Economics Letters 236 (2024) 111598

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Shorter can be better: Balancing length and predictive power when measuring noncognitive skills to predict academic outcomes

Shuaizhang Feng [a], Yu Gan [a], Yujie Han [a,*], Tim Kautz [b]

[a] Institute for Economic and Social Research, Jinan University, No. 601, West Huangpu Avenue, Guangzhou 510632, China
[b] Mathematica, Inc., Princeton, NJ 08540, USA

ARTICLE INFO

JEL codes:
C83
I20

Keywords:
Abbreviated surveys
Big Five
Predictive power
Noncognitive skills
Personality traits

ABSTRACT

We develop shorter versions of a Big Five survey designed to measure students' noncognitive skills and predict students' later academic outcomes. We find that measures with fewer items can better predict students' outcomes, suggesting that using shorter versions of a Big Five Inventory may be cost-effective in large-scale social surveys.

1. Introduction

Recent evidence demonstrates that noncognitive skills—such as persistence and self-control—predict later life outcomes and can be shaped by interventions (Heckman et al., 2021). These findings have generated widespread interest in measuring noncognitive skills to track students' progress in schools (Feng et al., 2022; Kautz et al., 2021), evaluate the impacts of interventions (Heckman et al., 2013), assess school performance for accountability purposes (West et al., 2018), and investigate the development of human capital (Cunha et al., 2010). Noncognitive skills are typically measured using surveys in which respondents rate their skills (Duckworth and Yeager, 2015). However, such surveys are often long, making them costly and burdensome to administer, leading to efforts to develop shorter versions of the surveys. For example, several widely used surveys in economics—including the National Longitudinal Study of Youth 1997 and German Socio-Economic Panel—used abbreviated measures of noncognitive skills.[1]

Past efforts to shorten surveys of noncognitive skills have focused on selecting a subset of items that perform well based on the internal psychometric properties of the items or other subjective considerations rather than selecting items that best predict later outcomes (Gosling et al., 2003; Soto and John, 2017b). However, predictive power matters more than other psychometric properties for many practical applications. For example, many interventions are designed to improve noncognitive skills to boost future outcomes, such as academic performance or educational attainment (Kautz et al., 2014). If the noncognitive measures used to evaluate interventions are unrelated to such outcomes, they will not provide a way to assess whether the intervention works as intended. Predictive power is also essential when using noncognitive measures to identify whether students are at risk of poor future outcomes (Kautz et al., 2021). For reasons like these, past research has highlighted the need to develop measures of noncognitive skills that predict a variety of students' outcomes (McAbee and Oswald, 2013).

In this note, we use longitudinal data to develop shorter measures of noncognitive skills with a focus on predicting students' academic outcomes. We use a 60-item survey that measures a commonly used set of noncognitive skills called the Big Five, which some describe as the

Feng acknowledges support from National Natural Science Foundation of China Grants 72073052. Han acknowledges support from National Natural Science Foundation of China Grants 72303081 and the Fundamental Research Funds for the Central Universities of China 23JNQN25. This research has been approved by the Institutional Review Board of Social Sciences and Humanities at Jinan University (A2009/001-002).

* Corresponding author.
E-mail address: hanyujie2013@outlook.com (Y. Han).

[1] The National Longitudinal Study of Youth 1997 used a 10-item version of personality measure (Ten Item Personality Measure, TIPI) to measure noncognitive skills. The German Socio-Economic Panel used a 15-item version (The Big-Five Inventory Short, BFI-S) to measure noncognitive skills.

https://doi.org/10.1016/j.econlet.2024.111598
Received 3 January 2024; Received in revised form 4 February 2024; Accepted 7 February 2024
Available online 9 February 2024
0165-1765/

Table 1
Timing and grade of the survey administration for the Longitudinal Study of Children's Development (LSCD).

Cohort	Grade during each survey year 2017	2018	2019	2020	2021
Grade 4 in 2017	Grade 4	Grade 5	Grade 6		
Grade 5 in 2017	Grade 5	Grade 6			Grade 9
Grade 6 in 2017	Grade 6			Grade 9	Grade 10
Grade 4 in 2018		Grade 4	Grade 5	Grade 6	Grade 7

Notes: After the initial survey for each cohort, the sample includes students who progressed through the grades as expected.

"longitude and latitude" of personality (noncognitive) skills (Costa and McCrae, 1992). To develop shorter measures, we selected items from a student survey of the Big Five to maximize predictive power for students' academic outcomes measured one year later. We focus on how the predictive power relates to the number of items. This relationship is relevant because researchers and practitioners face limited time to administer surveys of noncognitive skills, so they frequently decide the number of survey items to use based on whether the benefits of including more items outweigh the costs.

We address the following questions: (1) To what extent do noncognitive measures based on student-self reports of varying lengths predict academic outcomes? (2) Which survey items lead to the most predictive measures? We find that the predictive power of noncognitive measures displays an inverted U-shaped pattern as a function of the number of items, suggesting that employing shorter measures of noncognitive skills may improve the predictive power of academic outcomes. The approach we adopt could be applied in other settings to select items when designing surveys.

2. Research design

2.1. Data

Our analysis uses the Longitudinal Study of Children's Development (LSCD), which was designed to track children's development in a predominantly rural county in China (Mianzhu) through school administrative data and surveys of students, their guardians, and teachers. We used measures of students' noncognitive skills based on student reports. Students completed the Big Five Inventory-2 (BFI-2) (Soto and John, 2017a). The BFI-2 includes 12 items to measure each dimension of the Big Five: openness to experience (openness), conscientiousness, extraversion, agreeableness, and emotional stability (or neuroticism).[2] The LSCD translated the original surveys into Mandarin. To ensure that the translated survey items performed well, the LSCD pretested the surveys with 469 students and made minor adjustments.

To assess the predictive power of the noncognitive measures, we focus on academic outcomes, drawing on administrative data on students' scores on semiannual Chinese and math exams.[3] In particular, we use the average Chinese and math scores measured one or two years after the noncognitive measures. In the Online Appendix, we present similar results for behavioral outcomes measured by teachers' reports on students' learning ability and mental health status.

We used five waves of the LSCD survey that span from 2017 to 2021 (Table 1). The baseline survey was conducted in 2017, covering around 6,000 students in grades four to six. In 2018, the LSCD followed the sampled fifth- and sixth-grade students and added another cohort of fourth-grade students. Follow-up surveys were conducted from 2019 to 2021 for selected cohorts of students. The total sample includes 12,941 student-year observations of students aged between ten and sixteen. The sample sizes for our analyses vary based on the availability of the survey and administrative data.

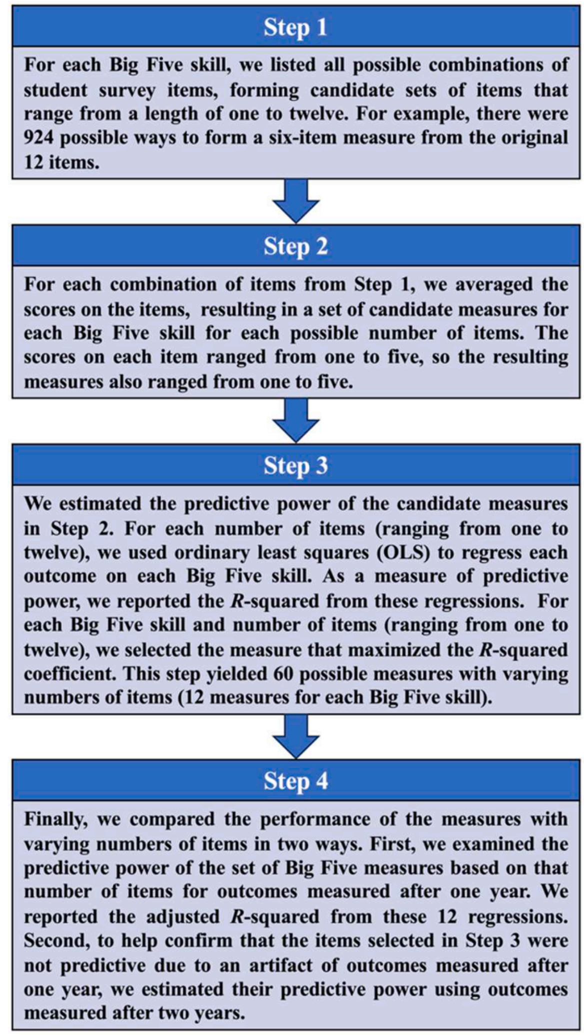

Fig. 1. The four steps of the analysis.

2.2. Methods

To develop predictive, student-reported measures of the Big Five, we selected items from the student survey to maximize the predictive power for academic outcomes measured one year later. Our analysis followed four steps (Fig. 1).

3. Results

When selecting Big Five items to best predict academic achievement after one year, the predictive power (*R*-squared coefficient) of the resulting measures displays an inverse U-shaped pattern as a function of the number of items (Fig. 2). As the number of items increases, the predictive power first increases and then decreases. Depending on the

[2] The Big Five are defined as follows: openness is the tendency to be curious and pursue intellectual interests; conscientiousness is the tendency to be hardworking and organized; extraversion is the tendency to be outgoing and sociable; agreeableness is the tendency to be unselfish and friendly; and emotional stability is the tendency to have consistency in emotional reactions.

[3] The academic outcomes are based on an average of the scores across the two exams. The behavioral outcomes were measured on five-point Likert scales.

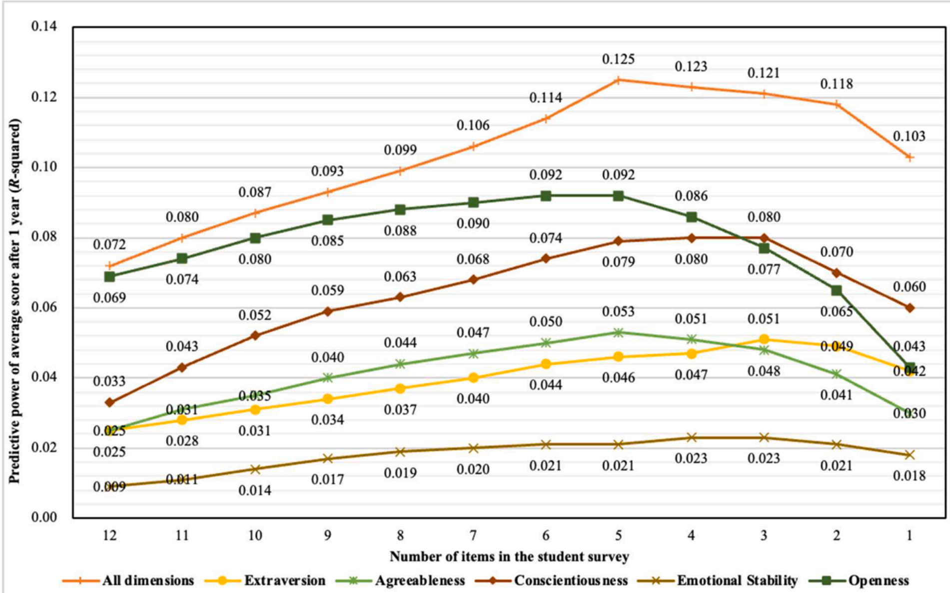

Fig. 2. Predictive power of student-reported measures of noncognitive skills with varying numbers of items for academic achievement measured one year later.
Notes: This graph displays the (adjusted) *R*-squared coefficients from an OLS regression of academic achievement on each dimension of student-reported noncognitive skills, as well as the five skills together, based on items from the Big Five Inventory-2 (Soto and John, 2017a). The outcome is an average score in Chinese and math collected through administrative records one year later. The number of observations is 12,941.

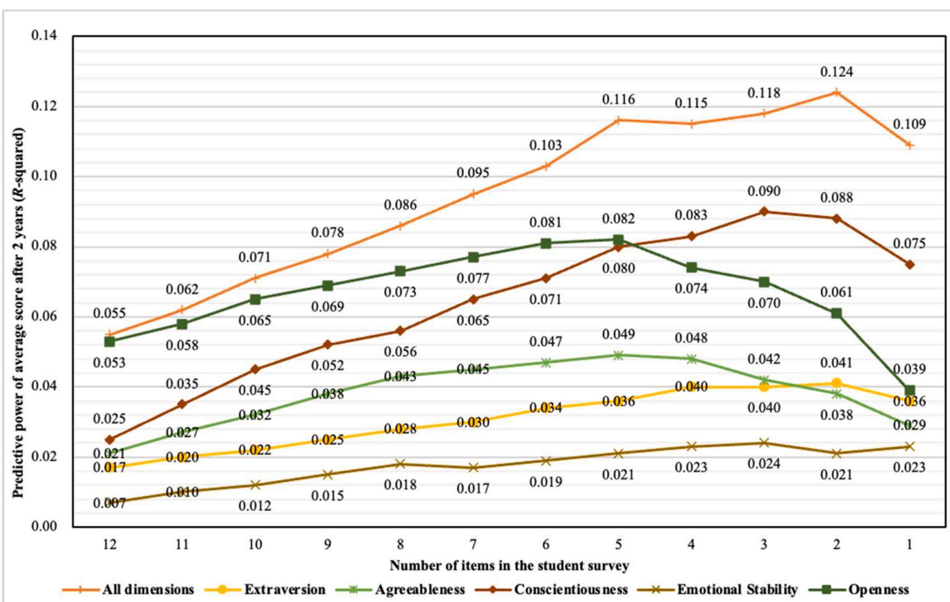

Fig. 3. Predictive power of the items selected in Fig. 2 for academic achievement measured two years later.
Notes: This graph displays the (adjusted) *R*-squared coefficients from an OLS regression of academic achievement on each dimension of student-reported noncognitive skills, as well as the five skills together, based on items selected in Fig. 2 and Table A1 in the Online Appendix. The outcome is an average score in Chinese and math collected through administrative records at least two years later. The number of observations is 12,941.

Big Five skill, the predictive power for future outcomes was maximized with three to five items from the student survey.[4] These findings suggest that, when selecting items to predict academic achievement, shorter surveys that use the most predictive items can outperform longer ones.

When we use the resulting student measures to predict academic achievement measured at least two years later, the predictive power displays a similar U-shaped pattern as a function of the number of items (Fig. 3). Depending on the dimension of the measures, the predictive power of the student-reported measures was maximized with between two and five items. These results help rule out that the selected items displayed in Fig. 2 were only predictive due to an artifact of the outcomes data collected after one year.

We conduct some sensitivity analyses and find similar results when using (1) students' average academic achievement measured at least two years later to select items (see Figs. A1 and A2 in the Online Appendix); (2) a cross-validation approach to select and test the student measures

[4] See Table A1 in the Online Appendix for a list of which items performed best for each possible number of items and Big Five skill.

with varying numbers of items (see Fig. A3 in the Online Appendix); and (3) other future outcomes (Chinese score, math score, learning ability, and mental health) to select items (see Figs. A4–A7 in the Online Appendix).

4. Conclusion

This is the first empirical study to develop and assess shorter measures of the Big Five noncognitive skills designed to predict later outcomes. We find that the predictive power of the resulting self-reported measures generally exhibited an inverted U-shaped pattern as a function of the number of items. Measures with between two to five items tended to perform well for the academic outcomes we examined.

The U-shaped pattern may have arisen through two competing factors. First, if each item measures the same skill, adding items generally reduces measurement error, increasing the estimated predictive power. Second, if the items for a given skill capture different facets of that skill with varying relationships to the outcomes, then adding items with weaker relationships to the outcome could reduce the predictive power. This second factor is consistent with previous evidence that shows that facets of individual Big Five skills can have opposing relationships with outcomes (Rustichini et al., 2016). These two factors might have counterbalanced one another. Adding more items may have first reduced measurement error but eventually led to a less predictive measure because it included items with weaker relationships to the outcomes.

Our findings suggest that using abbreviated measures of noncognitive skills may improve the predictive power for academic outcomes, which is consistent with other research (Ziegler et al., 2014). Using shorter measures may also have practical advantages, including reducing costs and burden and increasing response rates (Edwards et al., 2002). Our supplementary analysis of behavioral outcomes suggests similar patterns. Our approach to selecting survey items could be applied much more broadly to many other contexts where researchers must balance the predictive power of survey-based measures with the length of surveys. Future research could explore how this approach works in such contexts.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.econlet.2024.111598.

References

Costa, P.T., McCrae, R.R., 1992. Four ways five factors are basic. Personal. Individ. Differ. 13 (6), 653–665.
Cunha, F., Heckman, J.J., Schennach, S.M., 2010. Estimating the technology of cognitive and noncognitive skill formation. Econometrica 78 (3), 883–931.
Duckworth, A.L., Yeager, D.S., 2015. Measurement matters: assessing personal qualities other than cognitive ability for educational purposes. Educ. Res. 44 (4), 237–251.
Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., Kwan, I., 2002. Increasing response rates to postal questionnaires: systematic review. Br. Med. J. 324 (7347), 1183–1186.
Feng, S., Han, Y., Heckman, J.J., Kautz, T., 2022. Comparing the reliability and predictive power of child, teacher, and guardian reports of noncognitive skills. Proc. Natl. Acad. Sci. 119 (6), e2113992119.
Gosling, S.D., Rentfrow, P.J., Swann Jr, W.B., 2003. A very brief measure of the Big-Five personality domains. J. Res. Personal. 37 (6), 504–528.
Heckman, J.J., García, J.L., Bennhoff, F., Ermini Leaf, D., 2021. The Dynastic Benefits of Early Childhood Education. University of Chicago, Becker Friedman Institute for Economics Working Paper, p. 77, 2021.
Heckman, J.J., Pinto, R., Savelyev, P., 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. Am. Econ. Rev. 103 (6), 2052–2086.
Kautz, T., Feeney, K., Chiang, H., Lauffer, S., Bartlett, M., Tilley, C., 2021. Using a Survey of Social and Emotional Learning and School Climate to Inform Decisionmaking (REL 2021–114). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.
Kautz, T., Heckman, J.J., Diris, R., Ter Weel, B., Borghans, L., 2014. Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. OECD Education Working Papers, p. 110, 2014.
McAbee, S.T., Oswald, F.L., 2013. The criterion-related validity of personality measures for predicting GPA: a meta-analytic validity competition. Psychol. Assess. 25 (2), 532–544.
Rustichini, A., DeYoung, C.G., Anderson, J.E., Burks, S.V., 2016. Toward the integration of personality theory and decision theory in explaining economic behavior: an experimental investigation. J. Behav. Exp. Econ. 64, 122–137.
Soto, C.J., John, O.P., 2017a. The next Big Five Inventory (BFI-2): developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. J. Personal. Soc. Psychol. 113 (1), 117–143.
Soto, C.J., John, O.P., 2017b. Short and extra-short forms of the Big Five Inventory –2: the BFI-2-S and BFI-2-XS. J. Res. Personal. 68, 69–81.
West, M.R., Buckley, K., Krachman, S.B., Bookman, N., 2018. Development and implementation of student social-emotional surveys in the CORE Districts. J. Appl. Dev. Psychol. 55, 119–129.
Ziegler, M., Poropat, A., Mell, J., 2014. Does the length of a questionnaire matter? Expected and unexpected answers from generalizability theory. J. Individ. Differ. 35 (4), 250–261.