparability)	- NLS72: background and tests from 1973/74/76 and 1979, wages from 1979 - NELS88: background etc. surveyed in 1990/92/94 and 2000, wages from 2000	effect of internal control, self-esteem, importance of money/work and importance of people/family on log hourly wages at ages 24 (NELS88) and 25 (NLS72), controlling for cognitive skills, experience and other variables	OLS (Oaxaca-Ransom Type Decomposition)	NLS72: 6.4% of the .24 log wage gap due to noncognitive skills, particularly by importance of money/work NELS88: 5.3% of .18 gap, again mainly due to importance of money/work
Urzua (2008)	National Longitudinal Survey of Youth (NLSY79) representative sample and subsample for oversampling blacks	N = 3,423	annual assessment beginning 1979 on background variables, test scores only in 1979	relationship between cognitive (Armed Service Vocational Aptitude Battery subtests, ASVAB) and noncognitive (internal control, self-esteem) abilities, schooling choices, and black-white labor market differentials	factor structure model and Bayesian MCMC methods	noncognitive skills are strong predictors of schooling choices and wages for both groups, but explain little of the racial wage gap <sup>a</sup>
Braakmann (2009)	German Socio-Economic Panel (SOEP), 2005 wave for 25-55 years old full time employed respondents	N = 4,123	cross-section <sup>b</sup>	effects of Big Five, internal control and positive/negative reciprocity on log hourly wages	OLS (Blinder-Oaxaca Decomposition)	using males as reference, noncognitive skills explain up to 17.7% <sup>c</sup> of the .18 log wage gap, mainly due to women's higher degree of agreeableness and neuroticism
Mueller and Plug (2006)	Wisconsin Longitudinal Study (WLS), high school graduates in 1957	N = 5,025	background variables from 1964, 1975 and 1992; IQ from 1957; personality from 1992	effect of standardized Big Five measures on gender wage gap (controlling for background, IQ, education and tenure)	OLS (Oaxaca-Ransom Type Decomposition)	- differences in noncognitive skills explain 7.3% and differences in rewards 4.5% of the .58 log wage gap in favor of men (driven by women's higher degree of agreeableness and neuroticism) - the penalty for agreeableness and neuroticism is 2.9% higher in case of women

\*\*\*  $p \leq .01$ , \*\*  $p \leq .05$  and \*\*\*  $p \leq .1$

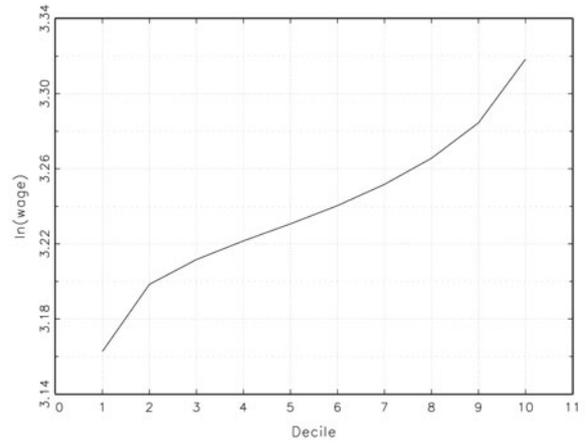
<sup>a</sup> The MCMC results are only presented graphically.

<sup>b</sup> A measure for risk-aversion was surveyed in the 2004 wave.

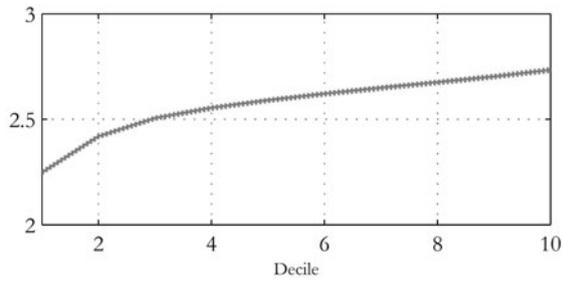
<sup>c</sup> Using the male coefficients as reference.



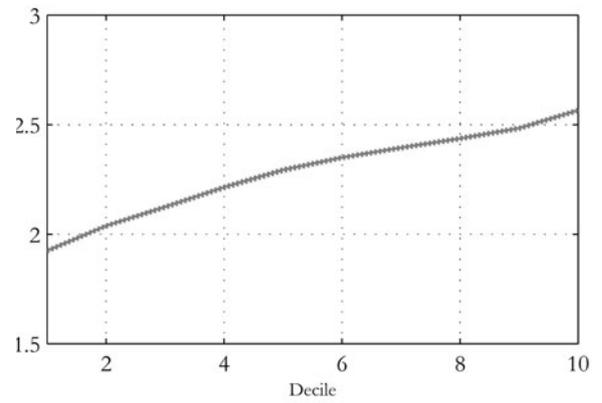
(a)



(b)



(c)



(d)

**Figure 1:** Net effect of noncognitive skills on log wages for 30-year old males and females in Germany and the United States. Upper panel: males (a) and females (b) in Germany. Lower panel: males (c) and females (d) in the United States. Sources: Flossmann, Piatek, and Wichert (2007) and Heckman, Stixrud, and Urzua (2006).