

Hanushek, Eric A.; Kinne, Lavinia; Sancassani, Pietro; Wößmann, Ludger

**Conference Paper**

## Patience and Subnational Differences in Human Capital: Regional Analysis with Facebook Interests

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2024: Upcoming Labor Market Challenges

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Hanushek, Eric A.; Kinne, Lavinia; Sancassani, Pietro; Wößmann, Ludger (2024) : Patience and Subnational Differences in Human Capital: Regional Analysis with Facebook Interests, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2024: Upcoming Labor Market Challenges, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/302338>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Patience and Subnational Differences in Human Capital: Regional Analysis with Facebook Interests<sup>\*</sup>

Eric A. Hanushek, Lavinia Kinne, Pietro Sancassani, and Ludger Woessmann<sup>†</sup>

## Abstract

Decisions to invest in human capital depend on people's time preferences. We show that differences in patience are closely related to substantial subnational differences in educational achievement, leading to new perspectives on longstanding within-country disparities. We use social-media data – Facebook interests – to construct novel regional measures of patience within Italy and the United States. Patience is strongly positively associated with student achievement in both countries, accounting for two-thirds of the achievement variation across Italian regions and one-third across U.S. states. Results are confirmed in an identification strategy that uses variation across ancestry countries from which ancestors of the current population of U.S. states migrated. Results also hold for six other countries with more limited regional achievement data.

Keywords: patience, human capital, student achievement, regions, social media, Facebook

JEL classification: I21, Z10

January 7, 2024

---

<sup>\*</sup> We gratefully acknowledge comments from Elliott Ash, Davide Cantoni, David Figlio, Philipp Lergetporer, Ömer Özak, Solomon Polachek, Paola Sapienza, Uwe Sunde, and seminar participants at Cornell University, the University of Rochester, the CESifo big data workshop, the IZA education workshop, the Berlin-Munich CRC retreat in Ohlstadt, the IWAE workshop in Catanzaro, the ifo Center for the Economics of Education, and the CESifo Area Conference on the Economics of Education. This work was supported by the Smith Richardson Foundation. The contribution of Woessmann is also part of German Science Foundation project CRC TRR 190.

<sup>†</sup> Hanushek: Hoover Institution, Stanford University; CESifo, IZA, and NBER, [hanushek@stanford.edu](mailto:hanushek@stanford.edu). Kinne: DIW Berlin, [kinne@diw.de](mailto:kinne@diw.de). Sancassani: ifo Institute, University of Munich, [sancassani.pietro@gmail.com](mailto:sancassani.pietro@gmail.com). Woessmann: University of Munich, ifo Institute; Hoover Institution, Stanford University; CESifo, and IZA, [woessmann@ifo.de](mailto:woessmann@ifo.de).

# 1. Introduction

An important determinant of individual investments in skills, recognized in the earliest human capital theory of Becker (1964), is differences in discount rates. Can then differences in people's time preferences – patience – also explain the large and longstanding regional differences in student achievement that exist in many countries? If so, differences in preferences provide a new perspective on regional income differences, since, for example, skill differences account for a substantial share of income differences across U.S. states (Hanushek, Ruhose, and Woessmann (2017)).<sup>1</sup> Subnational investigation of the role of discount rates has been stymied by a lack of representative region-specific measures of time preferences. Here, we combine the massive data available from social media – specifically Facebook interests – with machine-learning algorithms to derive new regional measures of patience. The analysis suggests a significant role for patience in accounting for within-country differences in student achievement and, by implication, geographical variations in incomes and economic growth. The differences in patience that we identify, for example, mirror the persistent and policy-resistant North-South disparities in economic development that exist in both Italy and the United States.

Many countries have large and persistent differences in student achievement across regions. Differences in average math achievement between the top- and bottom-performing U.S. states on the eighth-grade National Assessment of Educational Progress (NAEP) are equivalent to over two years of learning – roughly two-thirds of the achievement differences between top- and bottom-performing countries in the OECD. A similar magnitude is found between the top- and

---

<sup>1</sup> In both regional and international analyses, the type of direct achievement measures used here haven proven much more precise measures of human capital in explaining variations in income and growth than educational attainment measures (e.g., (Hanushek and Woessmann (2015))).

bottom-performing regions in Italy on the Istituto Nazionale per la Valutazione del Sistema Dell'Istruzione (INVALSI) test.

The role of discount rates in determining individual investment decisions is just part of the full impact of time preferences. Patience, the relative valuation of present versus future payoffs, appears in many closely-related decisions. At the individual level, students weigh current gratification such as play time with friends against study time that may lead to deferred rewards. At the group level, communities trade off present against future costs and benefits when deciding how much to invest in school quality, how strongly to motivate children to learn, and whether to design institutions to incentivize learning. Testing any hypothesized contribution of patience in affecting regional differences in educational achievement, however, requires representative regional measures of preferences.<sup>2</sup>

A major contribution of this paper is demonstrating how social-media data can be used to derive subnational measures of preferences that can be validated by international evaluations. The fundamental idea is that social-media data contain important information that can characterize people's underlying preferences such as patience and that allow analyses at more granular geographic levels than otherwise possible. This in turn permits investigation of previously unexplained heterogeneity of skills and income differences.

For marketing purposes, Facebook has developed an algorithm to classify the "interests" of over two billion people based on their self-reported interests, clicks and "likes" on Facebook, software downloads, clicks on advertisements that Facebook places on other sites, and additional inferences from overall behavior and location. We scrape Facebook's marketing application

---

<sup>2</sup> Data from the Global Preference Survey (GPS), which we use below, do have regional identifiers, but the small country samples cannot support regionally-representative samples. For example, 17 U.S. states have fewer than 10 individual observations in the GPS and just two states have over 50 observations.

programming interface (API) to identify the 1,000 Facebook interests with the largest audiences worldwide and use these raw data for describing preference differences.

Our derivation of regional patience measures builds on recent international analysis of culture in Obradovich et al. (2022). We extract data on the prevalence of Facebook interests in each country and region and reduce their dimensionality by fitting principal component analyses (PCA). Employing machine-learning techniques, we train an international model to predict the scientifically-validated patience measure of the Global Preference Survey (GPS) of Falk et al. (2018).

We validate the Facebook-derived measure of patience through an international analysis that mimics existing investigations of preferences and cross-country achievement differences (Figlio et al. (2019); Hanushek et al. (2022)). The Facebook-derived measure performs as well as the original GPS measure in predicting student achievement in the Programme for International Student Assessment (PISA), both in the original sample of 48 GPS countries and in a sample expanded to 80 countries by out-of-sample prediction. Importantly, our Facebook-derived measure also performs well in explaining the achievement of migrants in international estimation that addresses identification of the impact of patience.

We use the parameters estimated from international Facebook interests to construct subnational patience measures across 20 regions for Italy and across 50 states for the United States based on observed regional Facebook interests. Our interest in the two countries derives from their longstanding subnational differences that are not well explained, their large numbers of distinct regions, and their national testing facilities that provide achievement information for several grades.<sup>3</sup> Both countries show substantial North-South variation in the Facebook-derived

---

<sup>3</sup> The large North-South variation in Italy has raised substantial interest in policy and research (e.g., Putnam (1993); Ichino and Maggi (2000); Guiso, Sapienza, and Zingales (2004); Bigoni et al. (2018)).

measure of patience that coincide with the longstanding geographical disparities. We employ the newly-derived regional measures of patience in analyses of regional student achievement in the two countries. We use INVALSI test data from over 200,000 students in four grades for Italian regions and state NAEP data for fourth and eighth grades for U.S. states.

The Facebook-derived measure of patience is strongly associated with student achievement across both Italian regions and U.S. states. In Italy, a one standard deviation (SD) increase in regional patience is related to a 1.2-1.5 SD increase in eighth-grade math achievement, only slightly smaller than the cross-country estimate. In the United States, the equivalent estimate is statistically significant albeit only about one quarter in magnitude.

Regional differences in patience account for over two-thirds of the test score variation across Italian regions and for over one-third across U.S. states. The reasons behind the smaller role in the United States are not entirely clear but may reflect that the substantial internal mobility of the U.S. population lessens the geographical concentrations of preferences and alters the intergenerational transmission of cultural traits.

We provide two analyses to explore causality of the descriptive baseline association. Our cross-country analysis indicates limited bias when we assign migrant students the preference parameters of their origin country. This analysis allows conditioning on fixed effects for residence countries to shield against unobserved features of students' residence countries and reverse causation from education to preferences. At a basic level, the within-country estimation is less prone to confounding from unobserved national traits such as languages, constitutions, and institutional factors that has hampered prior cross-country analyses. But our instrumental-variable approach goes further to address potential confounders. We exploit the fact that the ancestors of the current populations of the different U.S. states migrated from different countries,

giving rise to variation in patience that is not jointly determined with other current state characteristics. Using the weighted patience level of the ancestry countries as an instrument for our Facebook-derived measure of states' patience confirms a significant effect of patience on student achievement of the same order of magnitude as the baseline model.

Consistent with skill development as a cumulative process, the association between patience and student achievement is stronger the higher the grade level. In the Italian INVALSI tests, estimates grow steadily across the four testing occasions from second to tenth grade. Similarly, estimates for the U.S. NAEP grow from fourth to eighth grade. This accumulation of skill differences likely comes from the joint impacts of preferences on student behavior, on parental behavior, and on societal investments in educational institutions, but it is not possible to pinpoint the precise contribution of each.

All results account for regional variation in risk-taking, another preference entering intertemporal decisions, but the machine-learning model predicting risk-taking from Facebook interests does not perform very well at the regional level. The poor measurement of risk-taking suggests that the estimates of patience are lower bounds; patience and risk-taking are positively associated and prior work suggests a negative association of risk-taking with student achievement (Hanushek et al. (2022)).

Results are stable in robustness analyses that include using reading achievement, differentiating by gender or wave, adding trust as an additional cultural-trait control, and employing the regionally representative PISA participation of Italy in 2012. Moreover, results are consistent for six additional countries where regional achievement data cover fewer grades or regions. The positive association between regional student achievement and Facebook-derived patience holds in a pooled sample of 190 regions in eight countries. The association is separately

significant in all additional countries – Brazil, Canada, Germany, Kazakhstan, and Mexico – except Spain.

Our main contribution is showing how the well-established theoretical and empirical relationship between patience and human capital accumulation can explain important parts of subnational differences in skills. Time preferences are important for economic development (Galor and Özak (2016); Sunde et al. (2022)), and previous analyses have shown that they are an important determinant of individual educational outcomes (Sutter et al. (2013); Golsteyn, Grönqvist, and Lindahl (2014); Polachek, Das, and Thamma-Apiroam (2015); De Paola and Gioia (2017); Castillo, Jordan, and Petrie (2019); Galor, Özak, and Sarid (2020); Angerer et al. (2023)) and of international achievement differences (Figlio et al. (2019); Hanushek et al. (2022)). But these analyses of geographic differences stop at the country level, leaving persistent differences in skills and income at a more granular level unexplained.

Past studies consider proximate causes of regional skill differences such as family background, school spending, and institutional settings (e.g., Hanushek and Raymond (2005); Woessmann (2010); Dee and Jacob (2011)). Yet, most stop without providing convincing explanations of more fundamental causes of the substantial on-going geographical variations in outcomes (e.g., Hanushek (2016)).

Our derivation of the regional patience measures further supports the use of social-media data in analyzing culture and social networks (e.g., Obradovich et al. (2022); Chetty et al. (2022); Bailey et al. (2022)) and in analyzing culture and economic outcomes more broadly (e.g., Guiso, Sapienza, and Zingales (2006); Alesina and Giuliano (2015); Giavazzi, Petkov, and Schiantarelli (2019)). With our better quantitative measures of more fundamental preference differences, it



becomes possible to understand portions of educational and economic outcomes that were previously listed as unexplained heterogeneity.

The next section describes our method to derive regional measures of patience, which we use in the following section to study the subnational association of patience with student achievement.

## **2. Deriving Regional Patience Measures from Facebook Interests**

We introduce the Facebook interest data (Section 2.1), validate their suitability for predicting international differences in patience (Section 2.2), and describe the derivation of regional patience measures (Section 2.3).

### **2.1 Facebook Interests**

With 2.9 billion monthly active users, Facebook is the world’s largest social network.<sup>4</sup> Facebook’s core business consists of selling advertising space which provides 97.5 percent of its revenues.<sup>5</sup> Hence, Facebook’s business model depends on its ability to keep users engaged on the platform while advertisers promote their products and services. To this purpose, Facebook puts considerable effort into inferring users’ interests (Thorson et al. (2021)).

Facebook determines users’ interests using a variety of sources, both inside the Facebook platform and on external websites (Cabañas, Cuevas, and Cuevas (2018); Obradovich et al. (2022)). Inside the Facebook platform, these sources include personal information that users share on Facebook as well as users’ activity on Facebook, such as page likes, group membership,

---

<sup>4</sup> Source: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> (accessed 23 February 2023).

<sup>5</sup> Figures about Facebook’s users and revenues are reported by its parent company Meta drawing on third-quarter 2022 results ([https://s21.q4cdn.com/399680738/files/doc\\_financials/2022/q3/Meta-09.30.2022-Exhibit-99.1-FINAL.pdf](https://s21.q4cdn.com/399680738/files/doc_financials/2022/q3/Meta-09.30.2022-Exhibit-99.1-FINAL.pdf), accessed 2 January 2023) and the 2021 annual report (<https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/14039b47-2e2f-4054-9dc5-71bcc7cf01ce.pdf>, page 58, accessed 2 January 2023).

and content with which users engage. Outside the platform, Facebook tracks users’ visited websites, installed apps, and purchasing behavior.<sup>6</sup> Facebook uses these data to deliver content and recommendations based on users’ interests and to allow advertisers to target users whose interests are relevant for the products or services that they want to sell.<sup>7</sup>

The hundreds of thousands of interests classified by Facebook are organized in nine main categories: business/industry, entertainment, family/relationships, fitness/wellness, food/drink, hobbies/activities, shopping/fashion, sports/outdoors, and technology. Interests can be very broad, such as “Entertainment” or “Music”, or very narrow, such as “Caribbean Stud Poker”, a casino game. Figure 1 shows the 1,000 Facebook interests with the largest worldwide audience, where larger fonts correspond to larger audience sizes. Interests often relate to leisure activities such as sports and beauty, but also to broader categories such as education and politics.

Following Obradovich et al. (2022), we proceed in two steps to retrieve data on Facebook interests for countries and subnational units. First, we obtain a comprehensive list of Facebook interests by querying Facebook’s marketing API, the interface that allows advertisers to configure their advertisement campaigns. For any given text query, a tool within the API returns a collection of closely related Facebook interests with their estimated worldwide audience and a unique identifier, which makes them language-independent. We iteratively feed this function with all 25,322 terms of an English dictionary<sup>8</sup> and 2,000 randomly selected titles of Wikipedia

---

<sup>6</sup> Aguiar et al. (2022) estimate that Facebook can track 55 percent of off-platform websites visited by U.S. Facebook users, amounting to 41 percent of browsing time.

<sup>7</sup> In a U.S. survey, 59 percent of Facebook users say that their assigned Facebook interests reflect their real-life interests (<https://www.pewresearch.org/internet/2019/01/16/facebook-algorithms-and-personal-data/>, accessed 23 February 2023).

<sup>8</sup> We use a dictionary of popular English words available at <https://github.com/dolph/dictionary/blob/master/popular.txt> (accessed 3 January 2023).

articles, each of which can yield several Facebook interests. This procedure produces 41,513 unique interests from which we select 1,000 with the largest worldwide audience.<sup>9</sup>

Second, for each of the 1,000 interests, we again use Facebook’s marketing API to obtain the estimated audience size separately for each country in which Facebook has a presence, as well as for each Italian region and U.S. state. For each geographical entity, this process yields a vector of size 1,000 with the estimated audience for the 1,000 interests. Finally, we standardize the estimated audience across the 1,000 interests in each geographical entity to make it independent of the total number of Facebook users or assigned interests in the entity.<sup>10</sup>

## **2.2 Cross-country Validation of Using Facebook Interests to Measure Patience**

We assess the suitability of the Facebook interest data for measuring patience through cross-country validation. We first reduce the dimensionality of the country-level Facebook interests by a principal component analysis (PCA) fitted on the sample of all 216 countries/entities featured by Facebook. The first 10 principal components (PCs) capture 70 percent of the total cross-country variance contained in the Facebook interests, 20 PCs capture 80 percent, and 48 PCs capture 90 percent (Appendix Figure A1). This pattern suggests that many PCs are required to capture the full cross-country variance in Facebook interests (see Obradovich et al. (2022)).

Next, we train a machine-learning model to characterize the relationship between the country-level PCs of the Facebook interests and the independently constructed measure of patience contained in the Global Preference Survey (GPS). The GPS collected scientifically-validated survey measures of several preference parameters from representative samples in 76

---

<sup>9</sup> We use 1,000 interests to make the data collection manageable. During data collection between April 2022 and May 2023, the API allowed a maximum of 300 queries per hour. For example, the over 50,000 queries for the U.S. states take over seven days of uninterrupted queries.

<sup>10</sup> Alternatively, dividing the Facebook audience count by the population or Facebook users in each geographic entity (to express it as a share of individuals holding an interest) yields the same qualitative results (not shown).

countries (Falk et al. (2018)).<sup>11</sup> The measure of patience combines a qualitative survey item and a hypothetical choice scenario that were chosen based on their predictive capacity for incentivized choices in an ex-ante laboratory setting.<sup>12</sup> Our 74-country training sample (Appendix Table A1) excludes Iran and Russia for which Facebook data are not available. We use a 10-fold cross-validated least absolute shrinkage and selection operator (LASSO) model for the cross-country training. The performance of the model is quite satisfactory: Independent of whether 10, 20, or 50 PCs are used, the  $R^2$  of the in-sample prediction of patience by the reduced-dimensionality Facebook interests is quite stable between 0.65 and 0.70 (Appendix Figure A2).

We use the parameter estimates of the machine-learning model to predict patience for all 80 countries with both PISA and Facebook data (Appendix Table A1). Given the limited size of the sample used to train the machine-learning model, we rely on the parsimonious specification with 10 PCs for the out-of-sample predictions to avoid overfitting.<sup>13</sup>

The beauty of this approach for measuring patience is its ability to capture latent components of users' underlying preferences (as described by the 1,000 most widespread Facebook interests). While the latent factors cannot readily be identified from any specific interest, survey question, or reported activity, they can be validated against the international preference survey. The procedure does not lend itself well to identifying specific interests as main predictors of the GPS patience measure, because any single Facebook interest can load positively or negatively (and in various combinations with other interests) on the different PCs

---

<sup>11</sup> The GPS measure is standardized to have mean zero and standard deviation one across individuals in the GPS countries, so that estimates in our subsequent analyses can be interpreted in terms of standard deviations.

<sup>12</sup> Brañas-Garza et al. (2023) provide lab, field, and online evidence that incentivized and hypothetical elicitations of time preferences yield broadly the same results.

<sup>13</sup> Less parsimonious models tend to obtain better in-sample performance (although this is hardly the case for patience, see Appendix Figure A2) but can lead to worse out-of-sample performance especially with small samples.

that enter the patience prediction.<sup>14</sup> However, if we correlate our Facebook-derived patience measure with the individual Facebook interests, terms such as dogs, outdoor recreation, adventure, holiday, painting, wildlife, and garden consistently show up among the strongest positive correlations, whereas the strongest negative correlations include terms such as free software, WhatsApp, Facebook Messenger, and massively multiplayer online role-playing games. Still, not all interests with strong correlations are easily interpretable, which aligns with the idea of picking up latent factors that are not well measured by individual Facebook interests. The Facebook-derived measure of patience can obviously be associated with regional differences in all sorts of other measures such as money, intelligence, or motivation – but, by construction, only to the extent that these are associated with patience as measured in the GPS.

We perform the same training and prediction models for risk-taking, another intertemporal preference contained in the GPS that has been found to enter into international student achievement. The  $R^2$  of the in-sample prediction for risk-taking is lower than for patience (Appendix Figure A2), indicating that risk-taking is harder to predict from Facebook interests.<sup>15</sup>

To validate our Facebook-derived measures of patience and risk-taking, we estimate their relationship with student achievement across countries. This validation, following Hanushek et al. (2022), uses math achievement on the PISA test over all seven available waves 2000-2018 to estimate the following OLS model:

$$T_{ict} = \beta_1 Patience_c + \beta_2 Risk_c + \alpha_1 B_{ict} + \mu_t + \varepsilon_{ict} \quad (1)$$

---

<sup>14</sup> Obradovich et al. (2022) show that direct prediction of cultural measures from individual Facebook interests generally provides worse predictions than using PCs.

<sup>15</sup> Unfortunately, the performance of the training model is even worse for the other preferences included in the GPS. For 10 PCs, the  $R^2$  of the in-sample prediction is about 0.35 for trust and at most 0.2 for altruism, positive reciprocity, and negative reciprocity (see also Obradovich et al. (2022)).

where  $T$ , the standardized PISA test score of student  $i$  in country  $c$  in year  $t$ , is a function of the country-level measures of patience and risk-taking, a vector of control variables  $B$  (student gender, age, and migration status), and an error term  $\varepsilon_{ict}$ . Fixed effects for test waves  $\mu_t$  account for time trends and idiosyncrasies of individual tests. The coefficients of interest are  $\beta_1$  and  $\beta_2$  that characterize the relationship of patience and risk-taking with student achievement. Regressions are weighted by students' sampling probability, giving equal weight to each country. Standard errors are clustered at the country level.

The Facebook-derived measures of patience and risk-taking perform very well in the cross-country validation exercise. Patience has a strong and significant positive relationship with student achievement when using the original GPS measure, whereas risk-taking has a strong and significant negative relationship (column 1 of Table 1, Panel A).<sup>16</sup> Column 2 substitutes our Facebook-derived measures for the GPS measures, using the same sample of countries.<sup>17</sup> The results are very much in line with those obtained using the original GPS measures, corroborating the validity of the Facebook-derived measures. Point estimates are in fact slightly larger (in absolute terms) than the original estimates.<sup>18</sup>

Out-of-sample predictions allow us to extend the analysis of the Facebook-derived measures of patience and risk-taking from 48 to 80 countries – all countries that participated in PISA and have Facebook data – encompassing over 2.6 million student observations. Results generalize very well to the extended sample, with increased precision and without significantly different

---

<sup>16</sup> This model replicates the main estimates of Hanushek et al. (2022) after dropping Russia with estimates hardly changed (see column 3 of their Table 1).

<sup>17</sup> The measures are obtained with 10 PCs of Facebook interests. Results are very similar when using additional (20-50) PCs (Appendix Table A2).

<sup>18</sup> All differences in coefficient estimates between columns 1 and 2 are statistically insignificant except for the coefficient on patience in Panel A.

estimates (column 3). Even in the 32 countries that were not part of the original GPS analysis, results are qualitatively the same and statistically highly significant (column 4).

In the international analysis, we further validate the Facebook measures by analyzing performance of migrants in a way that accounts for unobserved differences across residence countries. The analysis restricts the PISA sample to students with a migrant background and assigns them the values of patience and risk-taking of their home countries (see Figlio et al. (2019); Hanushek et al. (2022)), based on the evidence that migrant children inherit some of the preferences of their country of origin through family linkages (e.g., Guiso, Sapienza, and Zingales (2006); Bisin and Verdier (2011); Alesina and Giuliano (2014)). By observing migrant students from different origin countries who are schooled in the same residence country, we can include fixed effects for residence countries (as well as their full interaction with wave fixed effects). This migrant analysis addresses the most significant threats to identification of the preference effects by excluding the possibility that the relationships are driven by other factors of the country of schooling or by reverse causation, strengthening the interpretation of the cross-sectional analysis.

The migrant analysis further validates the informational content of the Facebook-derived measures. Results in Panel B of Table 1 show that the positive patience relationship and the negative risk-taking relationship again replicate very well when using the Facebook-derived rather than the original GPS measures.<sup>19</sup> The risk-taking coefficient is somewhat less precisely estimated but actually increases in (absolute) size. Estimates become quite imprecise (and larger)

---

<sup>19</sup> With the Facebook data, we expand the countries of origin considered in the migrant analysis from 56 to 93 (see Appendix Table A3). The destination countries increase only from 46 to 50 because some PISA countries do not report students' and parents' country of birth required to determine migrants' country-of-origin preferences.

when restricting the sample to non-GPS countries (column 4), indicating limited power of the migrant analysis in the smaller sample.<sup>20</sup>

Overall, the validation exercise shows that the preference measures from the Facebook data follow very closely the patterns from scientifically-validated GPS survey measures. This supports using Facebook interests to understand the role of preferences for geographical entities lacking alternative representative measures.

### **2.3 Predicting Regional Patience from Reduced-Dimensionality Facebook Interests**

Our construction of regional measures of patience from Facebook interests extends the method developed by Obradovich et al. (2022). First, we reduce the dimensionality of Facebook interests using a PCA, but this time we fit the PCA across regions *within* a given country. Fitting the PCA at the subnational level ensures that the PCs capture variance in dimensions of Facebook interests that are relevant for the specific country. For both Italian regions and U.S. states, the first 4 PCs capture over 70 percent of the regional variance in Facebook interests (Appendix Figures A3 and A4). 90 percent of variance is captured by 10 PCs in Italy and 15 PCs in the United States.

Second, we return to the international Facebook interest data to estimate the relationships between Facebook interests and the GPS preference measures but using within-country PC loadings from the previous step. We train a 10-fold cross-validated LASSO model to learn the relationship between these within-country PCs and the GPS measure of patience across countries.<sup>21</sup>

---

<sup>20</sup> Results of the migrant analysis are again stable for patience when using 20-50 PCs (Appendix Table A4). The significantly negative estimate on risk-taking also shows with 20 PCs, but not beyond, in line with the observation that risk-taking seems to be harder to predict from the Facebook data.

<sup>21</sup> For the training model, we again use all 74 countries with GPS and Facebook data to maximize the sample size for the machine-learning algorithm. However, results are quite similar if we instead use only the 25 OECD countries in the training model: the correlation among the two versions is 0.981 for Italian regions and 0.974 for



The model performs relatively well at predicting the GPS patience measure for both country-specific PC loadings. A small number of PCs capture a considerable portion of the variation in the GPS patience measure: with 10 PCs, the  $R^2$  of the in-sample prediction reaches 0.5 for PCs fitted across Italian regions and over 0.6 for PCs fitted across U.S. states (Appendix Figures A5 and A6). Given the limited number of subnational regions, we focus on models with fewer PCs in our main regional analyses because of the potential problem of overfitting out-of-sample predictions.

Third, we derive regional measures of patience by using the parameter estimates from the internationally trained model to construct patience measures from the Facebook interests observed for Italian regions and U.S. states. Figure 2 shows maps of the regional variation of the Facebook-derived patience measure in Italy and the United States. In Italy, the regions with the lowest patience measure are Sicily and Campania in the South. The region with the highest level of patience is Trentino-Alto-Adige in the North-East. Interestingly, parts of Trentino-Alto-Adige belonged to Austria and the former Austro-Hungarian empire for long periods of time, and large parts of the population speak German as their first language. The high level of predicted patience in Trentino-Alto-Adige is consistent with the fact that Austria has a much higher level of patience than Italy according to the country-level GPS measures, thus adding qualitative support for the Facebook-derived measure.<sup>22</sup> In the United States, the states that exhibit the highest level of patience are Vermont and Maine in the North-East. Both countries tend to show a North-South gradient in the Facebook-derived measure of patience.

---

U.S. states, and qualitative results do not change for our main regression analysis. This indicates that our results do not hinge on any heterogeneity across countries with very different development levels.

<sup>22</sup> The Austrian GPS measure of patience (0.61) is half a standard deviation higher than for Italy (0.11). A similar argument can be made for the Aosta Valley region in the North-West whose culture is deeply intertwined with neighboring France. France's GPS measure of patience is a quarter of a standard deviation higher than Italy's.

The prediction model following the same approach for risk-taking performs substantially worse. Both for Italian regions and for U.S. states, the  $R^2$  of the in-sample prediction is well below 0.2 for all models with up to 10 PCs and well below 0.4 even for a model with 20 PCs.<sup>23</sup> Consistent with prior investment studies, we include the measure of risk-taking as a control variable in our regional analysis throughout.<sup>24</sup> However, its poor measurement at the regional level means that the estimates for patience are likely lower bounds because patience and risk-taking are positively associated and risk-taking is negatively associated with achievement in the cross-country analysis (Hanushek et al. (2022)).

Another way of validating our Facebook-derived patience measure is to compare it to the regionally-identified GPS data. The GPS data contain regional identifiers from the underlying survey that allow construction of non-representative regional GPS measures of patience (Sunde et al. (2022)). These are obviously very noisy due to the small regional GPS sample sizes, averaging 50 individuals per Italian region and 20 per U.S. state. Nonetheless, they are positively correlated with our measure (weighted by the number of GPS observations per region): 0.49 (significant at the 5 percent level) across Italian regions and 0.23 (significant at the 10 percent level) across U.S. states. The smaller U.S. correlation is consistent with the added noise from the small GPS state samples.

### **3. The Importance of Patience for Subnational Student Achievement**

We report results for Italian regions (Section 3.1) and U.S. states (Section 3.2), robustness analyses (Section 3.3), exploration into causality using ancestry variation across U.S. states (Section 3.4), and results for additional countries (Section 3.5).

---

<sup>23</sup> See Appendix Figures A5 and A6. The performance with 20 PCs is a spike that likely reflects overfitting.

<sup>24</sup> Appendix Figure A7 shows maps of the regional distributions of risk-taking in Italy and the United States, but these should be interpreted with care because of the poor performance of the prediction model.

We think of patience as a deep determinant of student achievement, leading us to employ very parsimonious specifications of achievement differences. Proximate inputs often included in education production functions such as parental education or school resources would be bad controls in this setting as they are endogenous to regions' patience.

Regionally representative data on student achievement come from INVALSI for Italy and from NAEP for the United States and refer to the last waves before the COVID-19 pandemic (see Appendix A for details). Our primary analysis focuses on math achievement in eighth grade, the oldest cohort available in both countries and closest in age to PISA. The Italian data are available at the student level (59,034 eighth graders, 235,661 in total), the U.S. data are available at the state level. We divide test scores by the student-level standard deviation in the respective country, so that regression coefficients can be interpreted in terms of standard deviations.

### **3.1 Performance of Italian Regions**

The longstanding North-South divide among Italian regions invites investigation of fundamental driving forces. Because the schooling system is regulated mostly at the country level, the clear test score variations across regions are unlikely to be driven by the institutional structure of schools.<sup>25</sup>

Regional differences in patience prove to be strongly and significantly associated with student achievement. Panel A of Table 2 shows results of student-level analyses of math achievement in eighth grade using patience measures obtained with 4, 7, and 10 PCs of Facebook interests. The coefficient estimates are highly significant and are not very affected by the number of PCs used to derive the patience measure. A one standard deviation (SD) increase

---

<sup>25</sup> The matters in which the state has exclusive legislation are listed in Article 117 of the Italian Constitution (<https://www.governo.it/it/costituzione-italiana/parte-seconda-ordinamento-della-repubblica/titolo-v-le-regioni-le-province-e-i>; accessed 30 January 2023).

in patience is associated with an increase in math test scores of 1.35-1.51 SD, which is close to the cross-country estimates reported in Table 1.

Regional differences in patience can account for at least two-thirds of the regional variation in student achievement in Italy. Using student test scores aggregated to the regional level in Panel B of Table 2, point estimates are very similar, albeit slightly smaller than in the student-level analysis. The  $R^2$  shows that the model accounts for 0.68-0.80 of the region-level variation,<sup>26</sup> indicating the strength of patience in accounting for the large differences in student achievement across Italian regions.<sup>27</sup>

The estimates are consistent with the cumulative nature of the skill accumulation process as affected by the steady influence of patience on the flow of learning. The association of patience with student achievement increases strongly with higher grade levels; Panels A and B of Table 3 show results for all four grade levels available in INVALSI.<sup>28</sup> Coefficient estimates at the student level increase continuously from an insignificant 0.29 SD in second grade to a highly significant 1.77 SD in tenth grade. Region-level estimates are again quite similar. These results suggest that educational investments are cumulative and that the role of patience keeps adding up across grades (see also Figlio et al. (2019)).

### 3.2 Performance of U.S. States

The United States provides a large regional sample of 50 states plus Washington, D.C. that feature large differences in student outcomes.<sup>29</sup> With data accessible only at the state level, Panel

---

<sup>26</sup> The  $R^2$  is virtually unchanged when wave fixed effects and risk-taking are excluded from the model.

<sup>27</sup> Appendix Figure A8 shows scatterplots of the associations of patience and achievement across Italian regions and U.S. states.

<sup>28</sup> The patience measure uses 4 PCs of Facebook interests; results are very similar for 7 or 10 PCs (not shown).

<sup>29</sup> Results are similar when excluding Washington, D.C. from the analysis (not shown).

C of Table 2 reports results of state-level regressions for math scores in eighth grade using Facebook-derived measures of patience obtained with 4, 7, and 10 PCs.

Patience is significantly associated with higher student achievement at the U.S. subnational level and accounts for slightly more than one-third of the variation in test scores across U.S. states. A one SD increase in the Facebook-derived measure of patience is associated with an increase of 0.17-0.29 SD in test scores across U.S. states.

While patience plays an important role in accounting for cross-state differences in student test scores in the United States, its role is weaker than in Italy. The point estimates are only about a quarter of those estimated for Italian regions. A possible explanation is that the population in the United States is substantially more mobile and mixed. In 2019, 42 percent of the U.S. population lives in a state different from their state of birth.<sup>30</sup> Because cultural traits such as patience are mostly transmitted across generations (e.g., Bisin and Verdier (2011); Alesina and Giuliano (2014); Giavazzi, Petkov, and Schiantarelli (2019)), such internal migration suggests that cultural differences across states are likely to lessen over time. An implication is that school quality may adjust to the changing preferences through a range of political decisions.

Consistent with the Italian evidence, the association of patience and student achievement is smaller in lower grades in the United States. While still statistically significant, the coefficient estimate in fourth grade is only about half the size as in eighth grade (Panel C of Table 3), corroborating that the role of patience accumulates.

---

<sup>30</sup> Own calculations based on the ACS 2019 table of state of residence by place of birth available at <https://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-of-residence-place-of-birth-acs.html> (accessed 25 February 2023).

### 3.3 Robustness Analysis

Results prove stable in a series of robustness analyses. Both in Italy and the United States, we find similar results for reading achievement, albeit with slightly smaller point estimates, and there are no significant gender differences. Results are also robust in the separate assessment waves available in each country and in models conditioning on trust as another cultural trait. From the individual-level data for Italy, we confirm that the estimates are larger for native students than for migrant students. The availability of individual-level data for Italy allows a more in-depth analysis than for the United States, where the analysis is constrained by the regional-level data.

**Italy.** We exploit the within-country nature and the richness of the INVALSI data to replicate our analysis using reading outcomes. Our main analysis focuses on math achievement, which is generally considered the most comparable subject across countries, while student reading outcomes are inherently language-specific and less suitable for cross-country analysis. But, within our separate countries, the significant positive association of patience with student achievement also holds for reading. Results show that a one SD increase in patience is associated with a 0.99-1.22 SD increase in student reading achievement in the Italian individual-level analysis (Appendix Table A3). At the regional level, a one SD increase in patience is associated with an increase of 0.71-0.91 SD in reading scores. The magnitude of the coefficients in reading is slightly smaller than in math but results clearly show in both subjects.

Results are also very robust across subsamples of waves and gender. Results do not depend on the year in which the assessment was conducted, suggesting that they are not driven by the

specific timing of the achievement observation (Appendix Table A4). Results also hold similarly for girls and boys, and the gender difference is not statistically significant.<sup>31</sup>

In line with a leading role of cultural traits as a deep determinant of student achievement, results are stronger for native students than for migrant students. A one SD increase in patience is associated with a 1.42-1.58 SD increase in achievement for native students, a 0.75-0.91 SD increase for students with a second-generation migrant background, and a 0.56-0.89 SD increase for students with a first-generation migrant background (Appendix Table A7). This pattern would be expected if it were indeed patience as a cultural trait that drives the achievement results, as the culture of the residence region is presumably less important for migrant students who have been less exposed to the regional culture.<sup>32</sup>

To confirm that the association captures patience and not other cultural traits, we also condition on additional preference measures. Risk-taking, the other preference with intertemporal bearing, is already included in the baseline model. Among the other four preference parameters contained in the GPS, only trust is reasonably well predicted by Facebook interests (see above). Results are robust to including trust as an additional control variable (Appendix Table A8).

Results also hold when excluding Trentino-Alto-Adige whose sample is not representative for the entire region and whose German-language population might limit comparability. In the INVALSI test of this region, only students in the autonomous municipalities of Bolzano and Trento are tested (see Appendix A.1). This sampling restricted to municipal areas may bias our estimates, not least because Trentino-Alto-Adige is the Italian region with the highest estimated

---

<sup>31</sup> Reported results are based on Facebook-derived measures obtained with 4 PCs, but results are qualitatively the same with 7 and 10 PCs (not shown).

<sup>32</sup> Hanushek et al. (2022) find a similar pattern in their analysis of international student achievement.

level of patience (see Section 2.3). Furthermore, we want to be sure that results are not driven by the Austrian history and the partially German-speaking population of the region. When omitting Trentino-Alto-Adige from the analysis, results are qualitatively the same and, if anything, slightly larger in magnitude (Appendix Table A9).

Results are also robust in an analysis of unobservable selection and coefficient stability proposed by Oster (2019). We compare our baseline model to a restricted model without control variables, following the standard procedure of setting  $R_{max} = 1.3\tilde{R}$ . Assuming an equal degree of selection between observables and unobservables,  $\delta = 1$ , the estimated bias-adjusted coefficient  $\beta^*$  for patience is between 1.487 and 1.705 (Appendix Table A10). In all cases, the bias-adjusted coefficient  $\beta^*$  is larger than our main estimates. The values  $\delta$  for which  $\beta = 0$  lie between -2.680 and -4.117. In all cases, these values are much larger (in absolute terms) than using the standard cutoff  $\delta = 1$ . These results imply that the selection on unobservables would need to be more than 2.6 times larger (and of opposite sign) than the selection on observables to push the coefficient of patience to 0.

Finally, we make use of the fact that Italy participated with a regionally representative sample in the international PISA test in 2012 to show that results hold equally well in this alternative achievement test. Intriguingly, the PISA results shown in Appendix Table A11 are very similar to the INVALSI results shown in Panel A of Table 2, indicating that a one SD increase in patience is associated with a 1.47-1.57 SD increase in the PISA math score.

**U.S. States.** The results for U.S. states also replicate well for reading outcomes. The results closely mirror the findings for Italy: the magnitude of the coefficient of patience is slightly smaller compared to the analysis of math achievement (Appendix Table A12). A one SD



increase in patience is associated with an increase of 0.14-0.23 SD in reading achievement.

Again, this analysis confirms that results do not depend on a particular subject.

Results also do not depend on the specific year in which student achievement is observed.

Results are qualitatively the same for all analyzed NAEP waves – 2015, 2017, and 2019

(Appendix Table A13). The magnitude of the patience coefficient tends to be smaller in the most recent wave, although not statistically significantly so.

The U.S. results are also similar across genders. Patience is significantly positively associated with student achievement of both boys and girls (Appendix Table A14). The coefficient estimates are somewhat larger for boys than for girls, but not significantly so.

Finally, the U.S. results are also robust to conditioning on trust (Appendix Table A8).

### **3.4 Exploration into Causality: Patience in Ancestry Countries of U.S. State Populations**

The analysis of patience and subnational differences in human capital is descriptive so far. Concerns about causal interpretation could emerge if patience of the regional population is affected by the region's human capital (reverse causation) or if regional patience is jointly determined with other regional factors that are important for educational achievement (omitted variables). For the cross-country analysis, our investigation of migrants in section 2.2 addresses the main concerns of endogeneity by exploiting variation in the home-country patience of migrant students living in the same destination country and shows that these confounding factors are not a significant concern in the cross-country estimation.

Compared to the cross-country analysis, the within-country analysis is less prone to bias that may arise from national factors such as languages, laws, and institutional settings to the extent that these do not vary across subnational regions. Furthermore, our robustness analysis showed that the subnational results are robust to conditioning on other cultural traits, reducing concerns

about fundamental omitted variables. Still, it is possible that the subnational association of patience and human capital captures aspects of reverse causation and joint determination with other regional characteristics.

One way to explore the empirical relevance of any remaining endogeneity bias is to exploit variation in the level of patience of the current population of different U.S. states that emanates from their different ancestries. The ancestors of current U.S. state populations migrated from different countries. Cultural traits such as patience tend to be slow-moving, being handed down from generation to generation within families (Bisin and Verdier (2011); Alesina and Giuliano (2014)). As a consequence, the time preferences of current populations have deep historical roots (Galor and Özak (2016); Becker, Enke, and Falk (2020)). The levels of patience of the current population in countries from which the ancestors migrated can potentially serve as partial proxies for the patience of each state's current population. Importantly, the level of patience in the ancestry countries of the current populations of U.S. states is neither affected by nor jointly determined with other current characteristics of U.S. states. We therefore employ an identification strategy that uses patience in the ancestry countries of U.S. states' populations as an instrumental variable (IV) for our Facebook-derived measure of patience in the U.S. states.

Data on the ancestry of the current population of each U.S. state are provided in the American Community Survey (ACS). We use the 2022 wave of the individual-level ACS data, retrieved from the Integrated Public Use Microdata Series (IPUMS, Ruggles et al. (2023)), which provides demographic information for a random sample of the U.S. population. Importantly for our purposes, the ACS asks respondents about their ancestry or ethnic origin. We use the first reported ancestry of each respondent as a measure of their historical ancestry.

Based on states' ancestry composition, we obtain the ancestry measure of patience for each U.S. state by assigning respondents the patience level of the respective ancestry country observed in the GPS. To obtain an ancestry-weighted patience measure, we calculate ancestry shares for each ancestry country by U.S. state and weight the ancestry patience levels accordingly.<sup>33</sup>

Results of the instrumental-variable model confirm a significant effect of patience on student achievement across U.S. states. The reduced-form estimate in the first column of Table 4 shows that states whose population migrated from ancestry countries with higher patience levels on average have significantly higher test scores. In the first stage, the ancestry-patience instrument, measured directly from the GPS data, strongly predicts patience as measured by our Facebook-derived measure, independent of the number of PCs of Facebook interests used to obtain the state patience measure (columns 2-4). This first-stage association provides further credence to our derivation of patience measures from Facebook interests.

The second-stage results indicate that the IV model provides statistically significant estimates that are of the same order of magnitude as the OLS estimates. For the different numbers of PCs, the IV point estimates range from 0.221 to 0.372 (columns 5-7), which is in the same ballpark as in our baseline model (0.172 to 0.293). That is, there is no indication that the baseline model is biased upwards by endogeneity from reverse causation or joint determination with other current state characteristics.

---

<sup>33</sup> For about a third of ACS respondents, the ancestry information is either missing or cannot be assigned to a specific country because it refers to an ethnic group (e.g., Kurdish) or a larger region (e.g., Eastern Europe). We assign these respondents the average U.S. patience level from the GPS and control for the share of missing ancestry information in the regressions. Results are very similar when coding missing ancestry information as missing, which is equivalent to assigning the average of the observed ancestries of the respective state (not shown).

In sum, the strong association of patience and student achievement across U.S. states is confirmed in an identification strategy that exploits the fact that the ancestors of the current populations in the different U.S. states migrated from different countries. While the IV model with the historical instrument cannot rule out all potential biases, its results do not suggest substantial bias in the basic cross-sectional model.

### **3.5 Regional Analysis in Additional Countries**

While we have focused on Italy and the U.S. as two countries with interesting regional variation and consistent test data at different grades for a substantial number of regions, we can assess the stability of our results in other contexts by extending the analysis to six additional countries with publicly available regional test data. We leverage regional indicators in the PISA data since 2012 for all countries with at least ten regions: Canada and Spain in 2012, 2015, and 2018, Brazil and Mexico in 2012, and Kazakhstan in 2018. The Institut zur Qualitätsentwicklung im Bildungswesen (IQB) provides regionally representative math achievement data for German ninth-grade students in 2012 and 2018. For each country, we implement the method described in Section 2.3 to obtain regional measures of patience from Facebook interests, consistently using only 3 PCs because of the small number of regions in some countries.<sup>34</sup>

The positive association between patience and student achievement in these additional countries indicates that the role of intertemporal preferences is not limited to specific contexts. In the pooled model of 190 regions in eight countries, the highly significant coefficient suggests that a one SD increase in patience is associated with a 0.27 SD increase in math scores (Table 5, column 1). Country-specific results are more tentative due to the limited regional information in

---

<sup>34</sup> Pooled results are similar using more PCs to derive the patience measure, but country-specific results are not stable at higher numbers of PCs in Brazil, Canada, and Germany (not shown).

several countries, but separate regressions show a positive regional association between patience and achievement that is statistically significant in each country except Spain (columns 2-9).<sup>35</sup> The magnitude of coefficient estimates varies considerably across countries, suggesting that the strength of the relationship might depend on country-specific features, but there are too few country observations to analyze these differences systematically.

## 4. Conclusions

Time preferences, while clearly important to individual investment decisions, have an even broader impact when looking at education decisions. Aggregate preferences, which are a component of cultural identities, also affect political perspectives and community decisions about educational institutions and, for example, the definition and importance of school quality.

This analysis investigates the importance of patience in determining historically significant but largely unexplained regional differences in student skills. These skill differences in turn have lasting consequences for incomes and for regional growth patterns.

We use the extensive compilations of social media information by Facebook to estimate preference differences for subnational regions within Italy and the United States. The measures of patience constructed from Facebook interests are validated by international comparisons where direct measures of time preferences are available.

Differences in patience across regions in Italy and across states in the U.S. provide a powerful explanation of human capital outcomes. This new perspective on student performance helps to explain why, for example, North-South differences in student outcomes in both countries have been very stable over time even in the face of national efforts to equalize performance. While causal identification of the preference-achievement relationship across

---

<sup>35</sup> Appendix Figure A9 shows scatterplots of the country-specific associations for the six added countries.

subnational regions is particularly challenging, results are confirmed in a migrant analysis that provides credible identification in the cross-country setting and in an instrumental-variable model that allows to exploit historical ancestry variation in the analysis of U.S. states.

Our findings imply that similar educational inputs can lead to substantially different outcomes due to differences in patience. When addressing within-country differences in student achievement, policymakers might look beyond such proximate factors as school spending or even family educational background to take possible differences in patience into account. Institutional features of schooling, such as reliance on parental choice or test-based accountability, appear less tied to aggregate preferences (Hanushek et al. (2022)). Thus, institutional reforms of school systems appear a viable policy mechanism for improvement that does not necessarily depend on changing preferences (Woessmann (2016)). Moreover, while cultural traits are considered hard to change (e.g., Guiso, Sapienza, and Zingales (2006); Bisin and Verdier (2011)), evidence shows that traits such as patience are malleable, especially at a young age, and can be improved through specific interventions (e.g., Bird (2001); Alan and Ertac (2018); Jung, Bharati, and Chin (2021)). Hence, policies aimed at increasing patience may also be an avenue for addressing educational investments and regional deficits in student outcomes.

## References

- Aguiar, Luis, Christian Peukert, Maximilian Schaefer, Hannes Ullrich (2022). Facebook Shadow Profiles. CESifo Working Paper 9571. Munich: CESifo.
- Alan, Sule, Seda Ertac (2018). Fostering Patience in the Classroom: Results from Randomized Educational Intervention. *Journal of Political Economy* 126 (5): 1865-1911.
- Alesina, Alberto, Paola Giuliano (2014). Family Ties. In *Handbook of Economic Growth, Vol. 2*, edited by Philippe Aghion, Steven N. Durlauf. Amsterdam: North Holland: 177-215.
- Alesina, Alberto, Paola Giuliano (2015). Culture and Institutions. *Journal of Economic Literature* 53 (4): 898-944.
- Angerer, Silvia, Jana Bolvashenkova, Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter (2023). Children's Patience and School-Track Choices Several Years Later: Linking Experimental and Field Data. *Journal of Public Economics* 220: 104837.
- Bailey, Michael, Drew M. Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, Johannes Stroebe (2022). The Social Integration of International Migrants: Evidence from the Networks of Syrians in Germany. NBER Working Paper 29925. Cambridge, MA: National Bureau of Economic Research.
- Becker, Anke, Benjamin Enke, Armin Falk (2020). Ancient Origins of the Global Variation in Economic Preferences. *AEA Papers and Proceedings* 110: 319-323.
- Becker, Gary S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York, NY: National Bureau of Economic Research.
- Bigoni, Maria, Stefania Bortolotti, Marco Casari, Diego Gambetta (2018). At the Root of the North-South Cooperation Gap in Italy: Preferences or Beliefs? *Economic Journal* 129 (619): 1139-1152.
- Bird, Edward J. (2001). Does the Welfare State Induce Risk-Taking? *Journal of Public Economics* 80 (3): 357-383.
- Bisin, Alberto, Thierry Verdier (2011). The Economics of Cultural Transmission and Socialization. In *Handbook of Social Economics*, edited by Jess Benhabib, Alberto Bisin, Matthew O. Jackson. Amsterdam: North-Holland: 339-416.
- Brañas-Garza, Pablo, Diego Jorrot, Antonio M. Espín, Angel Sánchez (2023). Paid and Hypothetical Time Preferences Are the Same: Lab, Field and Online Evidence. *Experimental Economics* 26 (2): 412-434.
- Cabañas, José González, Ángel Cuevas, Rubén Cuevas (2018). Facebook Use of Sensitive Data for Advertising in Europe. arXiv:1802.05030.
- Castillo, Marco, Jeffrey L. Jordan, Ragan Petrie (2019). Discount Rates of Children and High School Graduation. *Economic Journal* 129 (619): 1153-1181.
- Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebe, Nathaniel Hendren, et al. (2022). Social Capital I: Measurement and Associations with Economic Mobility. *Nature* 608 (7921): 108-121.

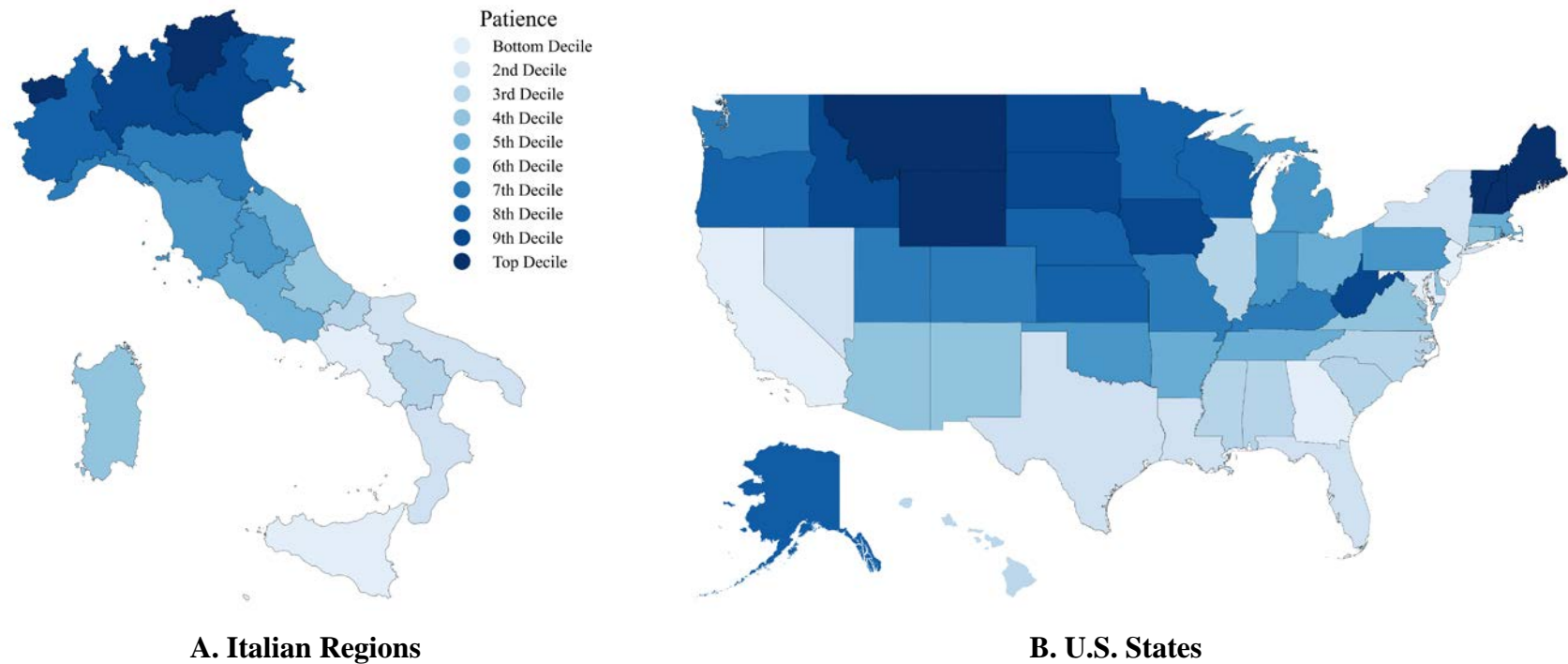
- De Paola, Maria, Francesca Gioia (2017). Impatience and Academic Performance. Less Effort and Less Ambitious Goals. *Journal of Policy Modeling* 39 (3): 443-460.
- Dee, Thomas S., Brian A. Jacob (2011). The Impact of No Child Left Behind on Student Achievement. *Journal of Policy Analysis and Management* 30 (3): 418-446.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, Uwe Sunde (2018). Global Evidence on Economic Preferences. *Quarterly Journal of Economics* 133 (4): 1645-1692.
- Falorsi, Piero Demetrio, Patrizia Falzetti, Roberto Ricci (2019). *Le Metodologie Di Campionamento E Scomposizione Della Devianza Nelle Rilevazioni Nazionali Dell'invalsi: Le Rilevazioni Degli Apprendimenti A.S. 2018-2019*. Milano: Franco Angeli.
- Figlio, David, Paola Giuliano, Umut Özek, Paola Sapienza (2019). Long-Term Orientation and Educational Performance. *American Economic Journal: Economic Policy* 11 (4): 272-309.
- Galor, Oded, Ömer Özkan (2016). The Agricultural Origins of Time Preference. *American Economic Review* 106 (10): 3064-3103.
- Galor, Oded, Ömer Özkan, Assaf Sarid (2020). Linguistic Traits and Human Capital Formation. *AEA Papers and Proceedings* 110: 309-313.
- Giavazzi, Francesco, Ivan Petkov, Fabio Schiantarelli (2019). Culture: Persistence and Evolution. *Journal of Economic Growth* 24 (2): 117-154.
- Golsteyn, Bart H.H., Hans Grönqvist, Lena Lindahl (2014). Adolescent Time Preferences Predict Lifetime Outcomes. *Economic Journal* 124 (580): F739-F761.
- Guiso, Luigi, Paola Sapienza, Luigi Zingales (2004). The Role of Social Capital in Financial Development. *American Economic Review* 94 (3): 526-556.
- Guiso, Luigi, Paola Sapienza, Luigi Zingales (2006). Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives* 20 (2): 23-48.
- Hanushek, Eric A. (2016). What Matters for Achievement: Updating Coleman on the Influence of Families and Schools. *Education Next* 16 (2): 22-30.
- Hanushek, Eric A., Lavinia Kinne, Philipp Lergetporer, Ludger Woessmann (2022). Patience, Risk-Taking, and Human Capital Investment across Countries. *Economic Journal* 132 (646): 2290-2307.
- Hanushek, Eric A., Margaret E. Raymond (2005). Does School Accountability Lead to Improved Student Performance? *Journal of Policy Analysis and Management* 24 (2): 297-327.
- Hanushek, Eric A., Jens Ruhose, Ludger Woessmann (2017). Knowledge Capital and Aggregate Income Differences: Development Accounting for U.S. States. *American Economic Journal: Macroeconomics* 9 (4): 184-224.
- Hanushek, Eric A., Ludger Woessmann (2015). *Universal Basic Skills: What Countries Stand to Gain*. Paris: Organisation for Economic Co-operation and Development.
- Ichino, Andrea, Giovanni Maggi (2000). Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm. *Quarterly Journal of Economics* 115 (3): 1057-1090.



- Jung, Dawoon, Tushar Bharati, Seungwoo Chin (2021). Does Education Affect Time Preference? Evidence from Indonesia. *Economic Development and Cultural Change* 69 (4): 1451-1499.
- Obradovich, Nick, Ömer Özak, Ignacio Martín, Ignacio Ortuño-Ortín, Edmond Awad, et al. (2022). Expanding the Measurement of Culture with a Sample of Two Billion Humans. *Journal of The Royal Society Interface* 19 (190): 20220085.
- Oster, Emily (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics* 37 (2): 187-204.
- Polachek, Solomon W., Tirthatanmoy Das, Rewat Thamma-Apiroam (2015). Micro- and Macroeconomic Implications of Heterogeneity in the Production of Human Capital. *Journal of Political Economy* 123 (6): 1410-1455.
- Putnam, Robert D. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
- Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, Megan Schouweiler (2023). IPUMS USA: Version 14.0 [Dataset]. <https://doi.org/10.18128/D010.V14.0>. Minneapolis, MN: IPUMS.
- Sunde, Uwe, Thomas Dohmen, Benjamin Enke, Armin Falk, David Huffman, Gerrit Meyerheim (2022). Patience and Comparative Development. *Review of Economic Studies* 89 (5): 2806-2840.
- Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rützler, Stefan T. Trautmann (2013). Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior. *American Economic Review* 103 (1): 510-531.
- Thorson, Kjerstin, Kelley Cotter, Mel Medeiros, Chankyung Pak (2021). Algorithmic Inference, Political Interest, and Exposure to News and Politics on Facebook. *Information, Communication & Society* 24 (2): 183-200.
- Woessmann, Ludger (2010). Institutional Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries. *Journal of Economics and Statistics* 230 (2): 234-270.



**Figure 2: Measure of Patience Derived from Facebook Interests for Italian Regions and U.S. States**



Notes: The figures show maps of the Facebook-derived measure of patience obtained with 4 PCs for Italian regions (Panel A) and U.S. states (Panel B), respectively. Each color corresponds to a decile of the distribution of patience within each country. Darker colors denote higher levels of patience.

**Table 1: Patience, Risk-taking, and Student Achievement: Cross-Country Validation Exercise**

	GPS measure	Facebook measure (10 PCs)		
	(1)	Original sample (2)	Extended sample (3)	Non-GPS sample (4)
<b>A. Cross-country analysis</b>				
Patience	1.225*** (0.132)	1.684*** (0.135)	1.722*** (0.119)	1.771*** (0.210)
Risk-taking	-1.229*** (0.188)	-1.359*** (0.310)	-1.537*** (0.254)	-1.660*** (0.388)
Control variables	Yes	Yes	Yes	Yes
Observations	1,954,840	1,954,840	2,660,408	705,568
Residence countries	48	48	80	32
R <sup>2</sup>	0.200	0.210	0.220	0.241
<b>B. Migrant analysis</b>				
Patience	0.957*** (0.115)	0.805*** (0.182)	0.902*** (0.205)	1.766*** (0.481)
Risk-taking	-0.315** (0.124)	-0.677** (0.278)	-1.221*** (0.350)	-3.531*** (0.549)
Control variables	Yes	Yes	Yes	Yes
Residence-country by wave fixed effects	Yes	Yes	Yes	Yes
Observations	78,403	78,403	90,983	12,580
Countries of origin	56	56	93	37
Residence countries	46	46	50	34
R <sup>2</sup>	0.280	0.272	0.298	0.310

Notes: Dependent variable: PISA math test score. Least squares regressions. Panel A: all PISA waves 2000-2018; weighted by students' sampling probability. Panel B: waves 2003-2018; students with both parents not born in the country where the student attends school; including 180 fixed effects for each residence-country by wave cell. Control variables: Panel A: student gender, age, and migration status; imputation dummies; and wave fixed effects; Panel B: student gender, age, dummy for OECD country of origin, imputation dummies. Robust standard errors adjusted for clustering at the country level (migrant analysis: country of origin) in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: PISA international student achievement test, 2000-2018; Falk et al. (2018); own elaboration of Facebook data.

**Table 2: Patience and Student Achievement: Subnational Analysis for Italy and the United States**

	4 PCs (1)	7 PCs (2)	10 PCs (3)
<b>A. Italy (individual level)</b>			
Patience	1.505*** (0.197)	1.351*** (0.114)	1.437*** (0.117)
Control variables	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes
Observations	59,034	59,034	59,034
Regions	20	20	20
$R^2$	0.092	0.099	0.099
<b>B. Italy (regional level)</b>			
Patience	1.245*** (0.193)	1.135*** (0.095)	1.207*** (0.099)
Wave fixed effects	Yes	Yes	Yes
Observations	42	42	42
Regions	20	20	20
$R^2$	0.679	0.790	0.795
<b>C. United States (state level)</b>			
Patience	0.293*** (0.089)	0.172* (0.096)	0.291** (0.132)
Wave fixed effects	Yes	Yes	Yes
Observations	153	153	153
Regions	51	51	51
$R^2$	0.360	0.348	0.360

Notes: Dependent variable: Panels A and B: INVALSI eighth-grade math test score in waves 2018 and 2019; Panel C: NAEP eighth-grade math test score in all NAEP waves 2015-2019. Least squares regressions with wave fixed effects. Unit of observation: Panel A: student; Panel B: region-wave combination; Panel C: state-wave combination. Patience measured at the regional/state level throughout. Col. 1-3 use the patience measure computed with 4, 7, and 10 principal components (PCs), respectively. Regressions control for risk-taking computed with the equivalent number of PCs. Additional control variables (Panel A): student gender, age, and migration status; imputation dummies. The Italian region of Trentino-Alto-Adige is represented by the two municipalities of Bolzano and Trento. Robust standard errors adjusted for clustering at the regional (state) level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: INVALSI mathematics achievement test, 2017-2019; NAEP mathematics achievement test, 2015-2019; own elaboration of Facebook data.

**Table 3: Patience and Student Achievement at Different Grade Levels**

	Grade 2 (1)	Grade 4/5 (2)	Grade 8 (3)	Grade 10 (4)
<b>A. Italy (individual level)</b>				
Patience	0.291 (0.193)	0.533* (0.286)	1.505*** (0.197)	1.766*** (0.236)
Control variables	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes
Observations	48,812	50,608	59,034	77,207
Regions	20	20	20	20
R <sup>2</sup>	0.028	0.032	0.092	0.151
<b>B. Italy (regional level)</b>				
Patience	0.182 (0.202)	0.364 (0.237)	1.245*** (0.193)	1.466*** (0.247)
Wave fixed effects	Yes	Yes	Yes	Yes
Observations	42	42	42	42
Regions	20	20	20	20
R <sup>2</sup>	0.044	0.075	0.679	0.678
<b>C. United States (state level)</b>				
Patience	—	0.156** (0.064)	0.293*** (0.089)	—
Wave fixed effects		Yes	Yes	
Observations		153	153	
Regions		51	51	
R <sup>2</sup>		0.158	0.360	

Notes: Dependent variable: Panels A and B: INVALSI math test score in waves 2018 and 2019; Panel C: NAEP math test score in all NAEP waves 2015-2019. Grade level indicated in column headers (col. 2 refers to fifth grade in Italy and fourth grade in the United States). Least squares regressions with wave fixed effects. Unit of observation: Panel A: student; Panel B: region-wave combination; Panel C: state-wave combination. Patience measured at the regional/state level throughout. Patience measure computed with 4 principal components (PCs). Regressions control for risk-taking computed with 4 PCs. Additional control variables (Panel A): student gender, age, and migration status; imputation dummies. The Italian region of Trentino-Alto-Adige is represented by the two municipalities of Bolzano and Trento. Robust standard errors adjusted for clustering at the regional (state) level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: INVALSI mathematics achievement test, 2017-2019; NAEP mathematics achievement test, 2015-2019; own elaboration of Facebook data.

**Table 4: Patience and Student Achievement across U.S. States: Using Ancestry Information as Instruments**

	Reduced form	First stage			Second stage		
		4 PCs	7 PCs	10 PCs	4 PCs	7 PCs	10 PCs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ancestry patience	0.945*** (0.349)	3.617*** (0.306)	3.605*** (0.283)	2.143*** (0.229)			
Patience					0.261*** (0.096)	0.221** (0.090)	0.372** (0.156)
Wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cragg-Donald Wald <i>F</i> statistic		291.218	443.142	228.963			
Observations	153	153	153	153	153	153	153
Regions	51	51	51	51	51	51	51
(Centered) <i>R</i> <sup>2</sup>	0.518				0.503	0.502	0.469

Notes: Dependent variable: col. 1 and 5-7: NAEP eighth-grade math test score in all NAEP waves 2015-2019; col. 2-4: patience. Unit of observation: state-wave combination. Patience measured at the state level throughout. Column headers indicate number of principal components (PCs) used to compute patience measure. Regressions control for risk-taking computed with the equivalent number of PCs and for the share of missing ancestry information. Robust standard errors adjusted for clustering at the state level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: NAEP mathematics achievement test, 2015-2019; ACS; own elaboration of Facebook data.

**Table 5: Patience and Student Achievement: Subnational Analysis in Eight Countries**

	Eight countries pooled	Brazil	Canada	Germany	Italy	Kazakhstan	Mexico	Spain	United States
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Patience (3 PCs)	0.266*** (0.066)	1.556*** (0.206)	0.383** (0.155)	2.093*** (0.624)	1.322*** (0.263)	0.458** (0.194)	0.667*** (0.161)	0.060 (0.108)	0.218* (0.113)
Grade/age	–	Age 15	Age 15	Grade 9	Grade 8	Age 15	Age 15	Age 15	Grade 8
Wave fixed effects	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes
Country fixed effects	Yes	No	No	No	No	No	No	No	No
Observations	383	27	30	32	42	16	32	51	153
Regions	190	27	10	16	20	16	32	17	51
$R^2$	0.282	0.719	0.726	0.577	0.648	0.177	0.517	0.362	0.300

Notes: Dependent variable: math test scores. Least squares regressions with wave fixed effects. Unit of observation: region-wave combination. Test and wave information: Brazil and Mexico: PISA 2012; Canada and Spain: PISA 2012, 2015, and 2018; Germany: IQB 2012 and 2018; Italy: INVALSI 2018 and 2019; Kazakhstan: PISA 2018; United States: NAEP 2015, 2017, and 2019. Regressions control for risk-taking computed with 3 PCs. Robust standard errors adjusted for clustering at the state level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: PISA, IQB, INVALSI, and NAEP mathematics achievement tests; own elaboration of Facebook data.



## **Appendix A: Data on Regional Student Achievement**

This appendix describes the regionally representative assessment data used for Italy and the United States: INVALSI (Appendix A.1) and NAEP (Appendix A.2), respectively.

### **A.1 Italy: INVALSI**

Since 2007, the Istituto Nazionale per la Valutazione del Sistema Dell'Istruzione (INVALSI) assesses a random sample of Italian students in math and Italian every year. Furthermore, INVALSI administers student, teacher, and principal questionnaires to collect background information about the educational environment. We use data on math achievement in the school years 2017-2018 and 2018-2019, the last years before the COVID-19 pandemic. In our main analysis, we focus on the sample of 59,034 eighth-grade students because they are closest in age to the students in PISA and NAEP, but we subsequently expand the analysis to include students in grades 2, 5, and 10, yielding a sample size of 235,661 students.

The sample of students is drawn following a two-step procedure, where a varying number of classes is randomly selected within a random sample of schools stratified at the regional level. Crucially for our analysis, the sample is representative at the regional level for 19 of the 20 regions in Italy (Falorsi, Falzetti, and Ricci (2019)). The exception is Trentino-Alto-Adige, where only students in the autonomous municipalities of Bolzano and Trento are tested. The difference between the lowest and highest performing region in Italy in eighth-grade math amounts to roughly three quarters of a standard deviation, equivalent to the average learning of almost three school years.

In robustness checks, we complement the INVALSI analysis using Italian data from PISA 2012 where Italy oversampled students in each region to obtain a representative sample of students.

## A.2 United States: NAEP

We use data from the National Assessment of Educational Progress (Main-NAEP), the largest nationally representative assessment of students in the United States. In our primary analysis, we focus on NAEP mathematics test scores in grade eight, using data from the last three waves of NAEP before the COVID-19 pandemic, namely NAEP 2015, 2017, and 2019. The resulting dataset consists of state-level test scores for the 50 U.S. states and the federal district of Washington, D.C. Approximately 140,000 students take part in a typical NAEP assessment.<sup>36</sup> In additional analyses, we also use data on fourth-grade students. In the United States, the difference between the lowest and highest performing state in eighth-grade math is equivalent to almost three years of schooling, similar to what was found in Italy.

We divide both INVALSI and NAEP test scores by the student-level standard deviation in the respective country, so that regression coefficients can be interpreted in terms of standard deviations.

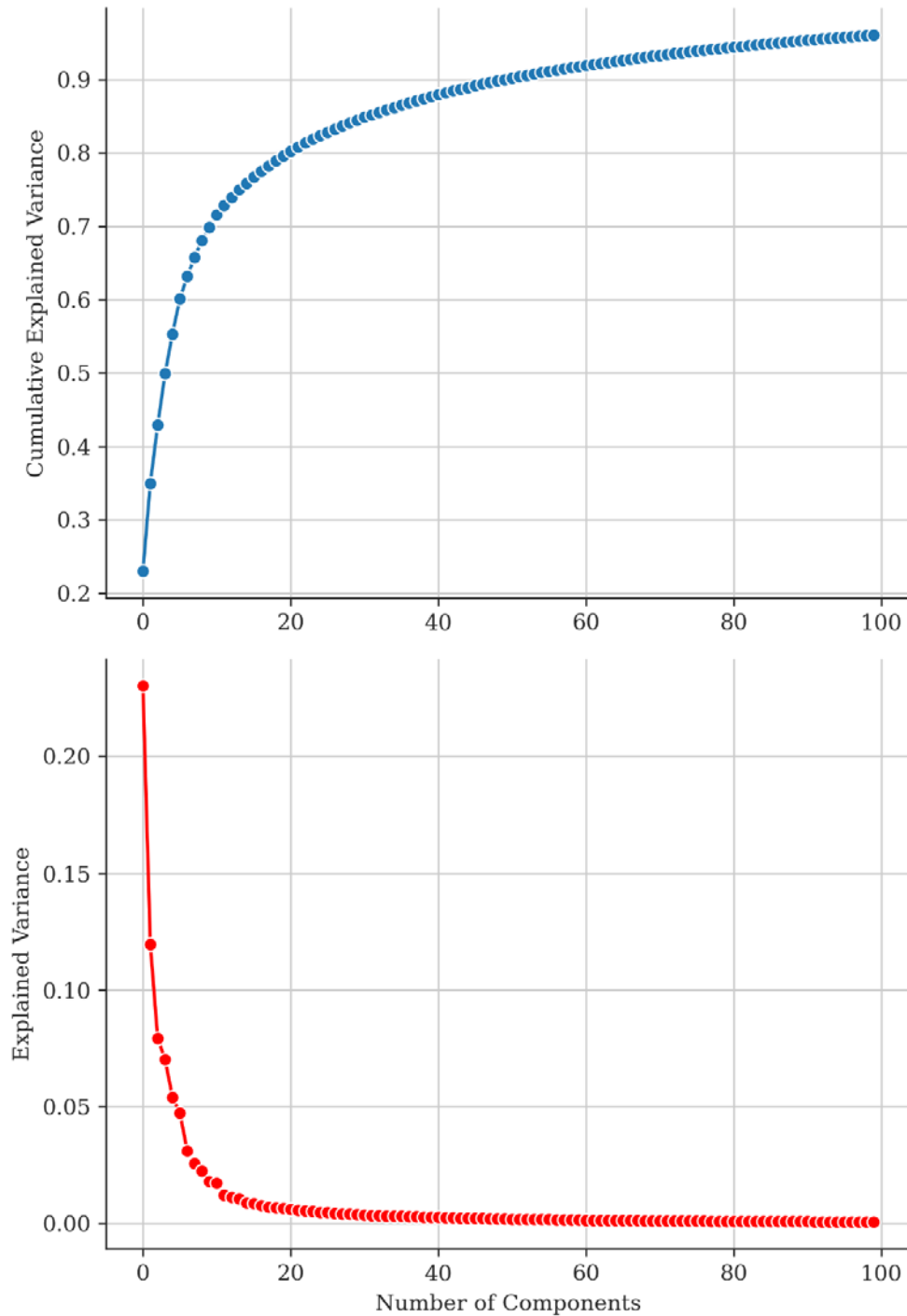
---

<sup>36</sup> Source:

<https://nces.ed.gov/nationsreportcard/guides/statsig.aspx#:~:text=A%20NAEP%20national%20assessment%20typically,samples%20of%20approximately%20140%2C000%20students> (accessed 23 February 2023).

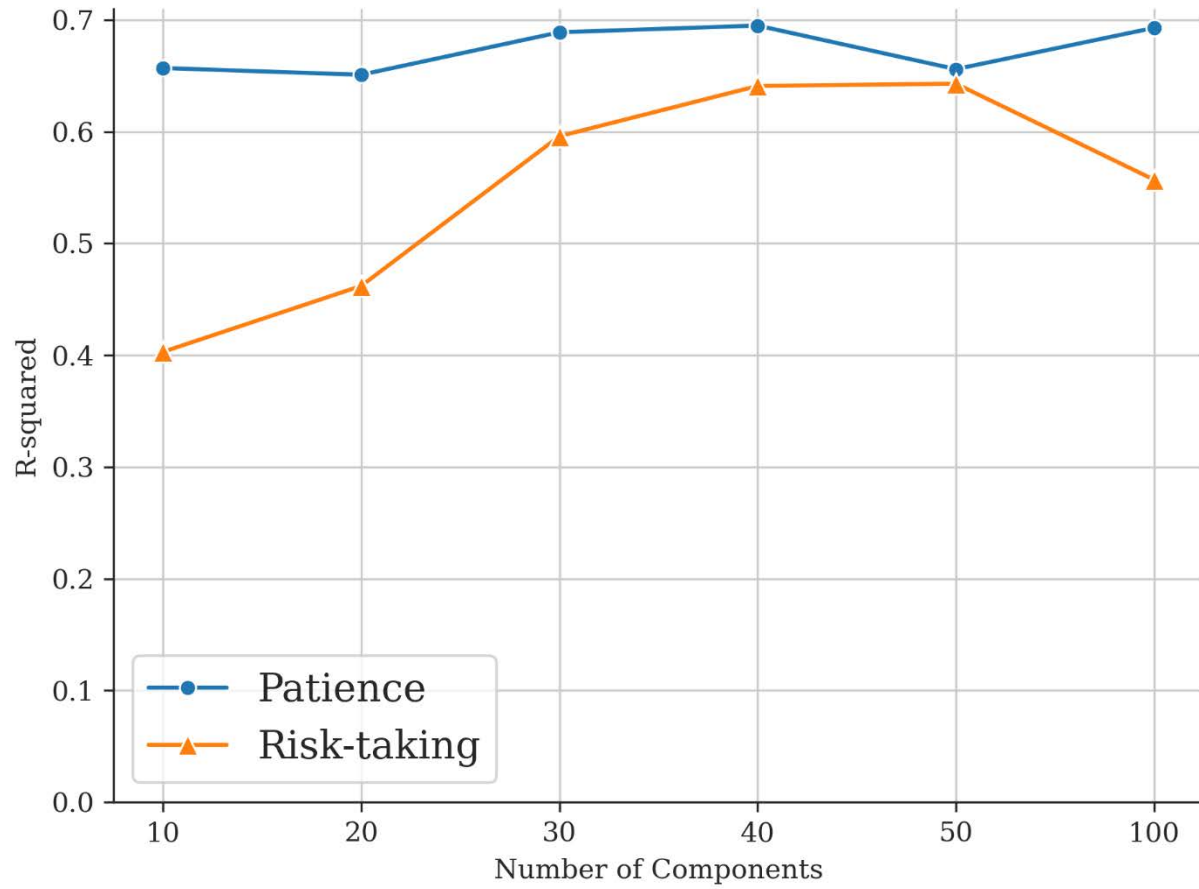
## Appendix Figures and Tables

**Figure A1: Variance in Facebook Interests Captured by PCs: International Sample**



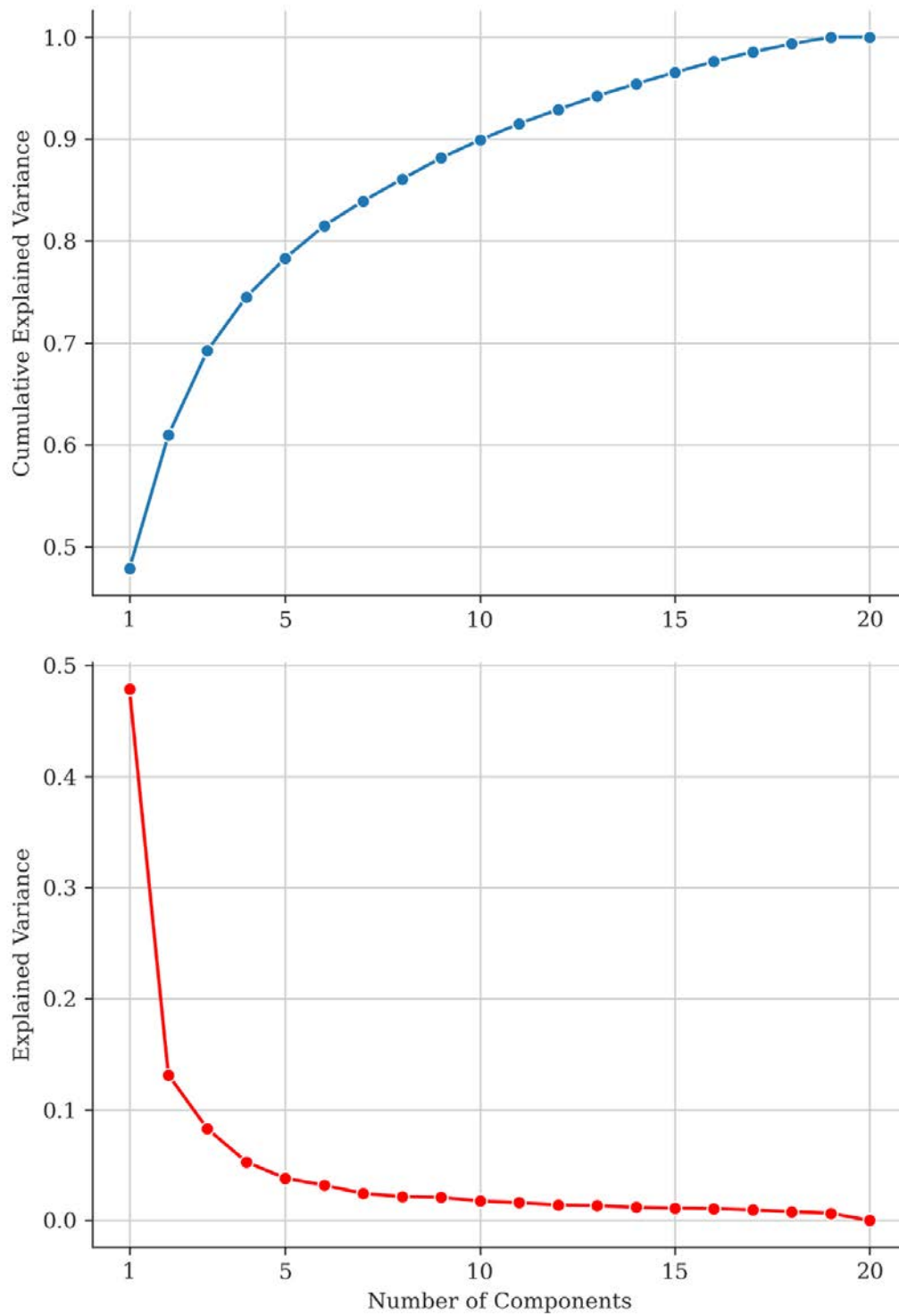
Notes: The top figure shows the cumulative variance in Facebook interests captured by the PCs of the Facebook interests in the international sample, the bottom figure shows the variance captured by each component.

**Figure A2: Performance of GPS Prediction with Facebook Interests: International Sample**



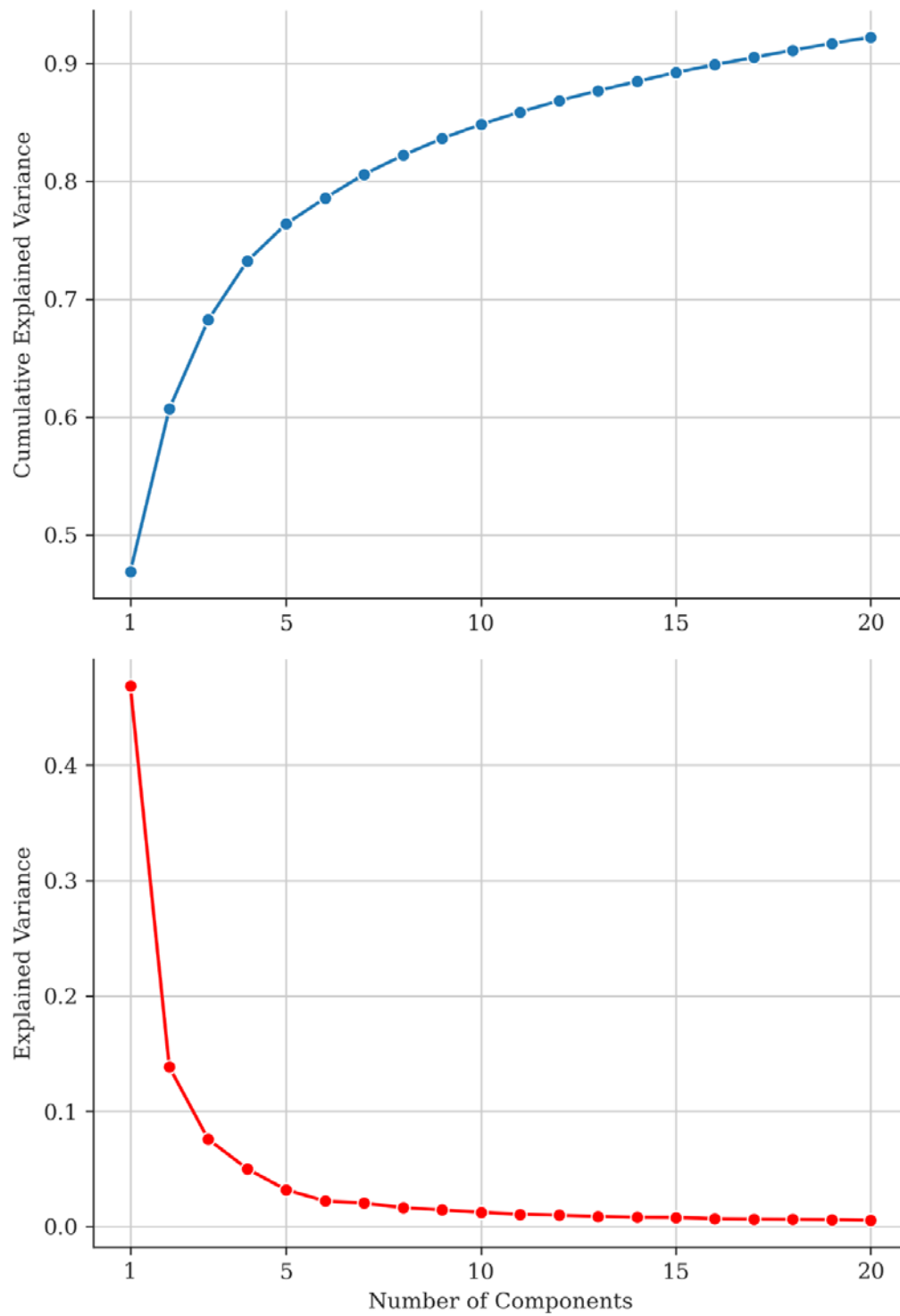
Notes: The figure shows the  $R^2$  of regressions of the GPS measures of patience and risk-taking, respectively, on the PCs of Facebook interests (obtained with PC loadings of country-level Facebook interests) for different numbers of PCs used in the regression. 10-fold cross-validated LASSO model. Sample: all 74 countries for which GPS and Facebook data are available.

**Figure A3: Variance in Facebook Interests Captured by PCs: Italian Regions**



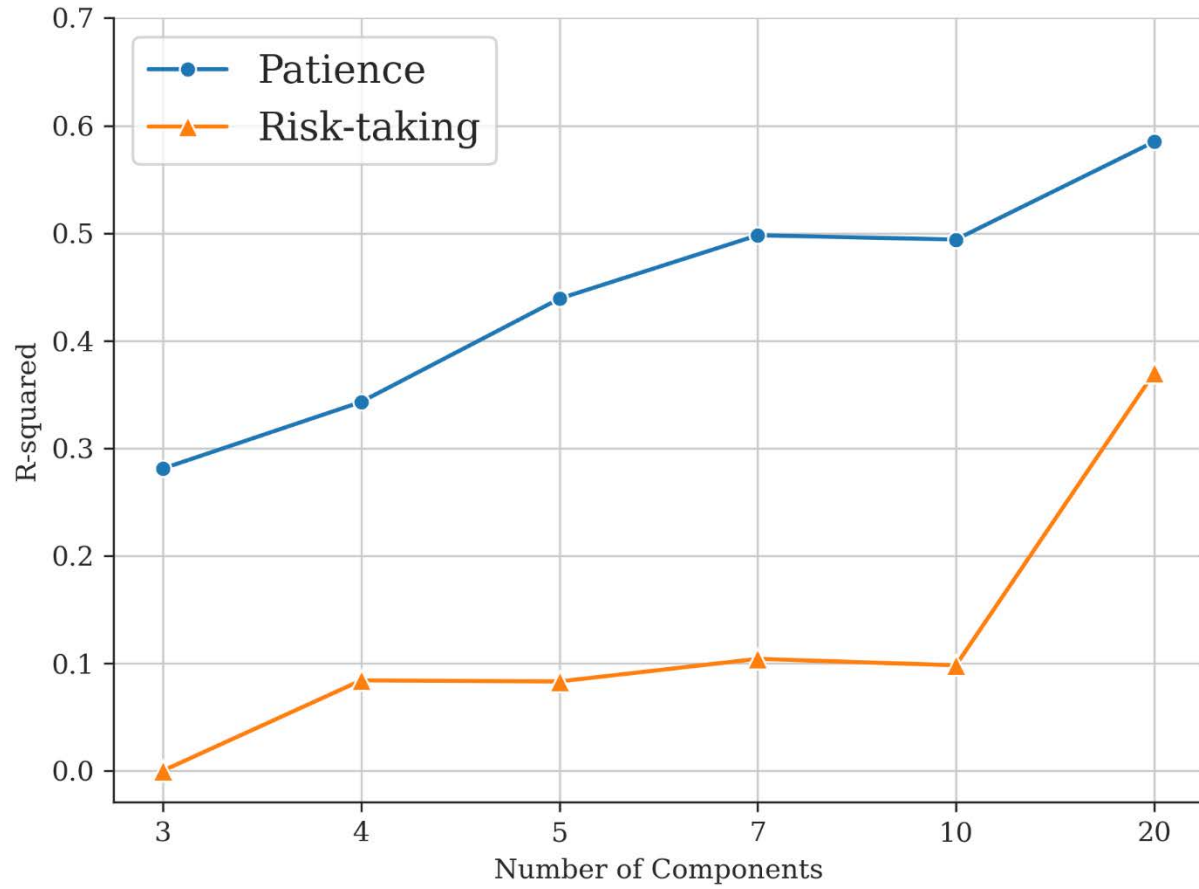
Notes: The top figure shows the cumulative variance in Facebook interests captured by the PCs of the Facebook interests in the Italian regions, the bottom figure shows the variance captured by each component.

**Figure A4: Variance in Facebook Interests Captured by PCs: U.S. States**



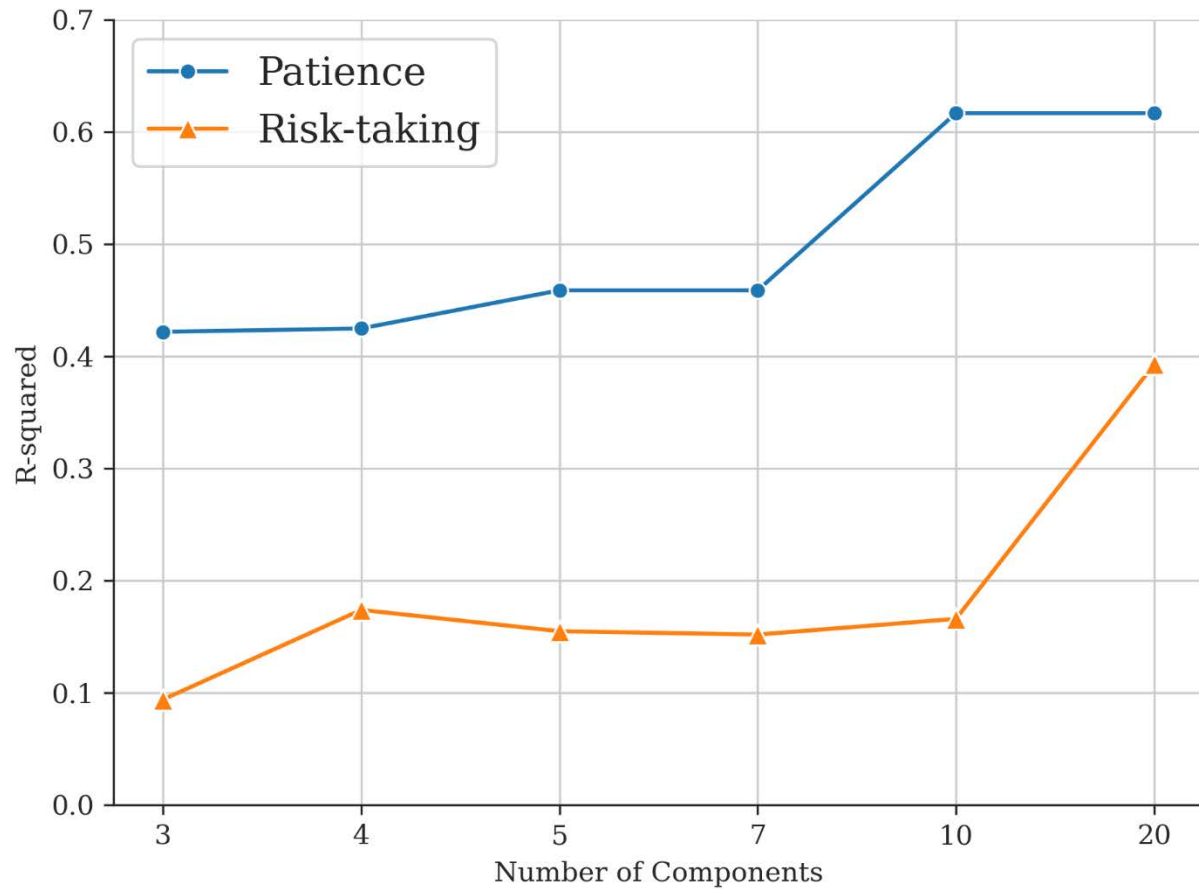
Notes: The top figure shows the cumulative variance in Facebook interests captured by the PCs of the Facebook interests in the U.S. states, the bottom figure shows the variance captured by each component.

**Figure A5: Performance of GPS Prediction with Facebook Interests: PC Loadings from Italian Regions**



Notes: The figure shows the  $R^2$  of regressions of the GPS measures of patience and risk-taking, respectively, on the PCs of Facebook interests (obtained with PC loadings of Italian-region-level Facebook interests) for different numbers of PCs used in the regression. 10-fold cross-validated LASSO model. Sample: all 74 countries for which GPS and Facebook data are available.

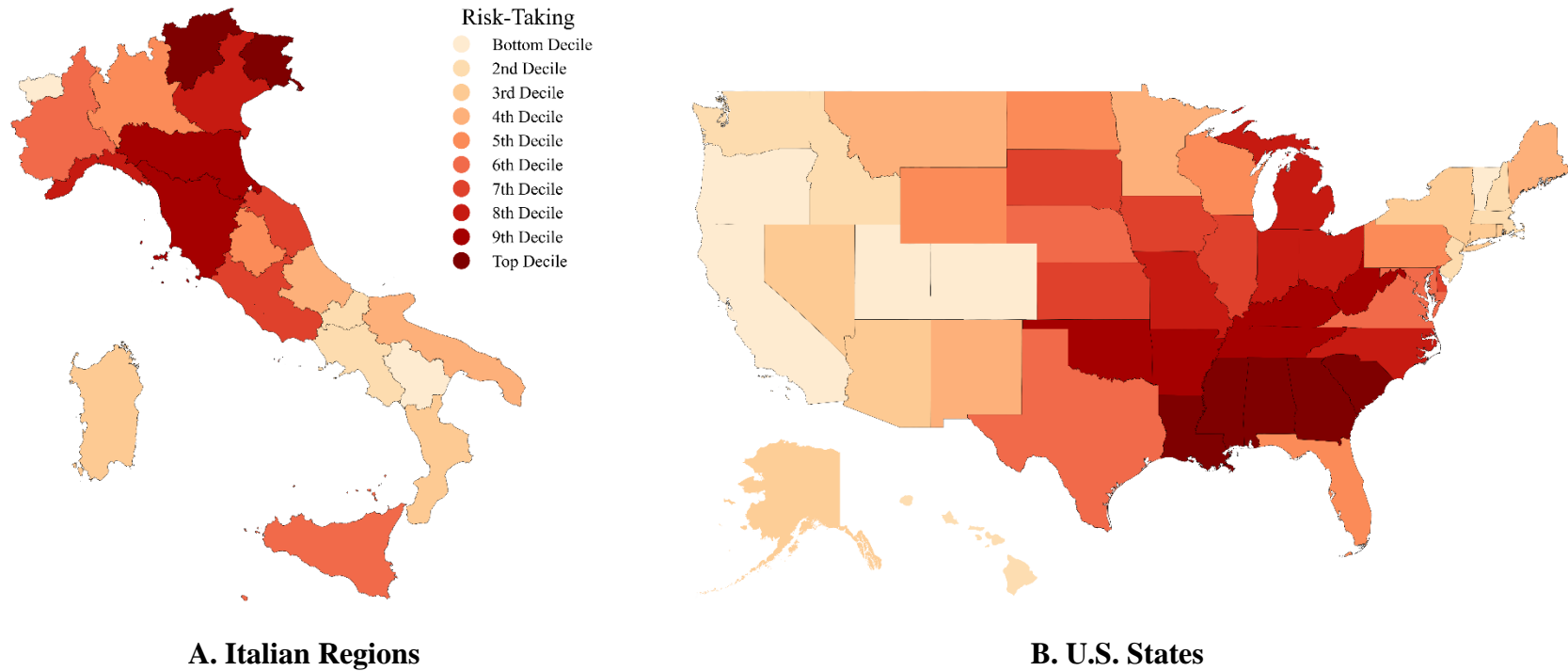
**Figure A6: Performance of GPS Prediction with Facebook Interests: PC Loadings from U.S. States**



Notes: The figure shows the  $R^2$  of regressions of the GPS measures of patience and risk-taking, respectively, on the PCs of Facebook interests (obtained with PC loadings of U.S. state-level Facebook interests) for different numbers of PCs used in the regression. 10-fold cross-validated LASSO model. Sample: all 74 countries for which GPS and Facebook data are available.

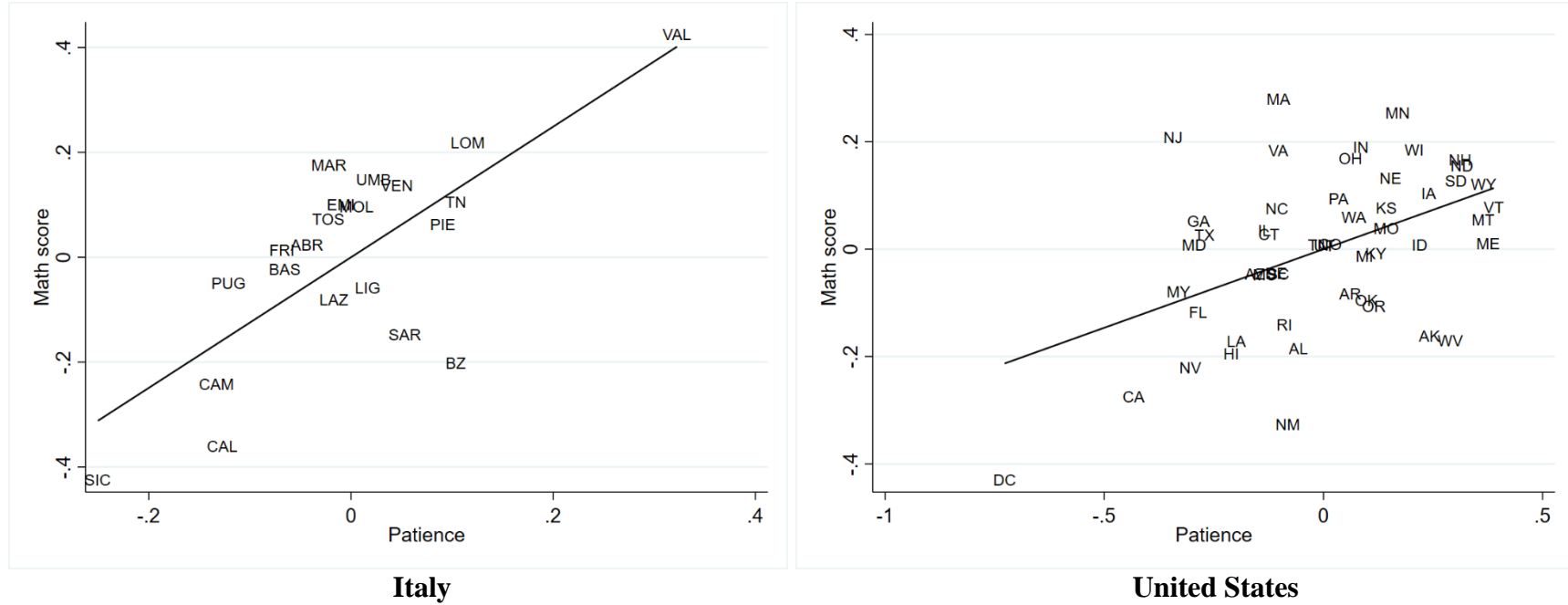


**Figure A7: Measure of Risk-Taking Derived from Facebook Interests for Italian Regions and U.S. States**



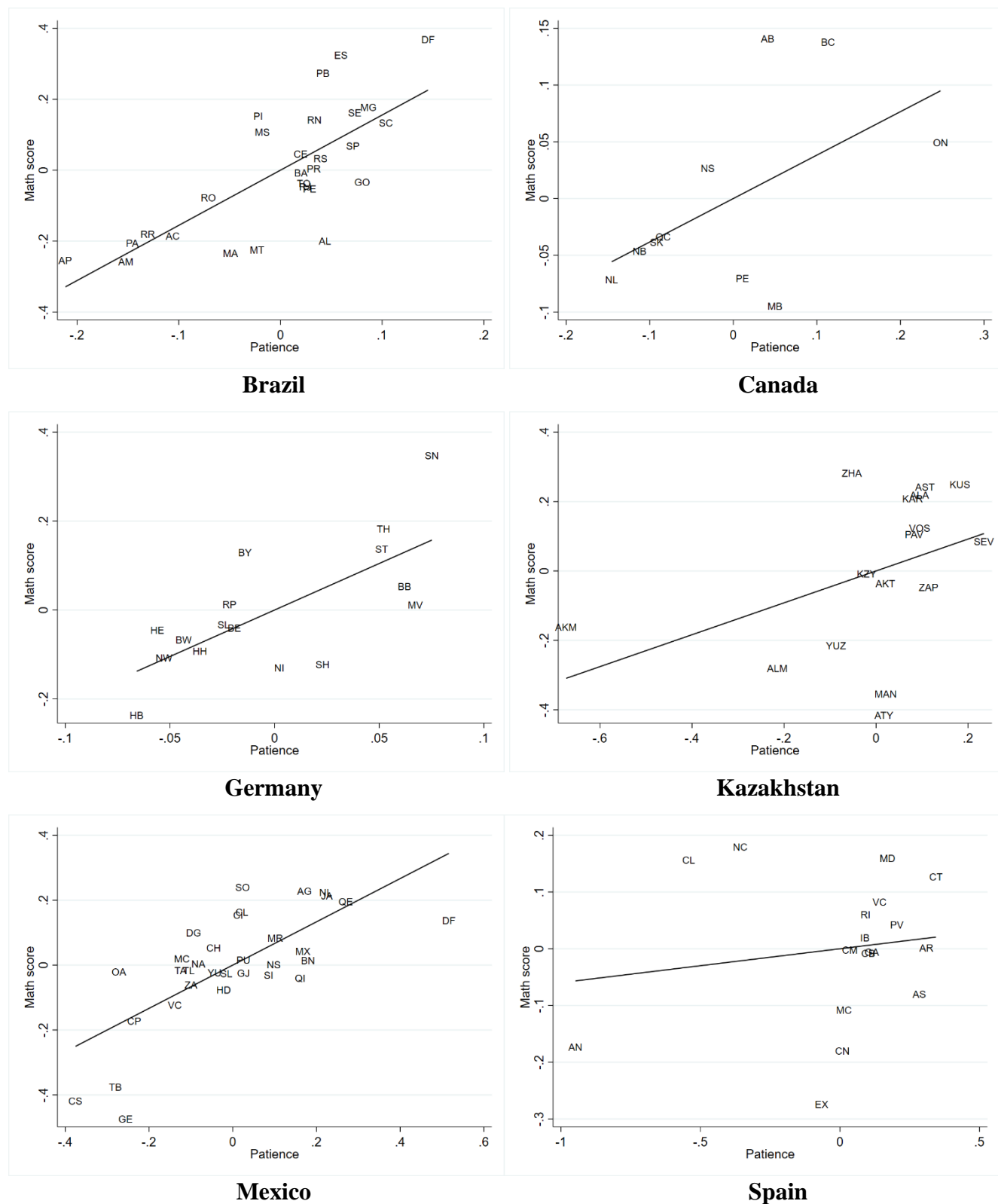
Notes: The figures show maps of the Facebook-derived measure of risk-taking obtained with 4 PCs for Italian regions (Panel A) and U.S. states (Panel B), respectively. Each color corresponds to a decile of the distribution of risk-taking within each country. Darker colors denote higher levels of risk-taking.

**Figure A8: Patience and Student Achievement across Italian Regions and U.S. States**



Notes: Scatterplots of math test scores (wave averages) against patience across regions, both conditional on risk-taking. See Table 2 for underlying measures.

**Figure A9: Patience and Student Achievement across Subnational Regions**



Notes: Scatterplots of math test scores (wave averages) against patience across regions, both conditional on risk-taking. See Table 5 for underlying measures.

**Table A1: Countries in the Cross-country Validation Exercise**

	PISA countries			Training sample
	Only Facebook (1)	Only GPS (2)	Facebook and GPS (3)	Facebook and GPS (4)
Afghanistan				x
Albania	x			
Algeria			x	x
Argentina			x	x
Australia			x	x
Austria			x	x
Azerbaijan	x			
Bangladesh				x
Belarus	x			
Belgium	x			
Bolivia				x
Bosnia and Herzegovina			x	x
Botswana				x
Brazil			x	x
Brunei Darussalam	x			
Bulgaria	x			
Cambodia				x
Cameroon				x
Canada			x	x
Chile			x	x
China				x
Colombia			x	x
Costa Rica			x	x
Croatia			x	x
Czech Republic			x	x
Denmark	x			
Dominican Republic	x			
Egypt				x
Estonia			x	x
Finland			x	x
France			x	x
Georgia			x	x
Germany			x	x
Ghana				x
Greece			x	x
Guatemala				x
Haiti				x
Hong Kong	x			
Hungary			x	x
Iceland	x			
India				x
Indonesia			x	x
Iraq				x
Ireland	x			
Israel			x	x
Italy			x	x
Japan			x	x
Jordan			x	x

(continued on next page)

**Table A1 (continued)**

	PISA countries			Training sample
	Only Facebook (1)	Only GPS (2)	Facebook and GPS (3)	Facebook and GPS (4)
Kazakhstan			x	x
Kenya				x
Korea			x	x
Kyrgyzstan	x			
Latvia	x			
Lebanon	x			
Liechtenstein	x			
Lithuania			x	x
Luxembourg	x			
Macao	x			
Malawi				x
Malaysia	x			
Malta	x			
Mauritius	x			
Mexico			x	x
Moldova			x	x
Montenegro	x			
Morocco			x	x
Netherlands			x	x
New Zealand	x			
Nicaragua				x
Nigeria				x
North Macedonia	x			
Norway	x			
Pakistan				x
Panama	x			
Peru			x	x
Philippines			x	x
Poland			x	x
Portugal			x	x
Qatar	x			
Romania			x	x
Russia		x		
Rwanda				x
Saudi Arabia			x	x
Serbia			x	x
Singapore	x			
Slovakia	x			
Slovenia	x			
South Africa				x
Spain			x	x
Sri Lanka				x
Suriname				x
Sweden			x	x
Switzerland			x	x
Tanzania				x
Thailand			x	x
Trinidad and Tobago	x			

(continued on next page)

**Table A1 (continued)**

	PISA countries			Training sample
	Only Facebook (1)	Only GPS (2)	Facebook and GPS (3)	Facebook and GPS (4)
Tunisia	x			
Turkey			x	x
Uganda				x
Ukraine			x	x
United Arab Emirates			x	x
United Kingdom			x	x
United States			x	x
Uruguay	x			
Venezuela				x
Vietnam			x	x
Zimbabwe				x
Total: 107 countries	32	1	48	74

Notes: Sample of countries: Col. 1-3: countries included in the cross-country validation exercise (Panel A of Table 1). Col. 4: countries included in training the machine-learning model. Country names are as reported in PISA codebooks or Facebook/GPS data and do not represent any political views of the authors.

**Table A2: Validation of Cross-Country Analysis: Different Numbers of Principal Components (PCs)**

	20 PCs (1)	30 PCs (2)	40 PCs (3)	50 PCs (4)
<b>A. Original country sample (GPS countries)</b>				
Patience	1.638*** (0.136)	1.588*** (0.141)	1.593*** (0.138)	1.660*** (0.146)
Risk-taking	-1.525*** (0.443)	-1.089*** (0.382)	-0.989*** (0.332)	-0.940*** (0.332)
Control variables	Yes	Yes	Yes	Yes
Observations	1,954,840	1,954,840	1,954,840	1,954,840
Residence countries	48	48	48	48
$R^2$	0.208	0.199	0.198	0.197
<b>B. Extended country sample (all Facebook countries)</b>				
Patience	1.682*** (0.125)	1.583*** (0.128)	1.588*** (0.129)	1.689*** (0.135)
Risk-taking	-1.531*** (0.330)	-1.422*** (0.331)	-1.224*** (0.281)	-1.089*** (0.279)
Control variables	Yes	Yes	Yes	Yes
Observations	2,660,408	2,660,408	2,660,408	2,660,408
Residence countries	80	80	80	80
$R^2$	0.204	0.203	0.198	0.194

Notes: Dependent variable: PISA math test score in all PISA waves 2000-2018. Least squares regressions weighted by students' sampling probability. Control variables: student gender, age, and migration status; imputation dummies; and wave fixed effects. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: PISA international student achievement test, 2000-2018; own elaboration of Facebook data.

**Table A3: Countries in the Migrant Analysis**

	GPS/Facebook country of origin			PISA destination country	
	Only GPS (1)	Only Facebook (2)	Both (3)	GPS analysis (4)	Facebook analysis (5)
Afghanistan			x		
Albania		x			
Algeria					
Argentina			x	x	x
Armenia		x			
Australia			x	x	x
Austria			x	x	x
Azerbaijan		x			
Bangladesh			x		
Belarus		x		x	x
Belgium		x		x	x
Bolivia			x		
Bosnia and Herzegovina			x	x	x
Brazil			x		
Brunei Darussalam				x	x
Bulgaria		x			
Cape Verde		x			
Canada			x	x	x
Chile			x		
China			x		
Colombia			x		
Costa Rica				x	x
Croatia			x	x	x
Czech Republic			x	x	x
Denmark		x		x	x
Dominican Republic		x		x	x
Egypt			x		
Estonia			x		
Ethiopia		x			
Fiji		x			
Finland			x	x	x
France			x		
Georgia			x		x
Germany			x	x	x
Greece			x		x
Haiti			x		
Hong Kong				x	x
Hungary			x		
Iceland		x			
India			x		
Indonesia			x	x	x
Iran	x				
Iraq			x		
Ireland		x		x	x
Israel				x	x
Italy			x		
Japan					
Jordan			x	x	x

(continued on next page)



**Table A3 (continued)**

	GPS/Facebook country of origin			PISA destination country	
	Only GPS (1)	Only Facebook (2)	Both (3)	GPS analysis (4)	Facebook analysis (5)
Kazakhstan			x		
Kuwait		x			
Kyrgyzstan					x
Latvia				x	x
Lebanon		x			
Libya		x			
Liechtenstein		x		x	x
Lithuania			x		
Luxembourg				x	x
Macao		x		x	x
Malaysia		x			
Mauritius				x	x
Mexico				x	x
Moldova			x	x	x
Montenegro		x		x	x
Morocco			x	x	x
Netherlands			x	x	x
New Zealand		x		x	x
Nicaragua			x		
Nigeria			x		
North Macedonia				x	x
Norway		x		x	x
Pakistan			x		
Palestine		x			
Panama		x		x	x
Paraguay		x			
Peru					
Philippines			x	x	x
Poland			x		
Portugal			x	x	x
Qatar		x		x	x
Romania			x		
Russia	x				
Samoa		x			
Saudi Arabia			x	x	x
Serbia			x		x
Singapore		x			
Slovakia		x		x	x
Slovenia		x		x	x
Somalia		x			
South Africa			x		
South Korea			x	x	x
Spain			x		
Suriname			x		
Sweden			x		
Switzerland			x	x	x
Tajikistan		x			
Thailand			x		

(continued on next page)

**Table A3 (continued)**

	GPS/Facebook country of origin			PISA destination country	
	Only GPS (1)	Only Facebook (2)	Both (3)	GPS analysis (4)	Facebook analysis (5)
Tonga		x			
Turkey			x	x	x
Ukraine			x	x	x
United Arab Emirates			x		
United Kingdom			x	x	x
United States			x		
Uruguay		x		x	x
Uzbekistan		x			
Venezuela			x		
Vietnam			x		
Yemen		x			
Zambia		x			
Total: 108 countries	2	37	56	46	50

Notes: Sample of countries that serve as countries of origin (col. 1-3) or destination countries (col. 4-5) in the migrant analysis (Panel B of Table 1). Country names are as reported in PISA codebooks or Facebook/GPS data and do not represent any political views of the authors.

**Table A4: Validation of Migrant Analysis: Different Numbers of Principal Components (PCs)**

	20 PCs (1)	30 PCs (2)	40 PCs (3)	50 PCs (4)
<b>A. Original sample (GPS countries of origin)</b>				
Patience	0.783*** (0.193)	0.876*** (0.197)	0.885*** (0.192)	0.875*** (0.216)
Risk-taking	-0.676** (0.306)	0.008 (0.367)	0.087 (0.322)	0.156 (0.371)
Control variables	Yes	Yes	Yes	Yes
Residence-country by wave fixed effects	Yes	Yes	Yes	Yes
Observations	78,403	78,403	78,403	78,403
Countries of origin	56	56	56	56
Residence countries	46	46	46	46
$R^2$	0.271	0.271	0.272	0.270
<b>B. Extended sample (all Facebook countries of origin)</b>				
Patience	0.838*** (0.211)	1.027*** (0.198)	1.033*** (0.191)	0.995*** (0.211)
Risk-taking	-1.155*** (0.422)	-0.067 (0.357)	0.064 (0.297)	0.154 (0.341)
Control variables	Yes	Yes	Yes	Yes
Residence-country by wave fixed effects	Yes	Yes	Yes	Yes
Observations	90,983	90,983	90,983	90,983
Countries of origin	93	93	93	93
Residence countries	50	50	50	50
$R^2$	0.295	0.294	0.294	0.291

Notes: Dependent variable: PISA math test score, waves 2003-2018. Least squares regressions, including 180 fixed effects for each residence-country by wave cell. Sample: students with both parents not born in the country where the student attends school. Control variables: student gender, age, dummy for OECD country of origin, imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: PISA international student achievement test, 2003-2018; own elaboration of Facebook data.

**Table A5: Patience and Reading Achievement: Analysis of Italian Regions**

	4 PCs (1)	7 PCs (2)	10 PCs (3)
<b>A. Individual level</b>			
Patience	1.218*** (0.201)	0.986*** (0.123)	1.050*** (0.128)
Control variables	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes
Observations	59,441	59,441	59,441
Regions	20	20	20
$R^2$	0.105	0.110	0.110
<b>B. Regional level</b>			
Patience	0.905*** (0.177)	0.716*** (0.094)	0.762*** (0.098)
Wave fixed effects	Yes	Yes	Yes
Observations	42	42	42
Regions	20	20	20
$R^2$	0.496	0.617	0.625

Notes: Dependent variable: INVALSI eighth-grade reading test score in waves 2018 and 2019. Least squares regressions with wave fixed effects. Unit of observation: Panel A: student; Panel B: region-wave combination. Col. 1-3 use the patience measure computed with 4, 7, and 10 principal components (PCs), respectively. Regressions control for risk-taking computed with the equivalent number of PCs. Additional control variables (Panel A): student gender, age, and migration status; imputation dummies. Robust standard errors adjusted for clustering at the regional level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: INVALSI reading achievement test, 2017-2019; own elaboration of Facebook data.

**Table A6: Patience and Math Achievement: Analysis of Italian Regions by Wave and Gender**

	2018 (1)	2019 (2)	Males (3)	Females (4)
<b>A. Individual level</b>				
Patience (4 PCs)	1.588*** (0.191)	1.422*** (0.217)	1.579*** (0.211)	1.427*** (0.198)
Control variables	Yes	Yes	Yes	Yes
Wave fixed effects	No	No	Yes	Yes
Observations	29,359	29,675	30,530	28,504
Regions	20	20	20	20
R <sup>2</sup>	0.095	0.089	0.097	0.082
<b>B. Regional level</b>				
Patience (4 PCs)	1.331*** (0.221)	1.161*** (0.241)	1.305*** (0.226)	1.185*** (0.227)
Wave fixed effects	No	No	Yes	Yes
Observations	21	21	42	42
Regions	20	20	20	20
R <sup>2</sup>	0.693	0.668	0.682	0.657

Notes: Dependent variable: INVALSI eighth-grade math test score in waves 2018 and 2019. Least squares regressions with wave fixed effects. Unit of observation: Panel A: student; Panel B: region-wave combination. Patience measure computed with 4 principal components (PCs). Regressions control for risk-taking computed with 4 PCs. Additional control variables (Panel A): student gender, age, and migration status; imputation dummies. Robust standard errors adjusted for clustering at the regional level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: INVALSI reading achievement test, 2017-2019; own elaboration of Facebook data.

**Table A7: Patience and Math Achievement: Analysis of Italian Regions by Migrant Status**

	4 PCs (1)	7 PCs (2)	10 PCs (3)
<b>A. Native students</b>			
Patience	1.581*** (0.188)	1.423*** (0.115)	1.514*** (0.118)
Control variables	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes
Observations	51,691	51,691	51,691
Regions	20	20	20
$R^2$	0.084	0.091	0.091
<b>B. Second-generation migrant students</b>			
Patience	0.909*** (0.237)	0.748*** (0.215)	0.820*** (0.220)
Wave fixed effects	Yes	Yes	Yes
Observations	3,572	3,572	3,572
Regions	20	20	20
$R^2$	0.033	0.035	0.035
<b>C. First-generation migrant students</b>			
Patience	0.565** (0.235)	0.842*** (0.112)	0.893*** (0.124)
Wave fixed effects	Yes	Yes	Yes
Observations	1,719	1,719	1,719
Regions	20	20	20
$R^2$	0.079	0.083	0.083

Notes: Dependent variable: INVALSI eighth-grade math test score in waves 2018 and 2019. Least squares regressions with wave fixed effects. Unit of observation: student. Col. 1-3 use the patience measure computed with 4, 7, and 10 principal components (PCs), respectively. Regressions control for risk-taking computed with the equivalent number of PCs. Additional control variables: student gender and age; imputation dummies. Robust standard errors adjusted for clustering at the regional level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: INVALSI mathematics achievement test, 2017-2019; own elaboration of Facebook data.

**Table A8: Patience and Math Achievement: Conditioning on Trust**

	4 PCs (1)	7 PCs (2)	10 PCs (3)
<b>A. Italy (individual level)</b>			
Patience	1.505*** (0.197)	1.319*** (0.106)	1.403*** (0.113)
Control variables	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes
Observations	59,034	59,034	59,034
Regions	20	20	20
$R^2$	0.092	0.101	0.101
<b>B. Italy (regional level)</b>			
Patience	1.245*** (0.193)	1.102*** (0.088)	1.171*** (0.093)
Wave fixed effects	Yes	Yes	Yes
Observations	42	42	42
Regions	20	20	20
$R^2$	0.679	0.827	0.828
<b>C. United States (state level)</b>			
Patience	0.312 (0.966)	0.167* (0.092)	0.274** (0.133)
Wave fixed effects	Yes	Yes	Yes
Observations	153	153	153
Regions	51	51	51
$R^2$	0.360	0.362	0.372

Notes: Dependent variable: Panels A and B: INVALSI eighth-grade math test score in waves 2018 and 2019; Panel C: NAEP eighth-grade math test score in all NAEP waves 2015-2019. Least squares regressions with wave fixed effects. Unit of observation: Panel A: student; Panel B: region-wave combination; Panel C: state-wave combination. Patience measured at the regional/state level throughout. Col. 1-3 use the patience measure computed with 4, 7, and 10 principal components (PCs), respectively. Regressions control for risk-taking and trust computed with the equivalent number of PCs. Additional control variables (Panel A): student gender, age, and migration status; imputation dummies. The Italian region of Trentino-Alto-Adige is represented by the two municipalities of Bolzano and Trento. Robust standard errors adjusted for clustering at the regional (state) level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: INVALSI mathematics achievement test, 2017-2019; NAEP mathematics achievement test, 2015-2019; own elaboration of Facebook data.

**Table A9: Patience and Math Achievement: Analysis of Italian Regions Excluding Trentino-Alto-Adige**

	4 PCs (1)	7 PCs (2)	10 PCs (3)
<b>A. Individual level</b>			
Patience	1.717*** (0.158)	1.412*** (0.122)	1.520*** (0.124)
Control variables	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes
Observations	55,437	55,437	55,437
Regions	19	19	19
$R^2$	0.095	0.098	0.098
<b>B. Regional level</b>			
Patience	1.462*** (0.171)	1.220*** (0.094)	1.314*** (0.097)
Wave fixed effects	Yes	Yes	Yes
Observations	38	38	38
Regions	19	19	19
$R^2$	0.783	0.835	0.846

Notes: Dependent variable: INVALSI eighth-grade math test score in waves 2018 and 2019. Least squares regressions with wave fixed effects. Unit of observation: Panel A: student; Panel B: region-wave combination. Students in the autonomous municipalities of Trento and Bolzano are dropped from the estimation sample. Col. 1-3 use the patience measure computed with 4, 7, and 10 principal components (PCs), respectively. Regressions control for risk-taking computed with the equivalent number of PCs. Additional control variables (Panel A): student gender, age, and migration status; imputation dummies. Robust standard errors adjusted for clustering at the regional level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: INVALSI mathematics achievement test, 2017-2019; own elaboration of Facebook data.



**Table A10: Analysis of Unobservable Selection and Coefficient Stability following Oster (2019): Analysis of Italian Regions**

	4 PCs		7 PCs		10 PCs	
	Restricted	Extended	Restricted	Extended	Restricted	Extended
	(1)	(2)	(3)	(4)	(5)	(6)
Patience	1.252*** (0.210)	1.505*** (0.197)	1.136*** (0.122)	1.350*** (0.114)	1.208*** (0.129)	1.437*** (0.117)
Control variables	No	Yes	No	Yes	No	Yes
Wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	59,034	59,034	59,034	59,034	59,034	59,034
Regions	20	20	20	20	20	20
$R^2$	0.043	0.092	0.049	0.099	0.050	0.099
Oster (2019) diagnostics						
Bound $\beta^*$ for $\delta = 1$		1.705		1.487		1.581
$\delta$ to match $\beta = 0$		-4.117		-2.687		-2.680

Notes: Dependent variable: INVALSI eighth-grade math test score in waves 2018 and 2019. Least squares regressions with wave fixed effects. Unit of observation: student. Students in the autonomous municipalities of Trento and Bolzano are dropped from the estimation sample. Patience measure computed with number of principal components (PCs) indicated in column header. Regressions control for risk-taking computed with the equivalent number of PCs. Odd columns: restricted model with wave fixed effects. Even columns: baseline models with wave fixed effects, student gender, age, and migration status; imputation dummies. Oster statistics computed using  $R_{max} = 1.3\tilde{R}$ , where  $\tilde{R}$  denotes the  $R^2$  reported in even columns. Robust standard errors adjusted for clustering at the regional level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: INVALSI mathematics achievement test, 2017-2019; own elaboration of Facebook data.

**Table A11: Patience and Math Achievement: Analysis of Italian Regions using PISA Data**

	4 PCs (1)	7 PCs (2)	10 PCs (3)
Patience	1.484*** (0.264)	1.473*** (0.132)	1.570*** (0.138)
Control variables	Yes	Yes	Yes
Observations	31,073	31,073	31,073
States	20	20	20
$R^2$	0.106	0.113	0.113

Notes: Dependent variable: PISA 2012 math test score. Least squares regressions. Unit of observation: student. Col. 1-3 use the patience measure computed with 4, 7, and 10 principal components (PCs), respectively. Regressions control for risk-taking computed with the equivalent number of PCs. Additional control variables: student gender, age, and migration status; imputation dummies. Robust standard errors adjusted for clustering at the regional level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: PISA student achievement test, 2012; own elaboration of Facebook data.

**Table A12: Patience and Reading Achievement: Analysis of U.S. States**

	4 PCs	7 PCs	10 PCs
	(1)	(2)	(3)
Patience	0.228*** (0.074)	0.141* (0.077)	0.227** (0.103)
Wave fixed effects	Yes	Yes	Yes
Observations	153	153	153
States	51	51	51
$R^2$	0.385	0.375	0.396

Notes: Dependent variable: NAEP eighth-grade reading test score in all NAEP waves 2015-2019. Least squares regressions with wave fixed effects. Unit of observation: state-wave combination. Col. 1-3 use the patience measure computed with 4, 7, and 10 principal components (PCs), respectively. Regressions control for risk-taking computed with the equivalent number of PCs. Robust standard errors adjusted for clustering at the state level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: NAEP mathematics achievement test, 2015-2019; own elaboration of Facebook data.

**Table A13: Patience and Math Achievement: Analysis of U.S. States by Wave**

	4 PCs (1)	7 PCs (2)	10 PCs (3)
<b>A. 2015</b>			
Patience	0.335*** (0.081)	0.194** (0.082)	0.346*** (0.119)
States	51	51	51
$R^2$	0.426	0.410	0.430
<b>B. 2017</b>			
Patience	0.309*** (0.084)	0.179** (0.085)	0.290** (0.125)
States	51	51	51
$R^2$	0.373	0.360	0.372
<b>C. 2019</b>			
Patience	0.235*** (0.077)	0.142* (0.077)	0.228* (0.114)
States	51	51	51
$R^2$	0.277	0.267	0.278

Notes: Dependent variable: NAEP eighth-grade math test score in all NAEP waves 2015-2019. Least squares regressions with wave fixed effects. Unit of observation: state-wave combination. Col. 1-3 use the patience measure computed with 4, 7, and 10 principal components (PCs), respectively. Regressions control for risk-taking computed with the equivalent number of PCs. Robust standard errors adjusted for clustering at the state level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: NAEP mathematics achievement test, 2015-2019; own elaboration of Facebook data.

**Table A14: Patience and Math Achievement: Analysis of U.S. States by Gender**

	4 PCs (1)	7 PCs (2)	10 PCs (3)
<b>A. Males</b>			
Patience	0.322*** (0.101)	0.194* (0.108)	0.305** (0.147)
Wave fixed effects	Yes	Yes	Yes
Observations	153	153	153
States	51	51	51
$R^2$	0.388	0.377	0.385
<b>B. Females</b>			
Patience	0.263*** (0.079)	0.147* (0.086)	0.258** (0.119)
Wave fixed effects	Yes	Yes	Yes
Observations	153	153	153
States	51	51	51
$R^2$	0.319	0.304	0.321

Notes: Dependent variable: NAEP eighth-grade math test score in all NAEP waves 2015-2019. Least squares regressions with wave fixed effects. Unit of observation: state-wave combination. Col. 1-3 use the patience measure computed with 4, 7, and 10 principal components (PCs), respectively. Regressions control for risk-taking computed with the equivalent number of PCs. Robust standard errors adjusted for clustering at the state level in parentheses. Significance level: \*\*\* 1 percent, \*\* 5 percent, \* 10 percent. Data sources: NAEP mathematics achievement test, 2015-2019; own elaboration of Facebook data.