International Review of Economics Education xxx (xxxx) xxxx

Contents lists available at ScienceDirect

International Review of Economics Education

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

# Economic competence in early secondary school: Evidence from a large-scale assessment in Germany

Luis Oberrauch[a], Tim Kaiser[a,b,*]

[a] University of Koblenz-Landau, 76829, Landau, Germany
[b] German Institute for Economic Research (DIW Berlin), 76829, Landau, Germany

ARTICLE INFO

*JEL Classification:*
A21
I21

*Keywords:*
Economic competence
Economic literacy
Pre-college economic education
Item response theory
Gender gap

ABSTRACT

We employ a psychometrically validated performance test to study economic competence among representative sample of 1,687 early secondary school students in Southwest Germany. The rich dataset allows us to study variation in economic competence across school types and observable student characteristics. Our results show that economic competence is significantly lower among female students, migrants, students with parents of low socio-economic status and those who do not attend the highest track school type. Additionally, quantile regression analyses suggest that the gender gap increases along the distribution of economic competence and that effects of parents with high socio-economic status are more pronounced above the median of the competence distribution.

## 1. Introduction

Research on the economic understanding of high school and undergraduate students has a long tradition in economics and economic education research (see Allgood et al., 2015; Walstad and Rebeck, 2001a; Becker et al., 1990; Siegfried and Fels, 1979 for reviews). Much of the evidence on economic literacy of school students comes from the U.S., as there is an increasing number of federal states implementing high school economic education and personal financial education mandates (see Urban et al., 2018 for a study using the most comprehensive database of mandates). While the recent attention to the topic of limited financial literacy (i.e., a subset of economic literacy) among various populations has spurred research around the globe, the number of studies from outside the U.S. remains limited (see Miller et al., 2015; Kaiser and Menkhoff, 2017 for reviews).

Evidence on the level of economic literacy of pre-university students in Germany is particularly scarce. This is partly due to the reason that economic education mandates in Germany have been very limited and virtually non-existing in the past (see Brückner et al., 2015a,b; Remmele and Seeber, 2012). While there are two rigorous studies evaluating school *financial* education interventions in Germany (Lührmann et al., 2015, 2018), explorative studies on *financial* literacy among high school and (first-year) university students relying on non-representative convenience samples (Schuhen and Schürkmann, 2014; Erner et al., 2016; Förster et al., 2017; Kaiser, 2017; Happ and Förster, 2018), and studies assessing economic literacy of (first-year) undergraduate university students in convenience samples (e.g. Happ et al., 2016), there is no representative study of knowledge and ability in the economic domain among secondary schools in Germany, to date.

This paper seeks to address this gap in the literature and is situated in context of a recent curriculum reform in the state of Baden-Württemberg, Germany, introducing mandatory economic education as a separate school subject from grades 7–10 for all general

* Corresponding author at: University of Koblenz-Landau and German Institute for Economic Research (DIW Berlin), 76829 Landau, Germany.
*E-mail addresses:* oberrauch@uni-landau.de (L. Oberrauch), kaiser@uni-landau.de (T. Kaiser).

https://doi.org/10.1016/j.iree.2019.100172
Received 18 December 2018; Received in revised form 17 June 2019; Accepted 21 November 2019

education school types. For this purpose, we employ a newly developed performance test (Seeber et al., 2018, 2019) which is derived from a competence model underlying economic education standards in Germany and the school economic education curriculum in the federal state our study is set in (see Retzmann and Seeber, 2016). Our paper presents evidence from a representative large-scale assessment of early secondary school students in Baden-Württemberg at the end of grade seven, belonging to the last cohort of students *without* mandatory economic education. Thus, we seek to provide an empirical investigation into the ex-ante levels of economic competence and to explore existing differences in economic competence across school types and observable student characteristics.

The study results in three main findings: First, we document a gender gap in economic competence favoring male students, which is in line with the literature on economic literacy in U.S. and U.K. high schools (e.g. Heath, 1989; Davies et al., 2005; Walstad, 2013), a recent study on financial literacy of secondary school students in Germany (Erner et al., 2016), international evidence on gender differences among undergraduate students (Brückner et al., 2015a,b), and the (adult) financial literacy literature (cf. Lusardi and Mitchell, 2008, 2014 Lusardi and Mitchell, 2008; Bucher-Koenen and Lusardi, 2011). Additionally, we document that the gender gap increases along the distribution of economic competence, i.e., the gender gap being especially evident among high performing students.

Second, we show that children of migrants and children of parents of lower socio- economic status (SES) have, on average, lower levels of economic competence. The effect of the parents' SES is more pronounced at the higher moments of the distribution. The effect of being a child of migrant parents is smaller at the top of the competence distribution. Additionally, we document positive correlations between self-reported math and reading skills and economic competence. However, these self-reported measures are less predictive of performance on the test for individuals at the bottom of the competence distribution.

Third, we document differences in average competence levels between the different school types in Germany. As the German system sorts secondary school children into different ability-tracks, we find that students in the lower track schools (*Realschule and Werkrealschule*) have, on average, substantially lower levels of economic competence than those in the highest track school type (*Gymnasium*). Comparing the highest track school type (*Gymnasium*) to a comprehensive school type accommodating students of all ability-levels (*Gemeinschaftsschule*) also results in a large and significant difference. Differences between the lower tier school types (*Werkrealschule and Realschule*) are relatively small. Overall, our analysis serves as an important baseline benchmark and informs about existing heterogeneity in economic competence among the target group of secondary school students in absence of formal instruction in economics.

This paper is structured into five further sections: Section 2 describes the sampling procedure (2.1) and provides descriptive statistics for the sample of schools and students covered (2.2). Section 3 describes the measurement of economic competence: Section 3.1 discusses the underlying conceptual model and the process of item development. Section 3.2 describes the details of the statistical measurement model, while our strategy to deal with missing values is outlined in Section 3.3. The regression models used to study observable predictors of economic competence are described in Section 4. Section 5 presents results and mentions robustness checks while Section 6 discusses these results with regard to prior literature and concludes.

## 2. Data

In the following, we describe the sampling procedures (2.1) and provide an overview of sample descriptive statistics (2.2).

### 2.1. Sampling

We initiate our sampling process by partitioning the whole population of interest, e.g., all students in 7th grade visiting a public school in the German federal state of Baden-Wuerttemberg, into subgroups by variables that indicate a strong relationship with our target variable (explicit stratification). We stratify by school type and degree of urbanization. Thus, our classification of the sample population results in twelve strata (four school types within three degrees of urbanization ("high", "medium", "low")). We account for the degree of urbanization by classifying administrative districts with regard to their population per square kilometers ratio: A population density of below 220 is defined as a "low degree of urbanization", between 220 and 485 is defined as "medium degree of urbanization" and above 485 is defined as a "high degree of urbanization" (see Figs. A1a and A1b in the Appendix A).

As administrative records of the entire student-population of interest are not available to us, we follow a two-stage cluster sampling procedure with selection of schools in the first stage and a random selection of one 7th grade class per school in the second stage. The number of selection elements (schools) in each stratum is adapted to the proportions of strata in the target population (*probability proportional to size*). Beside explicit stratification variables, we introduce *school size* as an implicit stratification variable. To ensure representation of schools of all sizes, we sized strata and deployed systematic sampling using sampling intervals in each stratum (e.g. Lohr, 2010). Thus, our final sample is comprised of 1,687 students in 84 classes (sampling proportions by stratum are listed in Table A1 in the Appendix A).

We compensate any remaining disproportionalities by means of including design weights in all of our analyses. These are defined as the inverse of selection probability $p^{(j,c)}$. The total selection probability is obtained by multiplying the school selection probability $p^{(j)}$ with the class selection probability $p^{(c)}$, e.g. $p^{(s,c)} = p^{(j)} \times p^{(c)}$. We calculate the school selection probability using the quotient of the number of selected schools $N^{(j)}$ and the number of total schools in strata s $N^{(s)}$ ($p^{(j)} = \frac{N^{(j)}}{N^{(s)}}$). Every available class at the end of 7th grade $C^{(j)}$ is selected with equal probability ($p^{(c)} = \frac{1}{\sum C^{(j)}}$). Student weights $w^{(j,c)}$ follow from the inverse of the total selection probability: $\frac{1}{w^{(j,c)}}$. Finally, student weights are normalized to the number of sampled participants, $w_{norm}^{(j,c)} = w^{(j,c)} \times \frac{1687}{\sum w^{(j,c)}}$.

**Table 1**
Descriptive statistics at the student and school level.

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| **Student level** | | | | | |
| WLE.500 | 1,687 | 500 | 100 | 87.308 | 884.689 |
| WLE | 1,687 | 0.095 | 0.984 | −3.967 | 3.882 |
| male | 1,687 | 0.551 | | 0 | 1 |
| age | 1,561 | 13.88 | 0.740 | 12 | 19 |
| mig | 1,561 | 0.411 | | 0 | 1 |
| native | | | | | |
| German | 1,563 | 0.627 | | 0 | 1 |
| bilingual | 1,563 | 0.251 | | 0 | 1 |
| foreign language | 1,563 | 0.122 | | 0 | 1 |
| book | 1,545 | 3.442 | 1.619 | 1 | 6 |
| selfread | 1,687 | 3.864 | 0.712 | 1 | 5 |
| selfmath | 1,681 | 3.512 | 0.850 | 1 | 5 |
| selfall | 1,677 | 3.663 | 0.638 | 1 | 5 |
| interesteco | 1,659 | 2.525 | 0.758 | 1 | 4 |
| interestimp | 1,653 | 2.970 | 0.752 | 1 | 4 |
| ownsalary | 1,536 | 0.667 | | 0 | 1 |
| time | 1,687 | 27.052 | 8.683 | 3.331 | 60.34 |
| bank.exp | 1,498 | −0.004 | 0.811 | −0.884 | 2.165 |
| **School level** | | | | | |
| GYM | 1,687 | 0.391 | | 0 | 1 |
| RS | 1,687 | 0.303 | | 0 | 1 |
| WRS | 1,687 | 0.145 | | 0 | 1 |
| GMS | 1,687 | 0.161 | | 0 | 1 |
| School size | 1,687 | 658.4 | 222.5 | 226 | 1292 |
| % non-natives/class | 1,687 | 34.56 | 20,672 | 0 | 100 |
| Urbanization | 1,687 | 1.939 | 0.808 | 1 | 3 |

### 2.2. Descriptive statistics

We capture demographic data at the individual level and school characteristics at the cluster level as shown in Table 1.

The dependent variable *WLE* represents the estimated person ability from the one-parametric Rasch Model on a logit scale (see section 3.2) and *WLE.500* is its transformation to have a mean of 500 and a standard deviation of 100 in the entire sample of students, as is the usual convention for large-scale assessments such as PISA (see e.g., OECD, 2014). For the variable *male*, "0" represents female and "1" represents male participants. 55 percent of students in our sample are male. Migration background *mig* is defined as at least one of the students' parents being born abroad (41 percent of students). In terms of mother tongue, represented by the categorical variable *native*, we asked participants for the primary language in their childhood: German (1) (63 percent), bilingual (2) (25 percent) or a foreign language (3) (12 percent). We also proxy for the socio-economic status of parents (*book)* by asking participants how many books (excluding school books and magazines) are available within the household on a scale from 1 (none) to 6 (several bookshelves) as this is the standard in educational large-scale assessments (e.g., OECD, 2014). The mean value of this variable is 3.44 with a standard deviation of 1.62.

In order to control for students' school performance, we ask for (self-reported) reading abilities (*selfread*) (M = 3.86, SD = 0.71), mathematical abilities (*selfmath*) (M = 3.51, SD = 0.85), and overall school performance (*selfall*) (M = 3.66, SD = 0.64) on a 1 (low) to 5 (high) scale. Moreover, we ask students about their interest in economics (*interesteco*) (M = 2.53, SD = 0.76) and their subjective assessment about the importance of economic knowledge in general (*interestimp*) (M = 2.97, SD = 0.75) on a 1 (low) to 4 (high) scale. We assess financial behavior by asking students whether they have a bank account, an ATM card (credit or debit) and the frequency of usage for both the account and ATM card. As all of the four variables capturing financial behaviors are strongly correlated, we built an equally weighted z-score summary index of all components (*bankexp*) based on the method described in Kling et al. (2007). We also measure labor income (*ownsalary*) from student-jobs (67 percent of students have a student-job). The average survey duration (*time*) is 27 min (SD = 8.68).

At the school level, we note that 39 percent of sampled students in the 7$^{th}$ grade in Baden-Wuerttemberg visit the most sophisticated school type (*Gym*) while 45 percent visit the lower tier school types (*WRS and RS*) and 16 percent visit a comprehensive school type that combines the traditional tracks into one school. Additional cluster-level variables are *school size* (average school size of 658 students, SD = 222.5), the ratio of students per class who don`t speak German as their primary language in childhood to those with German as their primary language in childhood (*prop.no.native)* (M = 34.56 %) and the degree of urbanization of the area the school is located in *(urbanization)* as discussed above (M = 1.94, SD = 0.81).

## 3. Measuring economic competence

In the following, we describe the underlying conceptual model of economic competence (3.1), provide an overview of the measurement model (3.2) and illustrate methods for dealing with missing values (3.3).

### 3.1. Competence model and item development

We measure economic competence by using an adapted version of a comprehensive performance test for German secondary school students in grades 7–10 developed by Seeber et al. (2018, 2019). The development of such as specific test was necessary, because the curriculum is largely based on a competence model described in Retzmann et al. (2010); Remmele and Seeber (2012), and Retzmann and Seeber (2016) where economic competence is defined as a domain-specific internal capability attributed to the student. Thus, widely used tests such as the "Test of Economic Literacy" (TEL) (Walstad and Rebeck, 2013) or the "Test of Understanding of College Economics" (TUCE) (Walstad and Rebeck, 2008) that are primarily measuring the *knowledge* of core economic concepts were not entirely applicable in our case. Instead, the definition of economic competence in Retzmann et al. (2010); Remmele and Seeber (2012), and Retzmann and Seeber (2016) is based on the concept of "competence" as cognitive and non-cognitive "prerequisites for meeting complex demands" (Weinert 2001) which is also the conceptual definition underlying the OECD's PISA assesments (OECD 2001, 6) (cf. Retzmann and Seeber, 2016, 13).

Concentrating on measurable cognitive components the conceptual model distinguishes three competence areas which are relevant in economically shaped life situations (see Fig. A2 in the Appendix A). The first competence area ("Decision-making and Rationality") focuses on the role of an individual who has to make rational choices among different alternatives. The second area focuses on social interactions, where the student has to consider the interests of others in a responsible way. The third area ("System and Order") focuses on a systemic analysis, which implies, among other components, insight into economic concepts, including the necessity of regulation of markets and the possibilities of changing the rules of regulation (cf. Remmele and Seeber, 2012).

Consequently, test items measuring this latent ability should require students to analyze consequences of economic decisions (*Decision-making and Rationality*), evaluate social interactions (*Relationship and Interaction*) and judge regulatory issues (*System and Order*). We established *content validity* by expert ratings and qualitative think-aloud-studies with selected students. Due to proper psychometric properties and conceptual overlap, we included two items from external sources (Lusardi and Mitchell, 2014; OECD INFE, 2012).

The complete process of items selection, validation, and measurement characteristics of this comprehensive test for students ranging from grade 7–10 are reported in full in Seeber et al. (2018, 2019) and Oberrauch (2019). We report the psychometric (re-) validation of our reduced item set in the following:

Our adapted version of the test uses a subset of 25 single choice and 5 open response items to measure economic competence among 7$^{\text{th}}$ grade students (see Table A2 in the Appendix A). The number of items is distributed evenly among the competence areas (see Table A2). With regard to reliability, Cronbach's alpha (Cronbach, 1951) was 0.81 indicating a consistent measure of economic competences. One key requirement for the subsequent IRT analysis is the unidimensionality of the underlying construct. Principal Component Analysis (PCA) revealed an eigenvalue of 10.42 for the first component, which accounts for 22.17 percent of total variance. The eigenvalue for the second component was merely 2.45 which shows the dominance of the first component and therefore the unidimensionality of our measurement scale.

### 3.2. Measurement model

One disadvantage of using raw item scores in the validation process is that they depend on the examined sample and vice versa ability estimates are dependent on the particular set of test-items administered. Thus, we validated the adapted competence test using probabilistic test theory (Baker and Kim, 2004; Hambleton and Swaminathan, 1985), where manifest variables (item responses) attribute to a latent ability (economic competence) in order to ensure sample-independent and item-independent estimations of the latent trait. The latent abilities and item difficulties are estimated with the One-Parameter-Logistic Model (1-PLM) (Rasch, 1960), which defines the probability a person $v$ of solving an item $i$ correctly as follows:

$$P(X_i = 1|\theta_v) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)} \tag{1}$$

$\theta_v$ denotes the ability of person $v$ and $\sigma_i$ the item difficultiy of item $i$ on a common logit scale of [-4:4]. The more person ability $\theta_v$ exceeds item difficulty $\sigma_i$, the more likely this person solves the item correctly and vice versa. In contrast to factor analysis, item

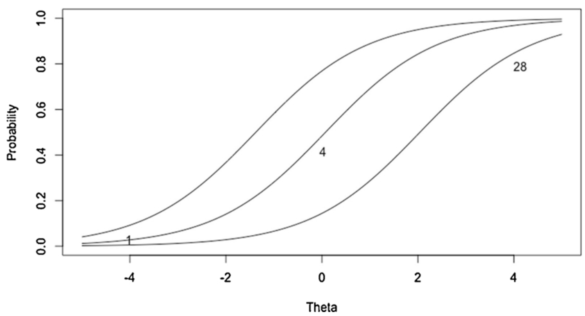
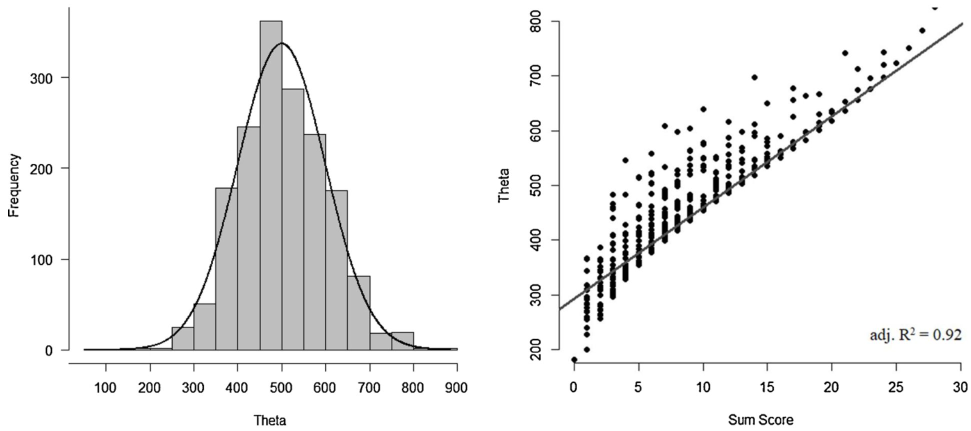



**Fig. 1.** ICCs for items 1, 4 and 28.




**a: Histogram of economic competence** **b: Correlation IRT ability and sum scores**

**Fig. 2.** Distribution of economic competence and correlation between sum scores and ability estimates. (a) Histogram of economic competence. (b) Correlation IRT ability and sum scores.
Notes: Results from 1-PL-IRT estimation. Fig. 2a shows the observed distribution of economic competence and a normal distribution as a reference. Fig. 2b shows the relationship between the estimated ability (economic competence) from the IRT model and the sum of correct responses.

response theory models assume a logistic relationship between a person's ability $\theta$ and solving an item correctly represented graphically by Item Characteristic Curves (ICC). Fig. 1 shows ICCs for three selected items (item1, item 4, and item 28; see also Appendix A Table A2). The difficulty parameter $\sigma$ represents the point on the logit scale where $P(X_i = 1|\theta_v) = 0.5$ applies, i.e., the point on the ability scale where a person $v$ has a greater possibility to score the item than indicated by chance. Since θ is assumed to have a mean of zero by definition, an item $i$ is relatively easy to solve if $\sigma_i < 0$, and an item $i$ is relatively hard to solve if $\sigma_i > 0$. Thus, item 1 is relatively easy, item 4 is of medium difficulty, and item 28 is relatively hard to solve. This interpretation requires the assumption of local independence among items (solving an item must not be conditional upon solving another item) and θ to be unidimensional.

By means of Eq. (2) we obtain person and item parameters by maximization of the likelihood function (MLE) with respect to $\theta$ and $\sigma$

$$\max_{\theta,\sigma} L = \prod_{v=1}^{N} \prod_{i=1}^{M} \frac{\exp(x_{vi}(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)} \tag{2}$$

where $x_{vi}$ denotes a data matrix with all single reactions of $n$ persons on $m$ items. However, MLE estimations tend to show statistical bias (Lord, 1983) which is why we rely on WLE estimators (Weighted Likelihood Estimation; Warm, 1989) as they allow more precise measurements through weighting individual item information.

Fig. 2a and b show the distribution of WLE estimates (economic competence) and their relationship to sum scores. For easier interpretation, we transformed the person ability estimates on the logit scale to have a mean of 500 with a standard deviation of 100 as is the convention in educational large-scale assesments. As the 1-PLM relies on raw scores for ability estimation (*sufficient statistics*), correlation between WLE and sum scores is, in case of model validity, strong (here: $r = .95$, $p < .01$).

Since the discrimination parameter (slope) is constrained to be fixed for all items of the scale, we assess item discrimination for single items by using the item total correlation from classical test theory. The item total correlation is defined as the point-biserial correlation *(rpb)* between sum scores and the observed item response. In general, we excluded items with *rpb* < 0.20 (see Walstad and Rebeck, 2001a,b) from further analysis. However, due to construct relevance some items slightly below our *rpb* threshold remain in the test (see Table A2). We exclude two items with *rpb* close to zero. Column 1 of Table A4 shows regressions estimates without item exclusions.

In order to assess test fairness among subgroups, we implement differential item functioning (DIF) analyses as the occurrence of item bias harms construct validity. One reason for the occurrence of DIF is that one unintentionally measures confounding abilities that are distributed unevenly among subgroups. Within an IRT context, an item exhibits DIF when students with the same ability level have different probabilities of solving the item correctly (Lord, 1980; Thissen et al., 1993). Thus, we scaled subgroups separately and tested the estimated parameters for statistical differences (Dorans and Holland, 1992). As DIF statistics are subject to interpretation difficulties, we follow the common classification from Educational Testing Services (ETS; Zieky, 1993).

The ETS approach transforms the Mantel-Haenszel odds ratio $\alpha_{MH}$ to the logistic definition of the delta scale using the formula specified in Holland and Thayer (1988), $\Delta - DIF_{MH} = -2.35\ln(\alpha_{MH})$. Based on effect size and significance, DIF is then classified into three categories, where $\Delta - DIF_{MH}$ assesses effect size and the Manthel-Haenszel statistic $MH_{\chi^2}$ evaluates significance by following a $\chi^2$ – distribution. If $\Delta - DIF_{MH}$ shows a non-significant, i.e., $MH_{\chi^2} < 3.84$, absolute value of less then 1, Items are allocated to negligible DIF (category A). Items with significant values ($MH_{\chi^2} > 3.84$) in range of $1.0 \leq \Delta - DIF_{MH} < 1.5$ are classified to show moderate DIF (category B). Items were allocated to category C, if $\Delta - DIF_{MH}$ shows significant values at least greater than 1.5. In the course of the validation process, we eliminated all items (n = 2) that fit into category C. Thus, our final scale and the economic competence estimates are based on a total of 30 items.

### 3.3. Missing data and plausible values

Since test participation is voluntary and not factored into the grade at school, the analysis of missing values reflects *item* non-response (reasons for *unit* non-response are not observable). We exclude those participants who show more than 50 percent item non-response from further analysis as the first half of the questionnaire consists mainly of demographic, self-assessment and attitudes questions. We deal with the remaining missing values by implementing various procedures based on variable type. In educational large-scale assessments omissions in the competence test could either treated as "not available" (missing) or "wrong" (0). Both approaches appear problematic: In the first case non-responses are treated as ignorable, i.e. participants had no opportunity to solve the item comparable to a missing by design in adaptive testing procedures. Application of this method yields to overestimation of competence since there could be several other reasons for non-response. The second approach assumes that participants did not know the correct answer and violates IRT model assumptions as every missing is determined to zero in advance. As statistical bias tends to be more severe in the second approach (e.g. Rose et al., 2010), this study treats omissions as missing. Appendix A Table A4 shows regression results for the second approach as a robustness exercise.

We dealt with item non-response in the covariate variables differently, since they are likely to correlate with the dependent variable and may therefore bias means and variances. The analysis of missing values involves examinations of missing occurrences and their effects on the dependent variable. Appendix A Table A3 shows fractions of missing data for variables with at least one missing value. The second column shows standardized mean differences (*Cohens' d*) of person abilities between groups with complete data and missing data.

Questions about migration status, mother tongue and parental educational background show over-proportionate fractions of missing data. Effect sizes that are different from 0 indicate a systematic relationship between missing values and competence measures, e.g., participants with missing data in the math self-assessment variable show a 0.92 standard deviations lower competence measure than participants with complete data for this variable. Thus, ignoring these missing data in regression analyses leads to a biased result.

To ensure proper statistical inference, we, thus, handle missing data in covariates with multiple imputation by chained equations (*MICE,* van Buuren and Oudshoorn, 2000) that allows conducting the imputation process for each variable separately. For each predictor variable $X_i$ with $i = 0,..,n$ we define a imputation model based on the conditional distribution $P(X_i \mid X_{-i})$. The imputation process is then conducted iteratively for all variables (see van Buuren, 2007). As imputation methods, we used logistic regression for binary variables and predictive mean matching (PMM) for numeric variables. The latter replaced conventional linear regression as PMM allows for relaxing homoscedasticity and normal distribution assumptions.

Next, based on the reasonable assumption that estimated competence measures from IRT modeling are subject to significant measurement error (i.e., $\hat{\theta} = \theta + \varepsilon$), ignoring such measurement error would lead to biased regression estimates. During recent years, a multiple imputation method for error term correction called "Plausible Values (PV)" (Marsman et al., 2016) has been established. The PV method regards the unknown variances of the estimated person ability as missing values and replaces them with multiple imputed values. Latent competence measures and covariates serve as predictor variables in this multiple imputation process in terms of a latent regression model.

Since demographic covariates are naturally correlated, we incorporated their interaction effects into our latent regression model. With high amounts of covariates, however, the imputation model tends to be unstable and therefore makes dimension reduction necessary. As a reduction method we choose partial least squares (PLS) with stepwise extraction of uncorrelated factors based on maximal covariance. We regressed the covariate matrix with all two-way-interactions on the competence values. Analysis showed that variance information remains stable after 15 PLS factors. Moreover, we extracted individual likelihood from person ability estimates that together with 15 covariate PLS factors specify the latent regression model from which 20 plausible values for each participant are drawn.

## 4. Regression models

We analyze determinants of economic competences by means of several regression methods. In order to accommodate the clustered sampling structure we extend the basic OLS regression model to a hierarchical model (Raudenbush and Bryk, 2002) Therefore, we allow intercepts to vary across schools by employing a the random intercept model. The model takes the form

$$\theta_{i,j}^{PV} = \beta_{0,j} + \beta_{i,j} X_{i,j} + \varepsilon_{i.j}^{*} \tag{3}$$

where $\beta_{0,j}$ represents a composition of mean competence value $\gamma_{00}$ and group dependent deviation $u_{0,j}$,

$$\beta_{0,j} = \gamma_{00} + u_{0,j} \tag{4}$$

$X_{i,j}$ is a covariate vector at the individual and school levels. For *schools* $j = 1, \ldots, J$ and student $i \neq k$, $\varepsilon_{i,j}^{*}$ represents cluster-robust standard errors with

$$E(\varepsilon_{i,j}, \varepsilon_{k,j'}) = \begin{cases} 0 \; if \; j = j' \\ \sigma_{(ik)j} \; if \; j \neq j' \end{cases} \tag{5}$$

**Table 2**
Estimates for the Random Intercept Model and Quantile Regression.

| Dependent Variable: Economic Competences (Plausible Values) | Random Intercept | Quantile Regression | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) 20th percentile | (3) 40th percentile | (4) 50th percentile | (5) 60th percentile | (6) 80th percentile |
| *male* | 14.840*** | 7.980 | 11.016** | 11.370** | 14.390*** | 23.259*** |
| | (3.92) | (5.40) | (4.80) | (5.14) | (5.02) | (6.73) |
| *age* | −7.474** | −6.710* | −8.374** | −8.453** | −7.376* | −8.565* |
| | (3.47) | (3.58) | (3.30) | (3.93) | (4.13) | (5.06) |
| *migrant* | −22.229*** | −28.199*** | −27.966*** | −28.315*** | −25.005*** | −18.169** |
| | (4.86) | (6.88) | (5.19) | (6.62) | (6.13) | (7.65) |
| *book* | 5.539*** | 3.243 | 4.520*** | 3.360* | 4.586** | 9.706*** |
| | (1.59) | (2.23) | (1.73) | (1.93) | (2.18) | (2.52) |
| *GYM* | 65.514*** | 84.377*** | 83.951*** | 88.156*** | 85.777*** | 80.104*** |
| | (13.03) | (10.54) | (9.02) | (9.88) | (10.54) | (11.74) |
| *RS* | 19.106* | 39.074*** | 30.417*** | 31.619*** | 27.408*** | 32.158*** |
| | (11.46) | (8.38) | (7.56) | (8.33) | (9.30) | (10.17) |
| *WS* | −8.676 | 15.558 | 6.362 | 6.069 | −0.562 | −6.088 |
| | (11.97) | (10.17) | (8.68) | (9.09) | (9.68) | (11.05) |
| *selfmath* | 10.200*** | 6.586** | 10.000*** | 10.398*** | 10.037*** | 9.422** |
| | (3.07) | (2.98) | (3.26) | (3.72) | (3.61) | (4.22) |
| *selfread* | 7.933** | 4.528 | 8.574** | 11.846*** | 10.867*** | 15.805*** |
| | (3.52) | (4.43) | (4.02) | (3.81) | (3.97) | (5.20) |
| *selfall* | −1.112 | 0.943 | 2.073 | 0.274 | −1.647 | 4.537 |
| | (4.19) | (4.96) | (4.05) | (4.92) | (5.13) | (6.15) |
| *interesteco* | 7.861** | 5.305 | 4.785 | 4.394 | 4.311 | 3.748 |
| | (3.63) | (4.70) | (3.87) | (4.09) | (3.90) | (5.52) |
| *interestimp* | 8.687** | 7.925 | 8.995** | 9.590** | 10.950** | 7.392 |
| | (3.43) | (4.89) | (3.83) | (4.13) | (4.48) | (5.24) |
| *urbanization* | 2.714 | 3.602 | 4.164 | 3.333 | 6.305* | 5.535 |
| | (3.79) | (3.70) | (2.69) | (3.18) | (3.37) | (4.09) |
| *school size* | −0.007 | −0.006 | −0.012 | −0.017 | −0.019 | −0.010 |
| | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) |
| *% non-natives/class* | −0.453*** | −0.344** | −0.234* | −0.270** | −0.316** | −0.422** |
| | (0.16) | (0.17) | (0.13) | (0.14) | (0.14) | (0.20) |
| *bank.exp* | −2.182 | 1.759 | 1.296 | 3.262 | 3.109 | −1.840 |
| | (2.88) | (3.45) | (2.90) | (3.38) | (3.66) | (4.34) |
| *time* | 2.001*** | 2.844*** | 2.434*** | 2.409*** | 2.131*** | 1.543*** |
| | (0.37) | (0.38) | (0.36) | (0.38) | (0.38) | (0.45) |
| *ownsalary* | −5.374 | 0.451 | −1.962 | −5.262 | −5.421 | −7.756 |
| | (5.19) | (5.68) | (5.02) | (5.57) | (5.83) | (7.47) |
| *constant* | 473.116*** | 395.182*** | 441.114*** | 460.152*** | 479.376*** | 524.791*** |
| | (10.53) | (9.40) | (8.86) | (9.17) | (9.28) | (11.61) |
| N | 1687 | 1687 | 1687 | 1687 | 1687 | 1687 |
| R-squ. Lev 1 | 0.77 | | | | | |
| R-squ. Lev 2 | 0.30 | | | | | |

*Note*: This table shows results for the Random Intercept Model specified in Eq. (2) and for Quantile Regression as discussed in Section 4. Covariates are drawn from 20 multiple imputations; dependent variable Economic Competences is multiple imputed from latent regression (Plausible Values); The standard errors for quantile regression estimates (column 2–6) are bootstrapped with 1,000 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in which we assume independence across schools and correlation within schools.

As model (3) only accounts for the average relationship between covariate matrices and our dependent variable based on conditional mean function $E(y|\ x)$, we are also interested in describing effects on different moments of the conditional distribution of economic competences. Therefore, we implement simultaneous quantile regression on the conditional distribution moments M (0.2, 0.4, 0.5, 0.6, 0.8), i.e. the 20th, 40th, 50th, 60th, and 80th percentile, allowing for richer analysis of the data by examination of effect sizes on the entire distribution. Estimation follows the minimization problem specified in Koenker and Bassett (1978). Median regression at moment M = 0.5 is included as it accounts properly for outlier robustness and may serve as an important reference point.

## 5. Results

### 5.1. Random intercept model

Table 2 shows results for the random intercept model specified in Eq. (1) while other models are discussed in robustness section Tables A3 and A4 in the Appendix A. Quantile regression estimates are shown in columns 2–6 of Table 2. The intercept-only model predicts a comparatively strong intra-cluster correlation (0.27). All non-categorial variables are mean-centered to ease the interpretation of the estimated intercepts. With regard to multicollinearity, we analyzed correlations between predictor variables. Only the variables mig and native show a strong correlation above 0.8, which is why we excluded the latter from regression analysis. All other variables show correlations below 0.5. Multiple regression with the complete variable set showed merely a maximum variance inflation factor (VIF) of 3.4 (mean VIF of 1.5). In order to account for residual collinearity, we add variables in blocks building on the four main demographic variables.

Model (1) shows a moderate and slightly significant gender gap in favor of male participants after controlling for survey duration (*time*), which corresponds with findings from an antecedent cross-sectional study in Baden-Württemberg (Seeber et al., 2018). Next, a one-year increase in age lowers competence levels by 7.47 points after controlling for the complete set of variables as younger students within one grade consist partially of repeaters and children of migrants with lower German-language skills. The effect size is inter alia confounded by school variables since students visiting a less sophisticated school type (*WS* and *GMS*) are older, on average, by (self-assessed) mathematical and reading abilities and by survey duration (*time*). Migration status (*mig*) is associated with a lower competence level by 22.23 points and remains relatively constant across different combinations of regression models. We complement the analysis of migration background by exchanging this variable with the variable *native* (primary language spoken at home) using two different definitions: In one specification we code *bilingual* students as native speakers and in one specification we code them as non-native speakers. Using these definitions of migration background, we find that the score difference increases to 25.3 points (bilingual = non-native) and 26.5 points (bilingual = native) respectively (available on request). Thus, differences in competence may be driven by differences in reading comprehension resulting from the fact that the primary language at home is different from the main language at school. On cluster level, an increase of one percentage-point of non-native speakers per class leads to lower competence level by 0.45 points.

The variable *book* is confounded with the multilevel structure as intra-class correlation in a one-way ANOVA with the dependent variable book reaches 0.24 (see also OLS-Regression - Model 2 in Table A4). The analysis of school type affiliation (tracks) relative to the comprehensive school including students of all ability levels (*Gemeinschaftsschule*) shows strongest overall effects for participants visiting a *Gymnasium* (GYM) followed by participants visiting a *Realschule* (RS) with a net effect of 19.11 on the competence scale. The negative effect of affiliation to type of school *Werkrealschule* (WS) becomes insignificant after controlling for variables capturing attitudes towards economics.

Self-assessment of mathematical ability and literacy show small but significant effects while the self-assessment of overall school performance has no significant influence. With regard to test *time*, a one-minute increase in total test duration is associated with an increase of 2 units on the competence scale. This effect can be interpreted as the returns to more effort on the test.

Overall, the variables school type (*GYM, RS, GMS* compared to *WRS*), *male, book, mig, age,* and *time* are most predictive of the outcomes. From a qualitative perspective, the results on gender differences (*male*), parental socioeconomic status (*book*) and migration background (*mig*) appear to be most important.

### 5.2. Quantile regression

In an additional step we conduct a quantile regression as described in chapter 4. Results for the complete models are shown in Table 2 (Columns 2–6).

The effect of the variable *male* locates within the OLS 95 % confidence intervals along all distribution moments and shows – relative to the OLS point estimates - a weaker gender gap at the first four distribution moments. At the $80^{th}$ percentile, however, male participants show a higher test result by 23.25 points (0.23 SD units) ($p < .05$).

These results correspond with results from PISA where gender gaps in favor of male students are even wider among high performers in the mathematics domain (OECD, 2013). Migration status shows stronger (negative) effects in lower moments at the distribution. Students with migration background (*mig*) at the $20^{th}$ percentile achieve a lower test result by 28.92 points (-0.29 SD units) ($p < .01$) while the effect shrinks to -18.00 points ($p < .05$) among high performers at the $80^{th}$ percentile. The importance of educational background of parents (*book*) remains within OLS confidence interval along the first distribution moments and only exceeds the OLS regression estimate at the $80^{th}$ percentile with an effect size of 9.71 points ($p < .01$) while all remaining distribution moments show smaller effects than the mean regression point estimate. In summary, gender differences are more pronounced among top performing students, differences between students with migration background and students with parents born in Germany are smaller at the higher moments of the competence distribution, and test results at the top of the competence distribution are more strongly dependent on parents' socio-economic status (*book*) than at the lower moments of the distribution.

### 5.3. Robustness exercises

As robustness exercises we present results from OLS regressions (with standard errors clustered at the school-level) as an alternative modelling approach in Table A4 as well as other specifications for the dependent variable Economic Competences. Results of the OLS estimations are near identical. Additionally, to account for uncertainty arising from model selection, we complement these analyses by conducting a LASSO regression. Table A5 (see Appendix A) shows CP-Values, $R^2$ and estimated coefficients along the sequence of models. Variables that enter the model early, i.e., at low shrinkage values, are the most predictive in the model. Results correspond widely with results from OLS regression. All significant OLS predictors enter the Lasso model relatively early. The strongest predictor is affiliation to type of school "*Gymnasium*" followed by *time*, SES of parents represented by the variable *book,* migration background *mig* and *age*. Non-significant OLS predictors either are excluded from the model (*selfall* or *bank_exp*) or enter the model comparatively late (*ownsalary* or *urbanization*). Thus, we conclude that our main results are insensitive to model selection.

## 6. Discussion

This paper has explored the economic competence of early secondary school students in absence of prior formal instruction in the economic domain. Using a psychometrically sound measure of economic competence, we have identified important heterogeneities in economic competence across school types and student characteristics. Our results on student-characteristics generally mirror the general from large-scale assessments in Germany in other domains such as mathematics achievement in PISA (see OECD, 2016): Female students perform worse than male students on our test with the gender gap increasing along the distribution of competence. The size of the average gender gap (0.15 SD units) is very similar to the observed gender gap in mathematics achievement (0.17 SD units) in Germany documented in PISA 2015 among 15 year old-students (mean age in our sample is 14 years) or to the male-female difference in mathematics at the end of fifth grade in the U.S. (cf. Fryer et al., 2010). The magnitude of the gender gap is also much larger than the gender gap in financial literacy: it corresponds to about two to two and a half times the average (conditional) gender gap across 13 OECD countries participating in the PISA financial literacy assessment 2012 (OECD, 2014, p. 80). As the employed economic competence test accounts for the possibility of (uniform) differential item functioning (DIF) with regard to gender (see Walstad and Robson, 1997), this finding is not the result of measurement error or differential item (format) responses. This result is generally in line with the literature on economic literacy in U.S. and U.K. high schools (e.g. Heath, 1989; Davies et al., 2005; Walstad, 2013). While the reasons for this gender gap have not been fully understood in the literature, yet, most studies point at a combination of cultural factors (such as gender specific socialization and differential parental expectations) and stereotypes (see also Grohmann, 2016; Driva et al., 2016). To get a qualitative sense of the magnitude of the gender gap one has to compare this result to either naturally occurring gains in competence over time or, alternatively, to treatment effects realized by interventions (e.g. increased instructional time) in the same domain. Unfortunately, data on natural progress in achievement in the economic domain is scarce. Thus, we compare this result to observations from adjacent domains.

Comparing effect sizes across educational domains is inherently difficult. However, empirical and normative benchmarks (from the U.S.) suggest that this difference may be equivalent to the gain in (standardized test scores on) reading achievement in the transition from grade nine to ten, ten to eleven, or the gain in mathematics achievement from grade ten to eleven in absence of any interventions (Hill et al., 2008, p.173).

Could an increase in formal instruction in high school economic education close this gap? While there is no systematic review on the effect of *economic* education on increases in economic literacy, knowledge or competence, a meta-analysis of a large body of work on personal *financial* education in schools exists (see Kaiser and Menkhoff, 2018). Comparing the size of the gender gap to the weighted average effect reported in 14 randomized experiments (RCTs) and 17 quasi experiments shows that the mean difference is equivalent to twenty hours of instruction (at average delay in measurement, class size, and student characteristics) (cf. Kaiser and Menkhoff, 2018, p. 27). Thus, the gender gap is quite substantial in size and amounts to the intensity of two thirds of a school year in instruction of the newly introduced economic education mandate in Baden-Württemberg for the following cohort (economics is taught as one 45-minute lesson per week).

Next, we document that thirteen percent of the variance in economic competence is explained by the socio-economic status of students' parents (proxied by the variable *book*). In our context, a one standard deviation increase in the variable *book* is associated with a 5 point 5 increase in economic competence (0.06SD). Thus, we find support that differences in parents education (socio-economic status) may be important in understanding the existing heterogeneities in economic competence in absence of formal instruction. This result is very similar to results from the 2012 (2015) PISA financial literacy assessment where fourteen (and ten) percent of the within-country variance in financial literacy scores can be attributed to parents' socio-economic status (see OECD, 2014, 2017). Thus, our research confirms the importance of parents as socialization agents (cf. Grohmann et al., 2015): Achievement in the economic domain is strongly dependent on parental influence, thus highlighting the need to offer assistance and tailored instruction to children of parents who do not actively discuss economic matters at home.

Additionally, we provide evidence that students with an migration background perform worse (-0.22 SD units) on the competence test than those whose parents have been born in Germany. These findings correspond to results from the 2012 and 2015 PISA financial literacy assessment where the (conditional) difference between students with migration background and those with parents born in the respective country of assessment amounts to about 0.19 to 0.26 SD units (OECD, 2014, 2017). The effect of the parents' SES are more pronounced at the higher moments of the distribution. The effect of being a child of migrant parents is smaller at the top of the competence distribution (0.18 SDs at the 80 th percentile) relative to the effect at the median (−0.27 SDs) or the 20 th percentile (−0.30 SDs). Similar to Happ and Förster (2018), we investigate the source of this difference in test performance and find that the effect of migration background increases when operationalized via the primary language spoken at home (instead of the place of birth of one parent). Our results are in line with the observation that differences in competence may be driven by differences in reading comprehension resulting from the fact that the primary language at home is different from the language used at school (cf. Cameron et al., 2014; Happ and Förster, 2018).

The analyses presented in our paper serve as important baseline information regarding the future impact evaluation of mandatory economic education introduced with the recent curriculum reform in Baden-Württemberg. Several large-scale evaluations of economic (and financial) education curricula have shown beneficial effects on knowledge and field behaviors (e.g. Bruhn et al., 2016; Frisancho, 2018). Our studied sample belongs to the last cohort without mandatory economic education and will continue without economics as a school subject throughout secondary school. Our large-scale longitudinal assessment will follow students of this cohort and the next cohort (with mandatory economics education) until the end of 10$^{th}$ grade to arrive at difference in difference estimates of the economic education impacts. The determinants of ex-ante levels of economic competence, thus, serve as important variables regarding the investigation of potential heterogenous effects in response to economic education. We will be particularly interested in the effects of mandatory economic education on the gap between males and females, the difference between students with migration background to those without a history of migration and the effects on students with parents of lower socio-economic status.

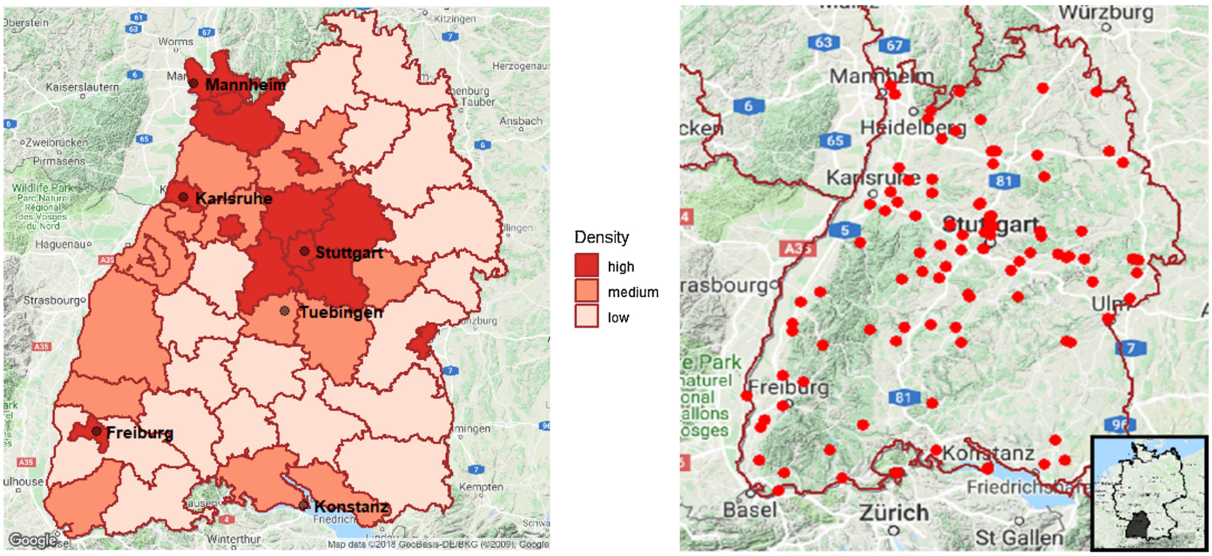
**Funding**

This work was supported by *Stiftung Würth*.

**Acknowledgments**

We would like to thank the schools and students participating in our study. Additionally, we appreciate comments by the editor (Edmund Cannon), three anonymous referees, and seminar participants in Hamburg, Freiburg, Tübingen, and St. Louis, especially Taiga Brahm, Günther Seeber, Miguel Luzuriaga, and Josh Staveley-O'Carroll, as well as research assistance by Ngoc Anh Nguyen and Lena Masanneck.

**Appendix A**




**a: Population density in Baden-Wuerttemberg** **b: Spatial distribution of sampled schools**

**Fig. A1.** Population density map and spatial distribution of sampled schools. (a) Population density in Baden-Wuerttemberg. (b) Spatial distribution of sampled schools.
Notes: A population density of below 220 residents per square kilometer is defined as a "low degree of urbanization", between 220 and 485 residents per square kilometer is defined as "medium degree of urbanization" and above 485 residents per square kilometer is defined as a "high degree of urbanization".

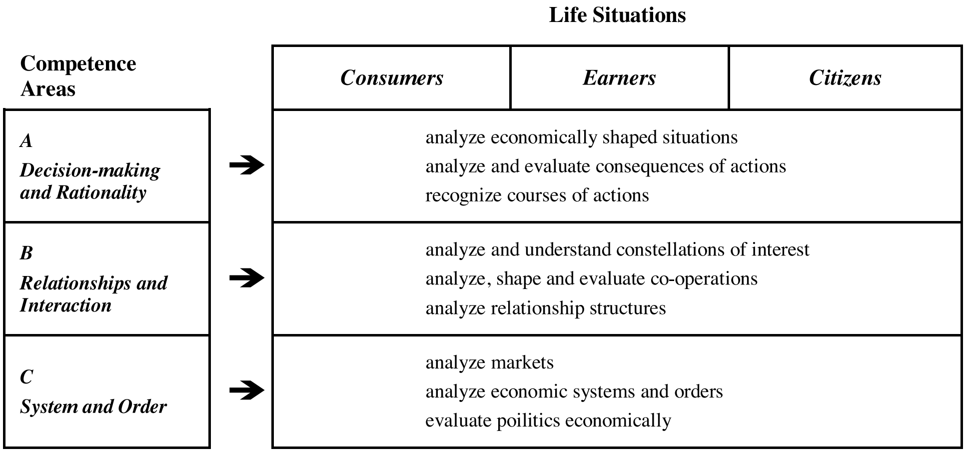



**Fig. A2.** Competence Model for economic education (cf. Retzmann and Seeber, 2016).

**Table A1**
Sampling proportions.

| Type of school | Degree of urbanization | Stratum | Students in Stratum | Percentage | Clusters | Students (sampled) | Percentage |
|---|---|---|---|---|---|---|---|
| GYM | high | 1 | 15742 | 15.92 | 606 | 213 | 12.54 |
| | medium | 2 | 9867 | 9.98 | 385 | 249 | 14.66 |
| | low | 3 | 8993 | 9.09 | 362 | 200 | 11.78 |
| RS | high | 4 | 10961 | 11.08 | 439 | 109 | 6.42 |
| | medium | 5 | 10193 | 10.31 | 425 | 165 | 9.72 |
| | low | 6 | 11958 | 12.09 | 507 | 191 | 11.25 |
| GMS | high | 7 | 5176 | 5.23 | 241 | 127 | 7.48 |
| | medium | 8 | 5592 | 5.65 | 259 | 116 | 6.83 |
| | low | 9 | 5435 | 5.50 | 267 | 73 | 4.30 |
| WRS | high | 10 | 3304 | 3.34 | 172 | 57 | 3.36 |
| | medium | 11 | 5478 | 5.54 | 271 | 53 | 3.12 |
| | low | 12 | 6190 | 6.26 | 304 | 145 | 8.54 |

**Table A2**
Item characteristics for the employed measurement scale of economic competence.

| | | | | | CTT | | IRT | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Source | Competence area | Type | n | RelFreq | rit_c | sigma | [SE] | Infit | Outfit |
| 1 | Seeber et al. (2018) | A | SR | 1689 | 0.738 | 0.258 | −1.103 | [0.058] | 1.006 | 0.987 |
| 2 | Oberrauch (2019) | A | SR | 1619 | 0.634 | 0.164 | −0.581 | [0.055] | 1.064 | 1.116 |
| 3 | Lusardi and Mitchell (2014) | A | SR | 1556 | 0.638 | 0.223 | −0.544 | [0.056] | 1.056 | 1.040 |
| 4 | Seeber et al. (2018) | B | SR | 1554 | 0.641 | 0.257 | −0.536 | [0.056] | 1.024 | 1.028 |
| 5 | Seeber et al. (2018) | A | SR | 1560 | 0.624 | 0.345 | −0.465 | [0.055] | 0.970 | 0.945 |
| 6 | Seeber et al. (2018) | A | SR | 1550 | 0.611 | 0.383 | −0.377 | [0.055] | 0.971 | 0.943 |
| 7 | Seeber et al. (2018) | A | SR | 1587 | 0.595 | 0.355 | −0.338 | [0.054] | 0.978 | 0.965 |
| 8 | Seeber et al. (2018) | B | SR | 1565 | 0.591 | 0.379 | −0.306 | [0.055] | 0.941 | 0.917 |
| 9 | Oberrauch (2019) | C | SR | 1584 | 0.569 | 0.397 | −0.171 | [0.054] | 0.943 | 0.920 |
| 10 | Oberrauch (2019) | A | SR | 1558 | 0.562 | 0.405 | −0.144 | [0.054] | 0.939 | 0.922 |
| 11 | Seeber et al. (2018) | C | SR | 1566 | 0.557 | 0.365 | −0.137 | [0.054] | 0.971 | 0.952 |
| 12 | Oberrauch (2019) | A | SR | 1620 | 0.550 | 0.376 | −0.114 | [0.053] | 0.972 | 0.954 |
| 13 | Oberrauch (2019) | A | SR | 1570 | 0.520 | 0.313 | −0.006 | [0.054] | 0.994 | 0.982 |
| 14 | Seeber et al. (2018) | C | SR | 1613 | 0.534 | 0.409 | 0.009 | [0.053] | 0.921 | 0.902 |
| 15 | Seeber et al. (2018) | A | SR | 1614 | 0.524 | 0.383 | 0.017 | [0.053] | 0.938 | 0.918 |
| 16 | OECD (2012) | A | CR | 1628 | 0.518 | 0.371 | 0.022 | [0.053] | 0.960 | 0.946 |
| 17 | Seeber et al. (2018) | A | SR | 1602 | 0.490 | 0.337 | 0.171 | [0.054] | 0.985 | 0.975 |
| 18 | Seeber et al. (2018) | C | SR | 1606 | 0.366 | 0.174 | 0.723 | [0.056] | 1.082 | 1.117 |
| 19 | Seeber et al. (2018) | B | CR | 1629 | 0.338 | 0.411 | 0.899 | [0.057] | 0.926 | 0.887 |
| 20 | Oberrauch (2019) | C | CR | 1571 | 0.327 | 0.399 | 0.981 | [0.059] | 0.931 | 0.886 |
| 21 | Seeber et al. (2018) | C | SR | 1619 | 0.289 | 0.371 | 1.121 | [0.059] | 0.946 | 0.918 |
| 22 | Seeber et al. (2018) | B | SR | 1623 | 0.290 | 0.315 | 1.161 | [0.06] | 0.985 | 0.986 |
| 23 | Seeber et al. (2018) | C | SR | 1628 | 0.284 | 0.207 | 1.181 | [0.06] | 1.056 | 1.104 |

(continued on next page)

**Table A2** (*continued*)

| | Source | Competence area | Type | n | CTT | | IRT | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | RelFreq | rit_c | sigma | [SE] | Infit | Outfit |
| 24 | Oberrauch (2019) | B | SR | 1629 | 0.257 | 0.37 | 1.313 | [0.061] | 0.945 | 0.898 |
| 25 | Seeber et al. (2018) | C | CR | 1609 | 0.251 | 0.101 | 1.326 | [0.062] | 1.107 | 1.198 |
| 26 | Seeber et al. (2018) | C | SR | 1618 | 0.224 | 0.174 | 1.496 | [0.064] | 1.068 | 1.201 |
| 27 | Seeber et al. (2018) | B | SR | 1614 | 0.206 | 0.344 | 1.648 | [0.067] | 0.956 | 0.912 |
| 28 | Seeber et al. (2018) | B | SR | 1628 | 0.176 | 0.212 | 1.799 | [0.069] | 1.030 | 1.087 |
| 29 | Oberrauch (2019) | C | CR | 1611 | 0.144 | 0.349 | 2.214 | [0.079] | 0.939 | 0.788 |
| 30 | Seeber et al. (2018) | B | SR | 1634 | 0.102 | 0.132 | 2.456 | [0.085] | 1.036 | 1.348 |

*Notes*: IRT analysis is conducted by means of the One-parametric IRT-Model specified in Eq. (1) in the main text. Validation criteria are based on item discrimination represented by corrected item total correlation from classical test theory (rit_c) and model fit by usage of weighted and summated mean-square residuals (INFIT) and unweighted FIT-indices (OUTFIT), where 1 indicates perfect item fit. Sigma denotes item difficulty on the logit scale where P(X = 1) = 0.5 applies (see Chapter 3.2). Column 3 classifies items into constructed response (CR) and selected response (SR, i.e. multiple choice). Scale reliability (Cronbach's alpha) from classical test theory is 0.81.

**Table A3**
Missing proportions and Cohens'd.

| Variable | Missing fraction | *Cohen's d* |
|---|---|---|
| *age* | 0.08 | −0.14 |
| *mig* | 0.08 | −0.12 |
| *native* | 0.08 | −0.12 |
| *book* | 0.09 | −0.18 |
| *selfmath* | 0.1 | −0.92 |
| *selfall* | 0.01 | −0.39 |
| *interesteco* | 0.02 | −0.37 |
| *interestimp* | 0.02 | 0.00 |
| *bankaccount* | 0.1 | −0.13 |
| *bankEC* | 0.1 | −0.18 |
| *ownsalary* | 0.09 | −0.15 |

*Notes*: Remaining covariates (see Table 1) are fully observed; Cohen's d effect size is expressed in standard deviation units and shows standardized mean difference between the group with missing values for this variable and group with fully observed data for each variable.

**Table A4**
Robustness exercises.

| *Dependent Variable: Economic Competencies* | | | | | |
|---|---|---|---|---|---|
| | (1) Random Intercept with NA = 0 | (2) Random Intercept (all items) | (3) OLS with clustered SEs | (4) School fixed Effects | (5) Random Intercept Sumscores |
| *male* | 17.985*** | 11.218*** | 15.183*** | 11.580** | 0.152*** |
| | (5.05) | (3.67) | (4.28) | (4.47) | (0.04) |
| *age* | −10.098*** | −8.258** | −7.645** | −8.439*** | −0.081*** |
| | (3.46) | (3.26) | (3.19) | (3.11) | (0.03) |
| *migrant* | −21.737*** | −23.278*** | −21.501*** | −24.747*** | −0.222*** |
| | (4.68) | (5.14) | (5.30) | (4.895) | (0.04) |
| *book* | 3.595** | 4.818*** | 5.536*** | 6.535*** | 0.040*** |
| | (1.70) | (1.56) | (1.67) | (1.56) | (0.01) |
| *GYM* | 83.153*** | 64.458*** | 67.381*** | | 0.925*** |
| | (8.94) | (12.66) | (8.91) | | (0.11) |
| *RS* | 31.822*** | 20.677* | 18.781** | | 0.357*** |
| | (7.98) | (11.38) | (7.56) | | (0.10) |
| *WS* | −6.261 | −4.786 | −10.471 | | −0.007 |
| | (10.49) | (11.91) | (8.47) | | (0.10) |
| *selfmath* | 7.214*** | 10.023*** | 10.466*** | 10.939*** | 0.074*** |
| | (2.72) | (3.21) | (2.97) | (2.77) | (0.03) |
| *selfread* | 10.279*** | 8.396** | 6.592* | 14.678*** | 0.146*** |
| | (3.26) | (3.36) | (3.64) | (3.35) | (0.03) |
| *selfall* | 3.005 | −0.519 | −1.153 | −4.429 | −0.003 |

(*continued on next page*)

**Table A4** (*continued*)

*Dependent Variable: Economic Competencies*

| | (1) Random Intercept with NA = 0 | (2) Random Intercept (all items) | (3) OLS with clustered SEs | (4) School fixed Effects | (5) Random Intercept Sumscores |
|---|---|---|---|---|---|
| | (4.53) | (3.91) | (4.09) | (3.90) | (0.04) |
| *interesteco* | 8.167** | 7.271** | 7.605** | 7.440** | 0.081*** |
| | (3.46) | (3.29) | (3.55) | (3.47) | (0.03) |
| *interestimp* | 5.403* | 8.556*** | 10.164*** | 8.602** | 0.036 |
| | (3.00) | (3.28) | (3.42) | (3.50) | (0.03) |
| *urbanization* | 1.819 | 3.738 | 1.983 | | 0.009 |
| | (3.79) | (3.89) | (2.77) | | (0.04) |
| *schoolsize* | −0.023 | −0.004 | −0.010 | | −0.000 |
| | (0.01) | (0.02) | (0.01) | | (0.00) |
| *%non-natives/class* | −0.208 | −0.509*** | −0.421*** | | −0.001 |
| | (0.14) | (0.17) | (0.13) | | (0.00) |
| *bank.exp* | −2.059 | −2.184 | −1.410 | −1.217 | −0.015 |
| | (2.80) | (2.75) | (2.83) | (2.70) | (0.02) |
| *time* | 3.358*** | 2.178*** | 1.849*** | 2.891*** | 0.033*** |
| | (0.49) | (0.39) | (0.30) | (0.31) | (0.00) |
| *ownsalary* | −1.690 | −5.793 | −7.190 | −0.823 | −0.042 |
| | (5.18) | (5.33) | (5.01) | (4.60) | (0.04) |
| *constant* | 457.832*** | 474.983*** | 473.662*** | 503.258*** | −0.445*** |
| | (8.38) | (10.50) | (8.06) | (30.67) | (0.10) |
| *N* | 1687 | 1687 | 1687 | 1433 | 1687 |
| *r-squ. (total)* | | | 0.39 | 0.287 | |
| *r-squ. (level 1)* | 0.210 | 0.170 | | | 0.210 |
| *r-squ. (level 2)* | 0.901 | 0.894 | | | 0.912 |

*Note*: This table shows robustness exercises with alternative specifications for the dependent variable and different estimation approaches (θ). In Column (1) all omissions are coded as wrong ( = 0). The dependent variable in Column (2) contains all available test items, i.e. items with item-total correlation close to zero or with severe DIF are included. Column (3) shows OLS estimated with clustered standard errors. Column (4) shows results of a school/class fixed-effects regression with exclusion of 'between'cluster variation. Column (5) shows the benchmark model results with (z-standardized) sum scores as the dependent variable. Note that Model (4) doesn't contain imputations as described in Section 3.3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table A5**
Lasso Regression – CP, R-squared, actions along the sequence of models and coefficients for minimum CP and estimated coefficients.

| Step | CP | $R^2$ | Action (adding variables) | Coefficient |
|---|---|---|---|---|
| 1 | 1064.4342 | 0.0000 | | |
| 2 | 807.1438 | 0.1039 | +C_GYM | 90.74 |
| 3 | 721.3954 | 0.1391 | +time_mc | 3.28 |
| 4 | 414.1109 | 0.2630 | +C_book_mc | 5.40 |
| 5 | 357.0209 | 0.2867 | +C_mig | −22.09 |
| 6 | 346.4903 | 0.2917 | +C_age_mc | −10.21 |
| 7 | 336.2391 | 0.2966 | +C_selfread_mc | 12.89 |
| 8 | 319.9682 | 0.3039 | +C_interestimp_mc | 7.92 |
| 9 | 226.9407 | 0.3420 | +C_selfmath_mc | 8.95 |
| 10 | 150.3639 | 0.3735 | +C_interesteco_mc | 6.13 |
| 11 | 130.7694 | 0.3822 | +C_WRS | −1.16 |
| 12 | 74.8562 | 0.4054 | +C_RS | 33.78 |
| 13 | 62.2315 | 0.4112 | +C_sex -C_WS | 11.28 |
| 14 | 48.7887 | 0.4174 | | |
| 15 | 25.4815 | 0.4276 | +S_schoolsize_mc | −0.03 |
| 16 | 21.8613 | 0.4298 | +C_WS | −1.16 |
| 17 | 17.8106 | 0.4322 | +urbanization_mc | 0.99 |
| 18 | 18.0636 | 0.4329 | +C_ownsalary | −0.52 |
| 19 | 19.1262 | 0.4333 | +bank_exp | . |
| 20 | 21.0792 | 0.4333 | +C_selfall_mc | . |
| 21 | 21.0000 | 0.4342 | +S_prop_n_native_mc | . |

**Item examples (translated from German)**

A: Item example for competence area “Decision and Rationality”

7) An employer, Angela Zapp, wants to compensate the employees of her company, “Zapp Estate”, with a higher salary.

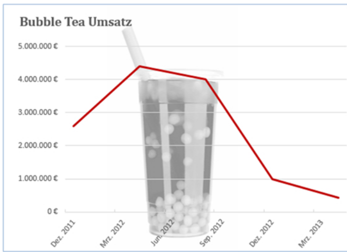
- What is a consequence she can expect?
- A decrease in manufacturing costs.
- A decrease in her profits.*
- A decrease in product quality.
- *A decrease in customer satisfaction.*

B: Item example for competence area “Relationships and Interaction”

28) Two friends, Emil and Kadir, go to the bank. Emil borrows €1,000 from the bank, Kadir deposits €1,000 into his savings account. After one year, Emil wants to pay back the loan, and Kadir wants to withdraw his money.

- Emil has to pay back €1,000. Kadir receives €1,000.
- Emil has to pay back €1,000. Kadir receives more than €1,000.
- Emil has to pay back more than €1,000. Kadir receives €1,000.
- Emil has to pay back more than €1,000. Kadir receives more than €1,000; the amount is the same for both of them.
- Emil has to pay back more than €1,000. Kadir receives more than €1,000; Emil’s amount is higher than Kadir’s. *

C: Item example for competence area “System and Order”




10) The above figure shows how the sales of bubble tea in Germany have developed in the course of 16 months.

What can you conclude from the figure about sales of bubble tea?

- Bubble tea was banned in Germany since August 2012
- Bubble tea continues to be sold profitably in Japan
- Bubble tea is dangerous to health.
- Bubble tea was sold relatively little since August 2012*

**Lasso Regression**

As a further check of robustness and accounting for uncertainty in model selection in the search of meaningful predictors, we choose the regularization method *Lasso (Least absolute shrinkage and selection operator;* Tibshirani, 1996). In this approach the OLS residual square sum is minimized with subject to lasso penalty $L_1 = \sum_{i=1}^{k} |\beta_i|$ smaller than tuning parameter $\vartheta$. The minimization problem follows the equation

$$\min_{\beta} \left[ (\theta_i - X_i\beta_i)^T (\theta_i - X_i\beta_i) \right] w.\ s.\ t.\ \sum_{i=1}^{k} |\beta_i| < \vartheta \tag{7}$$

where $\vartheta$ resembles power of shrinkage. If $\vartheta < \sum_{i=1}^{k} |\beta_i^{OLS}|$ , OLS estimators will be shrunk to zero, i.e. negligible predictors are excluded. Eq. (8) can also be written as the penalized maximum likelihood estimator

$$L^{Lasso}(\beta) = \min_{\beta} \left[ (\theta_i - X_i\beta_i)^T (\theta_i - X_i\beta_i) + \lambda \sum_{i=1}^{k} |\beta_i| \right] \tag{8}$$

with penalization term $\lambda \sum_{i=1}^{k} |\beta_i|$ and $\lambda$ representing a measure of shrinkage (Tibshirani, 1996, 269) which is in contrary relation to tuning parameter $\vartheta$, i.e. for sufficient high values of $\lambda$ estimators $\beta_i$ are pushed towards zero. Due to its strong correlation with migration status *mig* we exclude the variable *native* in the regression.

## References

Allgood, Sam, Walstad, William B., Siegfried, John J., 2015. Research on teaching economics to undergraduates. J. Econ. Lit. 53 (2), 285–325.
Baker, F., Kim, S., 2004. Item Response Theory: Parameter Estimation Techniques, 2nd. ed. CRC Press.
Becker, W., Greene, W., Rosen, S., 1990. Research on high school economic education. Am. Econ. Rev.: Pap. Proc. 80 (2), 14–22.
Brückner, S., Förster, M., Zlatkin-Troitschanskaia, O., Walstad, W.B., 2015a. Effects of prior economic education, native language, and gender on economic knowledge of first-year students in higher education. A comparative study between Germany and the USA. Stud. High. Educ. 40 (3), 437–453.
Brückner, S., Förster, M., Zlatkin-Troitschanskaia, O., Happ, R., Walstad, W.B., Yamaoka, M., Asano, T., 2015b. Gender effects in assessment of economic knowledge and understanding: differences among undergraduate business and economics students in Germany, Japan, and the United States. Peabody J. Educ. 90 (4), 503–518.
Bruhn, M., Leão, L., de, S., Legovini, A., Marchetti, R., Zia, B., 2016. The impact of high school financial education: evidence from a large-scale evaluation in Brazil. Am. Econ. J. Appl. Econ. 8 (4), 256–295.
Bucher-Koenen, T., Lusardi, A., 2011. Financial literacy and retirement planning in Germany. J. Pension Econ. Financ. 10 (4), 565–584.
Cameron, M.P., Calderwood, R., Cox, A., Lim, S., Yamaoka, M., 2014. Factors associated with financial literacy among high school students in New Zealand. Int. Rev. Econ. Educ. 16, 12–21.
Cronbach, Lee J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16 (3), 297–334.
Davies, P., Mangan, J., Telhaj, S., 2005. Bold, reckless and adaptable? Explaining gender differences in economic thinking and attitudes. Br. Educ. Res. J. 31 (1), 29–48.
Dorans, N.J., Holland, P.W., 1992. DIF detection and description: mantel-Haenszel and standardization. Ets Res. Rep. Ser. 1, i-40.
Driva, Anastasia, Lührmann, Melanie, Winter, J., 2016. Gender differences and stereotypes in financial literacy: off to an early start. Econ. Lett. 146, 143–146.
Erner, C., Goedde-Menke, M., Oberste, M., 2016. Financial literacy of high school students: evidence from Germany. J. Econ. Educ. 47 (2), 95–105.
Frisancho, V., 2018. The Impact of School-based Financial Education on High School Students and Their Teachers: Experimental Evidence from Peru. Inter-American Development Bank Working Paper No. 871.
Fryer, Roland Jr, G., Levitt, S.D., 2010. An empirical analysis of the gender gap in mathematics. Am. Econ. J. Appl. Econ. 2 (2), 210–240.
Förster, M., Happ, R., Molerov, D., 2017. Using the U.S. test of financial literacy in Germany—adaptation and validation. J. Econ. Educ. 48 (2), 123–135.
Grohmann, A., Kouwenberg, R., Menkhoff, L., 2015. Childhood roots of financial literacy. J. Econ. Psychol. 51, 114–133.
Grohmann, A., 2016. The gender gap in financial literacy: income, education, and experience offer only partial explanations. DIW Econ. Bull. 46, 531–537.
Hambleton, R.K., Swaminathan, H., 1985. Item Response Theory: Principles and Applications. Kluwer-Nijhoff, Boston.
Happ, R., Förster, M., Zlatkin-Troitschanskaia, O., Carstensen, V., 2016. Assessing the previous economic knowledge of beginning students in Germany: implications for teaching economics in basic courses. Citizsh. Soc. Econ. Educ. 15 (1), 45–57.
Happ, R., Förster, M., 2018. The relationship between migration background and knowledge and understanding of personal finance of young adults in Germany. Int. Rev. Econ. Educ. https://doi.org/10.1016/j.iree.2018.06.003.
Heath, J., 1989. An econometric model of the role of gender in economic education. Am. Econ. Rev.: Pap. Proc. 79 (2), 226–230.
Hill, C.J., Bloom, H.S., Black, A.R., Lipsey, M.W., 2008. Empirical benchmarks for interpreting effect sizes in research. Child Dev. Perspect. 2 (3), 172–177.
Holland, P.W., Thayer, D.T., 1988. Differential item performance and the Mantel-Haenszel procedure. In: Wainer, H., Braun, H.I. (Eds.), Test Validity. Lawrence Erlbaum, Hillsdale, NJ, pp. 129–145.
Kaiser, T., 2017. [Financial sozialization and financial knowledge] Finanzielle Sozialisation und finanzielles Wissen. Vierteljahrshefte zur Wirtschaftsforschung 86 (4), 17–30.
Kaiser, T., Menkhoff, L., 2017. Does financial education impact financial literacy and financial behavior, and if so, when? World Bank Econ. Rev. 31 (3), 611–630.
Kaiser, T., Menkhoff, L., 2018. Financial Education in Schools: A Meta-Analysis of Experimental Studies. CESifo Working Paper Series 7395, CESifo Group Munich.
Kling, J.R., Liebman, J.B., Katz, L.F., 2007. Experimental analysis of neighborhood effects. Econometrica 75 (1), 83–119.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.
Lohr, Sharon L., 2010. Sampling: Design and Analysis, 2. ed. Brooks/Cole, Boston, MA.
Lord, F.M., 1980. Applications of Item Response Theory to Practical Testing Problems. Lawrence Erlbaum Associates, NJ.
Lord, F.M., 1983. Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. Psychometrika 48 (2), 233–245.
Lusardi, A., Mitchell, O.S., 2008. Planning and financial literacy: how do women fare? Am. Econ. Rev.: Pap. Proc. 98 (2), 413–417.
Lusardi, A., Mitchell, O.S., 2014. The economic importance of financial literacy: theory and evidence. J. Econ. Lit. 52 (1), 5–44.
Lührmann, M., Serra-Garcia, M., Winter, J., 2015. Teaching teenagers in finance: does it work? J. Bank. Financ. 54, 160–174.
Lührmann, M., Serra-Garcia, M., Winter, J., 2018. The impact of financial education on adolescents' intertemporal choices. Am. Econ. J. Econ. Policy 10 (3), 309–332.
Marsman, M., Maris, G., Bechger, T., Glas, C., 2016. What can we learn from plausible values? Psychometrika 81 (2), 274–289.
Miller, M., Reichelstein, J., Salas, C., Zia, B., 2015. Can you help someone become financially capable? A meta-analysis of the literature. World Bank Res. Obs. 30 (2), 220–246.
Oberrauch, L., 2019. Economic competence among children and youth: development and validation of a performance test]. Ökonomische Kompetenzen von Heranwachsenden: Entwicklung und Validierung eines Testinstruments Mimeo.
OECD, 2013. Gender and socio economic disparities in students' engagement, Drive and self-beliefs. Ready To Learn: Students' Engagement, Drive And Self-Beliefs Vol. III OECD Publishing, Paris.
OECD, 2014. PISA 2012 Results: Students and Money. Financial Literacy Skills for the 21st Century. PISA, OECD Publishing, Paris.
OECD, 2016. PISA 2015 Results (Volume I): Excellence and Equity in Education. PISA, OECD Publishing, Paris.
OECD, 2017. PISA 2015 Results (Volume IV): Students' Financial Literacy. PISA, OECD Publishing, Paris.
Rasch, G., 1960. Probabilistic Models for Some Intelligence and Attainment Tests. Kopenhagen: Paedagogiske Institut.
Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed. Sage Publications, Inc, Thousand Oaks, CA.
Remmele, B., Seeber, G., 2012. Integrative economic education to combine citizenship education and financial literacy. Citizsh. Soc. Econ. Educ. 11 (3), 189–201.
Retzmann, T., Seeber, G., Remmele, B., Jongebloed, H.-C., 2010. educational standards for economic education at all types of General-education schools in Germany. Final Report to the Gemeinschaftsausschuss der Deutschen Gewerblichen Wirtschaft (Working Group, Economic Education).
Retzmann, T., Seeber, G., 2016. Financial education in General education schools: a competence model. In: Aprea, C., Wuttke, E., Breuer, K., Koh, N.K., Davies, P., Greimel-Fuhrmann, B. (Eds.), International Handbook of Financial Literacy. Springer, Singapore, pp. 9–24.
Rose, N., von Davier, M., Xu, X., 2010. Modelling Nonignorable Missing Data with Item Response Theory (IRT). ETS Research Report Series 2010.
Schuhen, M., Schürkmann, S., 2014. Construct validity of financial literacy. Int. Rev. Econ. Educ. 16 (Part A(0)), 1–11.
Seeber, G., Kaiser, T., Oberrauch, L., 2019. A Test of Economic Competence for Students: Development, Validation, and Results from a Large-Scale Assessment in Germany. Mimeo.
Seeber, G., Körber, L., Hentrich, S., Rolfes, T., Haustein, B., 2018. Economic competence among Adolescents in Baden-Württemberg. Results for Frades 9, 10, and 11 in Ferneral Education Schools]. Ökonomische Kompetenzen Jugendlicher in Baden-Württemberg. Testergebnisse für die Klassen 9,10 und 11 der allgemeinbildenden Schulen. Hg. v. Stiftung Stiftung Würth, Künzelsau: Swiridoff.
Siegfried, John J., Fels, R., 1979. Research on teaching college economics: a survey. J. Econ. Lit. 17 (3), 923–969.
Thissen, D., Steinberg, L., Wainer, H., 1993. Detection of differential item functioning using the parameters of item response models. In: Holland, P.W., Wainer, H. (Eds.), Differential Item Functioning. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, pp. 67–113.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. 58 (1), 267–288.

Urban, C., Schmeiser, M., Michael Collins, J., Brown, A., 2018. The effects of high school personal financial education policies on financial behavior. Econ. Educ. Rev. https://doi.org/10.1016/j.econedurev.2018.03.006.
van Buuren, S., 2007. Multiple imputation of discrete and continuous data by fully conditional specification. Stat. Methods Med. Res. 16 (3), 219–242.
van Buuren, S., Oudshoorn, K., 2000. Multivariate Imputation by Chained Equations: MICE V1.0 User's Manual, Volume Pg/Vgz/00.038. TNO Prevention and Health, Leiden.
Warm, T.A., 1989. Weighted likelihood estimation of ability in item response theory. Psychometrika 54 (3), 427–450.
Walstad, W., 2013. Economic Understanding in US High School Courses. Am. Econ. Rev.: Pap. Proc. 103 (3), 659–663.
Walstad, W., Rebeck, K., 2001a. Assessing the Economic Understanding of U.S. High School Students. Am. Econ. Rev.: Pap. Proc. 91 (2), 452–457.
Walstad, W., Rebeck, K., 2001b. Test of Economic Literacy. Examiner's Manual, 3rd ed. National Council on Economic Education., New York, NY.
Walstad, W., Robson, D., 1997. Differential item functioning and male-female differences on multiple-choice tests in economics. J. Econ. Educ. 28 (2), 155–171.
Zieky, M., 1993. Practical questions in the use of DIF statistics. In: development, test, Holland, P.W., Wainer, H. (Eds.), Differential Item Functioning. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, pp. 67–113.