

Oberrauch, Luis; Kaiser, Tim

Preprint

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

Suggested Citation: Oberrauch, Luis; Kaiser, Tim (2018) : Economic competence in early secondary school: Evidence from a large-scale assessment in Germany, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/190763>

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.

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

Luis Oberrauch^a and Tim Kaiser^{a,b *}

^a *University of Koblenz-Landau*

^b *German Institute for Economic Research (DIW Berlin)*

Abstract

We employ a psychometrically validated performance test to study economic competence among a large and representative sample of 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. Our analysis sets the stage for a long-term study of economic competence among secondary school students and the impact of a recent curriculum reform introducing mandatory economic education.

JEL-Classification: A21, I21

Keywords: Economic competence, economic literacy, pre-college economic education, item response theory, gender gap

December 18, 2018

Acknowledgments: We would like to thank the schools and students participating in our study. Additionally, we appreciate comments from seminar participants in Hamburg, Freiburg, and Tübingen, especially Taiga Brahm, Günther Seeber, and Miguel Luzuriaga, as well as research assistance by Ngoc Anh Nguyen and Lena Masanneck.

Funding: This work was supported by *Stiftung Würrth*.

Luis Oberrauch, Research Associate, University of Koblenz-Landau, 76829 Landau, Germany; oberrauch@uni-landau.de.

Tim Kaiser* ([corresponding author](#)), Junior Professor of Economics and Economic Education, University of Koblenz-Landau and German Institute for Economic Research (DIW Berlin), 76829 Landau, Germany; kaiser@uni-landau.de.

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

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, 2001; Becker, 1997; 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 understanding 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., 2015; 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 2018; Happ and Förster 2018), and studies assessing economic understanding of (first-year) undergraduate university students in convenience samples (e.g. Happ et al., 2016), there is no representative study of economic competence in 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 to 10 for all general education school types. Our paper presents evidence from a representative large-scale assessment of early secondary school students 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. We employ a newly developed performance test (Seeber et al., 2018) 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 2017).

The study results in three main findings: First, we document a gender gap in economic competence favoring male students (+ 0.19 SDs, on average), 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. 2015), and the (adult) financial literacy literature (cf. Lusardi and Mitchell 2008, 2014 Lusardi et al. 2008; Bucher-Koenen and Lusardi 2010;). The gender gap favoring male students is very similar in size to the gender gap in mathematics achievement of German students documented in PISA 2015 (cf. OECD 2016) and 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. Additionally, we document that the gender gap increases along the distribution of economic competence: While the gap amounts to approximately 0.1

SD units at the 20th percentile, the gender gap increases to 0.15 SD units at the median and amounts to 0.23 SD units at the top of the economic competence distribution (80th percentile). Thus, the gender gap is especially evident among high performing students.

Second, we show that children of migrants (-0.23 SDs) and children of parents of lower socio- economic status (SES) have, on average, lower levels of economic competence. 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 80th percentile) relative to the effect at the median (-0.27 SDs) or the 20th percentile (-0.30 SDs). We investigate in complementary analyses (using a different definition of migration background based on the primary language used at home) that this effect is likely to be driven by differences in language skills resulting from the fact that the primary language at home is different from the language used in school (cf. Cameron et al. 2014; Happ and Förster 2018). 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, 0.5 to 0.8 standard deviation units 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 of about 0.94 standard deviation units. Differences between the lower tier school types (*Werkrealschule and Realschule*) are relatively small (about 0.25 to 0.3 SD units). Interestingly, economic competence in *Realschule* (e.g. a lower track school) is about 0.4 SD units higher than in the comprehensive school type (*Gemeinschaftsschule*). Additionally, the variance in economic

competence within schools and school types is much larger than across schools and school types, thus questioning the efficacy of the aspired tracking of students of similar (economic) ability into the school types. 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 process (2.1) and provides descriptive statistics for the sample of schools and students covered (2.2). Section 3 describes our empirical strategy in assessing the levels of economic competence (3.1) and dealing with missing values (3.2), the regression models used to study observable predictors of economic competence (3.3), as well as the methods used to assess the sensitivity of our results to model selection and choice of estimation method (3.4). Section 4 presents results while section 5 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”. Figures 1a and 1b show the population density of administrative districts and the spatial distribution of sampled clusters in Baden-Württemberg.

Figure 1: Population density map and spatial distribution of sampled schools

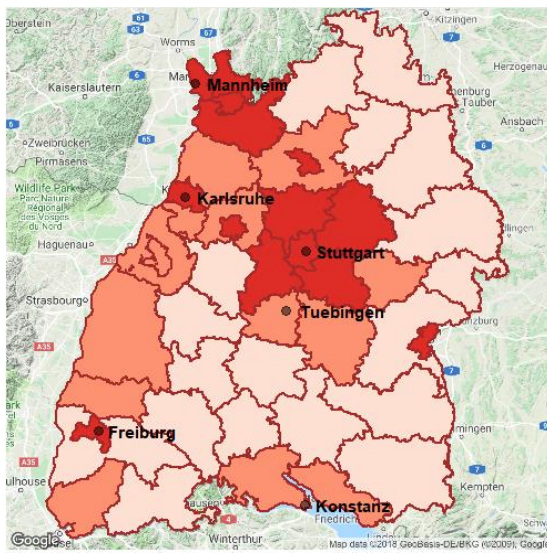


Fig. 1a: Population density in Baden-Wuerttemberg

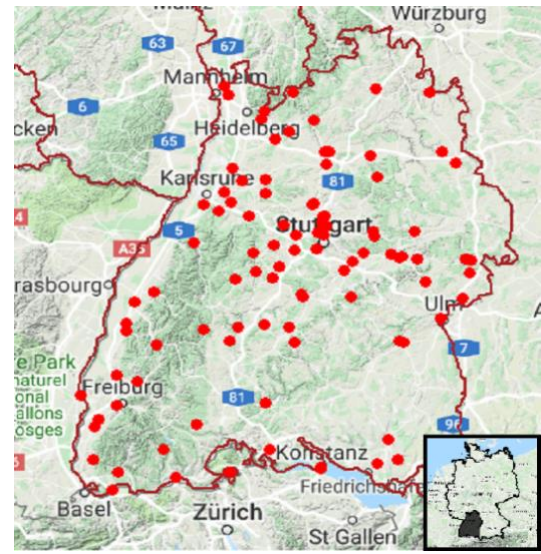


Fig. 1b: Spatial distribution of sampled schools

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”.

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 1689 students in 84 classes (sampling proportions by stratum are listed in [Table A1](#) in the appendix).

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, $w_{norm}^{(j,c)} = w^{(j,c)} \times \frac{1689}{\sum w^{(j,c)}}$.

2.2 Descriptive Statistics

Within our framework of the ongoing panel study *WiKo-BW* that measures longitudinal development of economic competences and evaluates the impact of the introduction of mandatory economic education, 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 3.1) 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 *sex*, “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). The mean value of this variable is 3.44 with a standard deviation of 1.62.

Table 1: Descriptive statistics at the student and school level

Statistic	N	Mean	St. Dev.	Min	Max
Student level					
WLE.500	1,687	500.000	100.000	87.308	884.689
WLE	1,687	0.095	0.984	-3.967	3.882
sex	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
WS	1,687	0.145		0	1
GMS	1,687	0.161		0	1
School size	1,687	658.4	222.5	226	1292
prop.no.native	1,687	0.346		0	1
Urbanization	1,687	1.939	0.808	1	3

In order to control for students' school performance, we ask for (self-assessed) 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 minutes ($SD=8.68$).

At the school level, we note that 39 percent of sampled students in the 7th grade in Baden-Wuerttemberg visit the most sophisticated school type (*Gym*) while 45 percent visit the lower tier school types (*WS 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 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*) ($M=0.346$) and the degree of urbanization of the area the school is located in (*urbanization*) as discussed above ($M=1.94$, $SD=0.81$).

3 Methods

This section describes the measurement of economic competence (3.1), the handling of missing data (3.2) and the regression models used to investigate predictors of economic competence among early secondary school students (3.3 and 3.4).

3.1 Assessing Economic Competence

We measure economic competence by using an adapted version of a comprehensive performance test for German secondary school students in grades 7 to 10 developed by Seeber et al. (2018). This test is based on a competence model described in Retzmann et al. (2010) and

Retzmann and Seeber (2017) where economic competence is divided into three competence areas and can be viewed from the perspective of three economic roles. The process of items selection, validation, and measurement characteristics of this comprehensive test for students ranging from grade 7 to 10 are reported in full in Seeber et al. (2018) and Oberrauch (2018). 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 7th grade students (see [Table A2](#) in the appendix). The number of items is distributed evenly among the competence areas. We validated the adapted competence test using probabilistic test theory, where manifest variables (item responses) attribute to latent abilities (economic competence). The latent ability is estimated with the One-Parameter-Logistic Model (1-PLM) (Rasch 1960):

$$P(x_{vi}) = \frac{\exp(x_{vi}(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)} \quad (1)$$

The maximum likelihood estimator maximizes the likelihood function resulting in person abilities and item difficulties on a common logit scale of [-4:4]. We rely on WLE estimators instead of the unweighted MLE estimators (Weighted Likelihood Estimation; Warm 1989) as they allow more precise measurements through weighting individual item information.

Figures 2a and 2b show the distribution of WLE estimates (economic competence) and their relationship to sum scores. 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 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 $rpb < 0.20$ 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.

Figure 2: Distribution of economic competence and correlation between sumscores and ability estimates

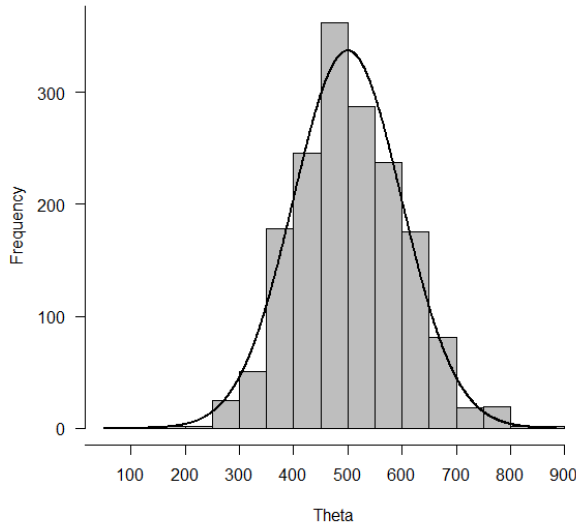


Fig.2a: Histogram of economic competence

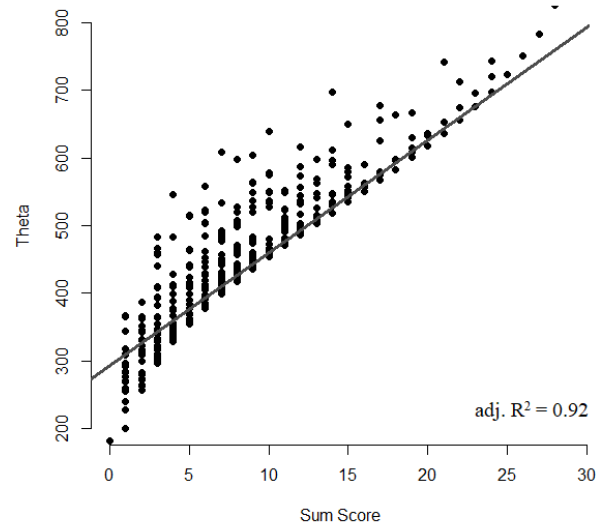


Fig. 2b: Correlation IRT ability and sum scores

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.

To assess test fairness among subgroups, we implemented differential item functioning (DIF) analyses by following the common classification of “educational testing services” (ETS) (Zwick, 2012). The ETS approach transforms the Mantel-Haenszel odds ratio α_{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 Mantel-Haenszel statistics MH_{χ^2} evaluates significance by following a χ^2 - distribution. If $\Delta - DIF_{MH}$ shows a non-significant, i.e. $MH_{\chi^2} < 3.84$, absolute value less than 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 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.

3.2 Missing Data and Plausible Values

Since test participation is voluntary, 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. We deal with the remaining missing values by implementing various procedures based on variable type. In educational large-scale assessments missing values in the competence test could either be 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. This method potentially leads to an underestimation of the levels of competence since the item could have been solved correctly anyways.

Thus, many educational assessments utilize mixed methods: The National Assessment of Educational Progress (NAEP) codes item non-response as missing in multiple choice item formats and as wrong in open item formats. PISA and TIMSS use mixed methods in terms of the different examination phases. Item non-responses are treated as missing for the item and test validation process while they are treated as wrong for population and regression analyses (see Pohl et al., 2010, 425). The test of economic competence (Seeber et al. 2018) uses mixed methods in terms of response time. If participants took more than three seconds for an item without choosing any answer, we code item-nonresponse as “wrong” (0). In this case, it is

plausible that participants did not know the correct answer. The remaining cases (i.e. response time < 3 seconds) are treated as “not available” (missing).

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 [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 & Oudshoorn 2000) that allows conducting the imputation process for each variable separately. For each predictor variable X_i with $i = 0, \dots, n$ we define a imputation model based on the conditional distribution $P(X_i | X_{-i})$. 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 softening 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)” 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.

3.3 Multiple Regression Models

We analyze determinants of economic competences by means of several multiple regression methods. The basic OLS regression model takes the form:

$$\theta_{i,j}^{PV} = \beta_{i,j} + \beta_{i,j}X_{i,j} + \varepsilon_{i,j} \quad (2)$$

where $\theta_{i,j}^{PV}$ represents the competence measure that is imputed and pooled from latent regression (plausible values) of participant i in school j . $X_{i,j}$ represents a covariates vector at the individual and school level. To ensure proper statistical inference, error term assumptions are adapted to the two-stage sampling process by weakening normal distribution assumptions on cluster levels

$$E(\varepsilon_{i,j}, \varepsilon_{i,j'}) = \begin{cases} 0 & \text{if } j = j' \\ \sigma & \text{if } j \neq j' \end{cases} \quad (3)$$

i.e. correlation between error terms within clusters is tolerable while error terms between clusters remain independent as unobserved factors like teacher influence, cognitive abilities or numeracy tend to be more similar within schools/classes than between schools.

In order to adapt estimation results to the clustered sampling structure we extend the basic OLS regression model to a hierarchical model. Therefore, we allow intercepts to vary across schools by usage of the random intercept model as varying regression coefficients between schools seem implausible. The extended multilevel model takes the form

$$\theta_{i,j}^{PV} = \beta_{0,j} + \beta_{i,j}X_{i,j} + \varepsilon_{i,j}^* \quad (4)$$

where $\varepsilon_{i,j}^*$ represents the cluster-robust standard errors specified in equation (2) and $\beta_{0,j}$ represents a composition of mean competence value γ_{00} and a group dependent deviation $u_{0,j}$

$$\beta_{0,j} = \gamma_{00} + u_{0,j} \quad (5)$$

3.4 Quantile Regression Models

As regression models (2) and (4) only account for the average relationship between covariate matrices and our target 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 process for the π th quantile with $0 < \pi < 1$ follows the minimization problem specified in Koenker and Bassett (1978)

$$\min_{\beta} \left[\sum_{(i:\theta_i > X_i\beta_i)} \pi |\theta_i^{PV} - X_i\beta_i| + \sum_{(i:\theta_i < X_i\beta_i)} (1 - \pi) |\theta_i^{PV} - X_i\beta_i| \right] \quad (6),$$

where θ_i^{PV} describes dependent variable economic competences from latent regression and X_i a vector of explanatory variables. Median regression at moment $M = 0.5$ is included as it accounts properly for outlier robustness and may serve as an important reference point.

3.5 Robustness exercises

For further check 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 ϑ . The minimization problem follows the equation

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

where ϑ 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. Equation (8) can also be written as the penalized maximum likelihood estimator

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

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

4 Results

4.1 Random intercept model

Table 2 shows results for the random intercept model specified in equation (5) while more naive OLS results are shown in [Table A4](#) in the appendix. The empty model predicts a comparatively strong intra-class correlation ($ICC = 0.27$). All non-categorical 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.

Models (1) – (6) show a moderate and slightly significant gender gap in favor of male participants, which corresponds with findings from an antecedent cross-sectional study in Baden-Württemberg (Seeber et al. 2018b). After controlling for survey duration (*time*), however, the effect size increases to significant 18.55 points in the complete model (see models 7 and 8). Next, a one-year increase in age lowers competence levels by almost 10 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 (Model 8). The initial effect size of -14.43 is inter alia confounded by school variables (Model 2) since students visiting a less sophisticated school type (*WS* and *GMS*) are older, on average, by (self-assessed) mathematical and reading abilities (Model 3) and by survey duration (*time*) (Model 7). Migration status (*mig*) is associated with a lower competence level by 23.01 points and remains relatively constant across 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.

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). An 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* (almost one standard deviation) followed by participants visiting a *Realschule* with a net effect of 45.66 on the competence scale. However, in Model (5) the effect of affiliation to type of school *Gymnasium* shrinks and the effect of *Werkrealschule* 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 survey *time*, a one-minute increase in total survey duration is associated with an increase of 2.95 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*, *WS* compared to *GMS*), *sex*, *book*, *mig*, *age*, and *time* are most predictive of the outcomes. From a qualitative perspective, the results on gender differences (*sex*), parental socioeconomic status (*book*) and immigration background (*mig*) appear to be most important.

Table 2: Regression estimates for the Random Intercept Model

<i>Dependent Variable: Economic Competence (Plausible Values)</i>								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
sex	12.732* (5.72)	13.247* (5.54)	10.717 (5.65)	12.754* (5.30)	11.127* (5.15)	13.074* (5.74)	19.840*** (4.97)	18.550*** (4.77)
age	-14.433*** (4.24)	-12.091** (3.97)	-13.287*** (4.02)	-13.997*** (4.15)	-10.767** (3.73)	-14.252** (4.35)	-12.914** (3.93)	-9.989** (3.61)
mig	-23.384*** (4.32)	-23.083*** (4.20)	-23.115*** (4.19)	-21.502*** (4.22)	-21.373*** (4.09)	-23.156*** (4.41)	-24.656*** (4.45)	-23.074*** (4.18)
book	9.169*** (1.61)	8.135*** (1.73)	7.494*** (1.62)	7.704*** (1.65)	5.629** (1.79)	9.007*** (1.60)	8.206*** (1.53)	4.855** (1.60)
GYM		101.499*** (8.50)			94.214*** (8.61)			93.904*** (9.13)
RS		45.656*** (7.74)			40.674*** (7.76)			41.291*** (7.36)
WS		21.268* (9.53)			14.936 (9.45)			12.140 (10.03)
selfmath			13.413*** (2.85)		11.614*** (2.79)			9.137*** (2.73)
selfread			12.032*** (3.17)		8.631** (2.97)			9.987** (3.34)
selfall			4.071 (4.80)		0.304 (4.66)			1.741 (4.63)
interesteco				11.735** (3.72)	10.090** (3.57)			8.161* (3.36)
interestimp				12.918*** (3.03)	11.427*** (2.92)			6.390* (3.15)
urbanization						10.484 (6.46)		-0.193 (3.48)
schoolsize						0.043 (0.02)		-0.020 (0.02)
prop_n_native						-18.000 (16.07)		0.549 (5.85)
bank_exp							-0.661 (2.56)	-1.083 (2.69)
time_mc							3.420*** (0.46)	2.948*** (0.43)
ownsalary							-2.918 (5.30)	-2.686 (5.06)
_cons	490.466*** (7.62)	443.068*** (7.39)	491.978*** (7.26)	490.028*** (7.35)	448.697*** (7.39)	489.974*** (7.20)	489.962*** (8.41)	448.898*** (7.86)
N	1687	1687	1687	1687	1687	1687	1687	1687
R ² (School)	0.08	0.815	0.105	0.093	0.809	0.294	0.233	0.885
R ² (Student)	0.08	0.066	0.101	0.112	0.113	0.071	0.223	0.231
R ² (total)	0.08	0.299	0.102	0.107	0.331	0.137	0.225	0.436

Notes: * p<0.05, ** p<0.01, *** p<0.001; R-squared based on Fisher's z transformation. Covariates are drawn from 20 multiple imputations; variance explained by missing values: 0.001. SE cluster are based on 84 Jackknife-Repeated Replication zones; Intra-class correlation: 0.27.

4.2 Quantile Regression

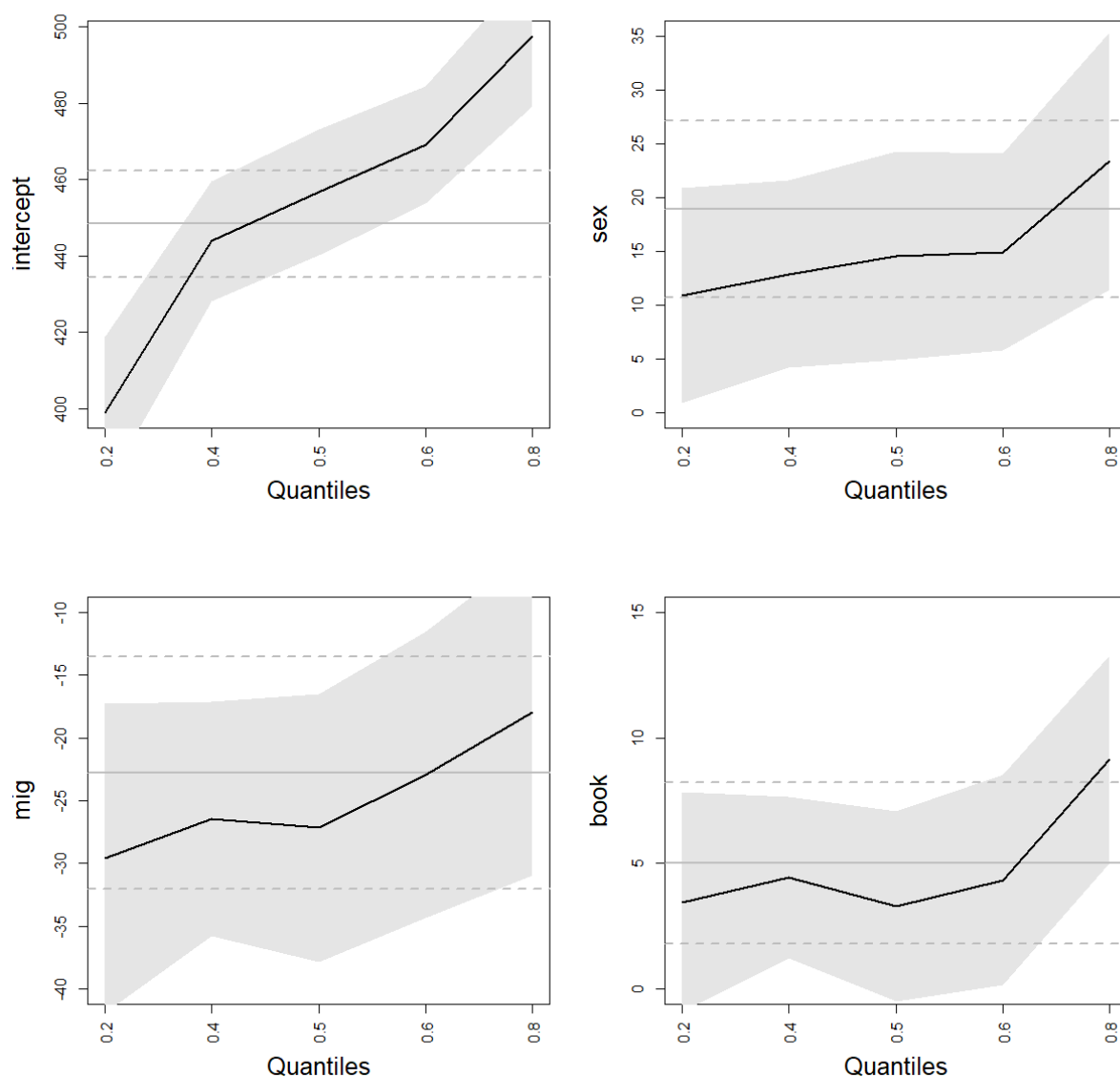
In an additional step we conduct a quantile regression as specified in equation (7). Results for selected variables of interest are displayed in Figure 3 while complete models are shown in [Table A5](#) in the appendix.

The plots show quantile regressions estimates across five different distribution moments (0.2, 0.4, 0.5, 0.6, 0.8) and the associated 95% confidence interval (CI) represented by the grey area. For comparisons, the solid horizontal line shows coefficient estimates from conventional mean regression (OLS) and the dotted lines indicate the bounds of its 95% CI.

The plot corresponding to the gender variable *sex* locates within OLS confidence bands along all distribution moments and show - in comparison to OLS estimates - a weaker gender gap at the first four moments. At the highest moment, however, male participants show a higher test result by 23.38 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 immigration background (*mig*) at $M=.2$ achieve a lower test result by -29.62 points (-0.296 SD units) ($p < .01$) while effect size shrinks almost by half to -18.00 points ($p < .05$) among high performers at distribution point $M = .8$. The importance of educational background of parents (*book*) remains within OLS confidence bands along the first distribution moments and only exceeds the OLS regression estimate at $M = .8$ with an effect size of 9.15 points ($p < .01$) while all remaining distribution moments show smaller effects than mean regression. Thus, gender differences are more pronounced among top performing students, differences between students with an immigration background and students with parents born in Germany are smaller at the higher moments at 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.

Figure 3: Quantile regression estimates for selected variables



Notes: Results from simultaneous quantile regressions as specified in table A5 in the appendix. The solid grey horizontal line shows the OLS (mean regression) result. The dashed lines show the upper and lower bound of the associated 95 percent confidence interval. The shaded grey area around the quantile regression estimates (solid black) line shows the corresponding confidence interval for the simultaneous quantile regression estimates. The standard errors for this estimations are bootstrapped with 1,000 replications.

4.3 Robustness exercises

To account for uncertainty arising from model selection, we complement these analyses by conducting a LASSO regression. [Table A6](#) (see appendix) 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 also 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. Additionally, we present results from OLS regressions (with standard errors clustered at the school-level) as an alternative modelling approach in [Table A4](#). Results are near identical.

5 Discussion & Conclusion

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 (0.19 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 and Levitt 2010). 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 increased 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 8 point 8 increase in economic competence (0.08SD). 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 immigration background perform worse (-0.23 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 an immigration 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). Similar to Happ and Förster (2018), we investigate the source of this difference in test performance and find that the effect of immigration 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.

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 10th 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 heterogeneous 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 immigration background to those without a history of migration and the effects on students with parents of lower socio-economic status.

References

- Allgood, Sam, William B. Walstad, and John J. Siegfried. (2015). Research on Teaching Economics to Undergraduates. *Journal of Economic Literature*, 53 (2): 285-325.
- Becker, W., Greene, W., & Rosen, S. (1990). Research on High School Economic Education. *American Economic Review: Papers & Proceedings*, 80(2): 14-22.
- Brückner, S., Förster, M., Zlatkin-Troitschanskaia, O., & Walstad, W. B. (2015). 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. *Studies in Higher Education*, 40(3): 437–453.
- Brückner, S., Förster, M., Zlatkin-Troitschanskaia, O., Happ, R., Walstad, W. B., Yamaoka, M., & Asano, T. (2015). Gender Effects in Assessment of Economic Knowledge and Understanding: Differences Among Undergraduate Business and Economics Students in Germany, Japan, and the United States. *Peabody Journal of Education*, 90(4): 503–518.
- Bruhn, M., Leão, L. de S., Legovini, A., Marchetti, R., and Zia, B. (2016). The Impact of High School Financial Education: Evidence from a Large-Scale Evaluation in Brazil. *American Economic Journal: Applied Economics*, 8(4), 256–295.
- Bucher-Koenen, T., & Lusardi, A. (2011). Financial literacy and retirement planning in Germany. *Journal of Pension Economics and Finance* 10 (4): 565–84.
- Cameron, M. P., Calderwood, R., Cox, A., Lim, S., & Yamaoka, M. (2014). Factors associated with financial literacy among high school students in New Zealand. *International Review of Economics Education*, (16): 12–21.
- Davies, P., Mangan, J., & Telhaj, S. (2005). Bold, Reckless and Adaptable? Explaining Gender Differences in Economic Thinking and Attitudes. *British Educational Research Journal*, 31(1): 29-48.

- Erner, C., Goedde-Menke, M., & Oberste, M. (2016). Financial literacy of high school students: Evidence from Germany. *Journal of Economic Education*, 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, G. Jr., & Levitt, S. D.. (2010). An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics*, 2 (2): 210-40.
- Förster, M., Happ, R. & Molerov, D. (2017). Using the U.S. Test of Financial Literacy in Germany—Adaptation and validation. *Journal of Economic Education*, 48(2): 123-135.
- Grohmann, A., Kouwenberg, R., & Menkhoff, L. (2015). Childhood roots of financial literacy. *Journal of Economic Psychology*, 51: 114–133.
- 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. *Citizenship, Social and Economics Education*, 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. *International Review of Economics Education*. <https://doi.org/10.1016/j.iree.2018.06.003>.
- Heath, J. (1989). An Econometric Model of the Role of Gender in Economic Education. *American Economic Review: Papers & Proceedings*, 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 Development Perspectives*, 2(3), 172–177.
- Holland, P. W., & Thayer, D. T. (1988): Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum.
- Kaiser, T. (2017). 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 Economic Review*, 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. Boston, MA: Brooks/Cole.
- Lusardi, A. & Mitchell, O. S. (2008): Planning and Financial Literacy: How Do Women Fare? *American Economic Review: Papers and Proceedings* 98 (2): 413-417.

- Lusardi, A., Mitchell, O. S., and Curto, V. (2010). Financial Literacy among the Young. *Journal of Consumer Affairs*, 44(2), 358–380.
- Lusardi, A. and Mitchell, O. S. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature*, 52(1), 5–44.
- Lührmann, M., Serra-Garcia, M., & Winter, J. (2015). Teaching teenagers in finance: Does it work? *Journal of Banking & Finance*, 54: 160–174.
- Lührmann, M., Serra-Garcia, M., & Winter, J. (2018). The Impact of Financial Education on Adolescents' Intertemporal Choices. *American Economic Journal: Economic Policy*, 10(3): 309–332.
- Miller, M., Reichelstein, J., Salas, C., & Zia, B. (2015). Can You Help Someone Become Financially Capable? A Meta-Analysis of the Literature. *World Bank Research Observer*, 30(2), 220–246.
- Oberrauch, L. (2018). Empirische Validierung eines Wirtschaftskompetenztests für Lernende der 7. Klassensstufe. *Mimeo*.
- OECD (2013). Gender and Socio Economic Disparities in Students' Engagement, Drive and Self-Beliefs. In: *Ready To Learn: Students' Engagement, Drive And Self-Beliefs – Volume 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.
- Pohl, S., Gräfe, L., & Rose, N. (2014). Dealing with omitted and not-reached items in competence tests. Evaluating approaches accounting for missing responses in item response theory models. *Educational and Psychological Measurement*, 74 (3): 423–452.
- Rasch, G. (1960): *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: Paedagogiske Institut.
- Remmele, B., & Seeber, G. (2012). Integrative Economic Education to Combine Citizenship Education and Financial Literacy. *Citizenship, Social and Economics Education*, 11(3): 189–201.
- Retzmann, T., Seeber, G., Remmele, B. & Jongbloed, 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).
- Schuhen, M., & Schürkmann, S. (2014). Construct validity of financial literacy. *International Review of Economics Education*, 16, Part A(0), 1–11.

- Seeber, G., Kaiser, T. and Oberrauch, L. (2018a): 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. (2018b): *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. *Journal of Economic Literature* ,17(3): 923–69.
- Tibshirani R (1996): Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 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. *Economics of Education Review*. <http://doi.org/10.1016/j.econedurev.2018.03.006>
- van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, 16(3):219-42.
- van Buuren,S. & Oudshoorn, K. (2000). *Multivariate Imputation by Chained Equations: MICE VI.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. *American Economic Review: Papers & Proceedings* , 103 (3): 659-63.
- Walstad, W., & Rebeck, K. (2001). Assessing the Economic Understanding of U.S. High School Students. *American Economic Review: Papers & Proceedings*, 91(2), 452-457.
- Walstad, W., & Robson, D. (1997). Differential Item Functioning and Male-Female Differences on Multiple-Choice Tests in Economics. *Journal of Economic Education*, 28(2): 155-171.
- Zwick, R. (2012): *A Review of ETS Differential Item Functioning Assessment Procedures: Flagging Rules, Minimum Sample Size Requirements, and Criterion Refinement*. Educational Testing Service. Princeton Research Report, ETS RR-12-08.

Appendix

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
WS	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

	Item	Source	n	CTT		IRT			
				RelFreq	rit_c	sigma	[SE]	Infit	Outfit
1	Broetchenuberangebot	Seeber et al. (2018a)	1689	0.738	0.258	-1.103	[0.058]	1.006	0.987
2	SunkCosts2	Oberrauch (2018)	1619	0.634	0.164	-0.581	[0.055]	1.064	1.116
3	Lusardi_1	Lusardi and Mitchell (2014)	1556	0.638	0.223	-0.544	[0.056]	1.056	1.040
4	Schulflohmarkt_1	Seeber et al. (2018a)	1554	0.641	0.257	-0.536	[0.056]	1.024	1.028
5	Inflation	Seeber et al. (2018a)	1560	0.624	0.345	-0.465	[0.055]	0.970	0.945
6	Lohnerhoehung	Seeber et al. (2018a)	1550	0.611	0.383	-0.377	[0.055]	0.971	0.943
7	BubbleTea_1	Seeber et al. (2018a)	1587	0.595	0.355	-0.338	[0.054]	0.978	0.965
8	Elastizitaet	Seeber et al. (2018a)	1565	0.591	0.379	-0.306	[0.055]	0.941	0.917
9	Arbeitslosigkeit	Oberrauch (2018)	1584	0.569	0.397	-0.171	[0.054]	0.943	0.920
10	Selbststaendigkeit	Oberrauch (2018)	1558	0.562	0.405	-0.144	[0.054]	0.939	0.922
11	Schuhe_1	Seeber et al. (2018a)	1566	0.557	0.365	-0.137	[0.054]	0.971	0.952
12	Preisstrategie	Oberrauch (2018)	1620	0.550	0.376	-0.114	[0.053]	0.972	0.954
13	SunkCosts	Oberrauch (2018)	1570	0.520	0.313	-0.006	[0.054]	0.994	0.982
14	StandortDeutsch_1	Seeber et al. (2018a)	1613	0.534	0.409	0.009	[0.053]	0.921	0.902
15	Opportunitaetskosten	Seeber et al. (2018a)	1614	0.524	0.383	0.017	[0.053]	0.938	0.918
16	Rechnung_2_CR	OECD (2012)	1628	0.518	0.371	0.022	[0.053]	0.960	0.946
17	Schule_1	Seeber et al. (2018a)	1602	0.490	0.337	0.171	[0.054]	0.985	0.975
18	Lohnabrechnung_3	Seeber et al. (2018a)	1606	0.366	0.174	0.723	[0.056]	1.082	1.117
19	Fairtrade_1_CR	Seeber et al. (2018a)	1629	0.338	0.411	0.899	[0.057]	0.926	0.887
20	Preisfunktion_CR	Oberrauch (2018)	1571	0.327	0.399	0.981	[0.059]	0.931	0.886
21	Steuerprogression_1	Seeber et al. (2018a)	1619	0.289	0.371	1.121	[0.059]	0.946	0.918
22	Ungarn_2	Seeber et al. (2018a)	1623	0.290	0.315	1.161	[0.06]	0.985	0.986
23	Verbraucher_2	Seeber et al. (2018a)	1628	0.284	0.207	1.181	[0.06]	1.056	1.104
24	Handyvertrag_CR	Oberrauch (2018)	1629	0.257	0.37	1.313	[0.061]	0.945	0.898
25	Konsum_1	Seeber et al. (2018a)	1609	0.251	0.101	1.326	[0.062]	1.107	1.198
26	Rente_1	Seeber et al. (2018a)	1618	0.224	0.174	1.496	[0.064]	1.068	1.201
27	KreditundSparen_1	Seeber et al. (2018a)	1614	0.206	0.344	1.648	[0.067]	0.956	0.912
28	Unternehmen_2B	Seeber et al. (2018a)	1628	0.176	0.212	1.799	[0.069]	1.030	1.087
29	ProdVerlagerung_CR	Oberrauch (2018)	1611	0.144	0.349	2.214	[0.079]	0.939	0.788
30	Unternehmen_1	Seeber et al. (2018a)	1634	0.102	0.132	2.456	[0.085]	1.036	1.348

Note: IRT analysis is conducted by means of One-parametric IRT-Model; 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). Scale reliability (Cronbach's alpha) from classical test theory is 0.81.

Table A3: Missing proportions and Cohens' d

Variable	Missing fraction	<i>Cohen's d</i>
<i>age</i>	0.08	-0.14
<i>mig</i>	0.08	-0.12
<i>native</i>	0.08	-0.12
<i>book</i>	0.09	-0.18
<i>selfmath</i>	0.1	-0.92
<i>selfall</i>	0.01	-0.39
<i>interesteco</i>	0.02	-0.37
<i>interestimp</i>	0.02	0.00
<i>bankaccount</i>	0.1	-0.13
<i>bankEC</i>	0.1	-0.18
<i>ownsalar</i>	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: OLS Regression estimates

Dependent Variable: Economic Competence (Plausible Values)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
sex	10.970* (4.92)	13.396** (4.47)	9.206 (4.82)	10.821* (4.78)	11.114* (4.36)	12.529* (4.94)	20.220*** (4.71)	18.902*** (4.20)
age	-24.973*** (3.40)	-12.047*** (3.23)	-22.742*** (3.37)	-23.944*** (3.39)	-10.583*** (3.20)	-23.817*** (3.49)	-23.207*** (3.20)	-10.304*** (3.07)
mig	-25.457*** (5.42)	-23.178*** (5.02)	-24.827*** (5.31)	-23.540*** (5.35)	-21.491*** (4.92)	-24.718*** (5.54)	-26.562*** (5.17)	-22.779*** (4.72)
book	18.124*** (1.73)	8.608*** (1.79)	15.391*** (1.78)	16.318*** (1.73)	6.110*** (1.76)	17.003*** (1.78)	15.734*** (1.63)	5.017** (1.63)
GYM		103.688*** (7.17)			97.036*** (6.86)			95.450*** (7.27)
RS		47.032*** (6.30)			42.457*** (6.15)			41.429*** (5.85)
WS		22.554** (8.39)			16.223* (8.25)			11.878 (8.10)
selfmath			15.340*** (3.35)		11.460*** (2.89)			9.238** (2.87)
selfread			14.308*** (4.04)		6.990 (3.70)			8.615* (3.49)
selfall			4.443 (4.34)		0.171 (4.16)			2.470 (4.02)
interesteco				10.673** (3.94)	9.639** (3.62)			7.673* (3.37)
interestimp				15.595*** (3.79)	12.500*** (3.50)			7.583* (3.41)
urbanization						8.333** (2.88)		-0.883 (2.52)
schoolsize						0.026* (0.01)		-0.023* (0.01)
prop.n.native						-13.295 (6.93)		1.304 (5.46)
bank.exp							1.531 (2.92)	-0.225 (2.65)
time.mc							3.506*** (0.39)	2.842*** (0.33)
ownsalary							-6.687 (5.32)	-3.362 (4.85)
constant	497.82*** (4.33)	441.21*** (6.60)	498.70*** (4.17)	497.31*** (4.24)	446.59*** (6.37)	496.89*** (4.24)	498.72*** (5.53)	448.44*** (7.12)
N	1687	1687	1687	1687	1687	1687	1687	1687
R ²	0.20	0.32	0.23	0.22	0.36	0.21	0.29	0.43

Notes: * p<0.05, ** p<0.01, *** p<0.001; R-squared based on Fisher's z transformation. Covariates are drawn from 20 multiple imputations; variance explained by missing values: 0.001.

Table A5: Quantile Regression – full models

Dependent Variable: Economic Competence (Plausible Values)					
	20 th percentile	40 th percentile	Median	60 th percentile	80 th percentile
<i>sex</i>	10.888* (5.11)	12.873** (4.44)	14.552** (4.95)	14.942** (4.68)	23.378*** (6.11)
<i>age</i>	-7.639* (3.59)	-10.219*** (2.92)	-10.828*** (3.15)	-8.057* (3.45)	-11.356* (4.69)
<i>mig</i>	-29.616*** (6.22)	-26.469*** (4.77)	-27.172*** (5.48)	-22.939*** (5.84)	-17.995** (6.66)
<i>book.</i>	3.453 (2.22)	4.438** (1.65)	3.295 (1.95)	4.339* (2.14)	9.151*** (2.10)
<i>GYM</i>	91.607*** (10.11)	85.283*** (7.55)	89.737*** (8.58)	94.666*** (8.31)	108.791*** (9.23)
<i>RS</i>	40.979*** (8.39)	30.587*** (6.95)	32.441*** (7.20)	35.753*** (6.71)	51.932*** (8.69)
<i>C_WS</i>	0.181 (9.59)	-5.492 (8.18)	-4.413 (8.32)	0.354 (8.11)	17.814 (9.75)
<i>selfmath</i>	5.905* (2.83)	8.749** (2.95)	11.161** (3.63)	10.982** (3.41)	10.613** (3.66)
<i>selfread</i>	7.020 (4.06)	10.133** (3.44)	10.278** (3.55)	10.337** (3.51)	14.714** (4.70)
<i>selfall</i>	3.645 (4.57)	4.341 (4.09)	2.566 (4.79)	0.024 (4.93)	3.926 (5.44)
<i>interesteco</i>	6.727 (4.91)	5.288 (3.63)	6.140 (3.77)	5.454 (3.61)	5.307 (5.14)
<i>interestimp</i>	4.719 (4.67)	5.812 (3.21)	5.985 (3.36)	7.405 (4.24)	3.795 (5.16)
<i>urbanization</i>	-1.832 (3.31)	2.570 (2.66)	1.322 (2.99)	1.362 (3.10)	1.276 (3.51)
<i>schoolsize</i>	-0.028 (0.02)	-0.018 (0.01)	-0.024* (0.01)	-0.020 (0.01)	-0.020 (0.02)
<i>prop.n.native</i>	8.169 (5.89)	2.067 (5.15)	-2.649 (5.57)	-1.588 (5.82)	2.582 (7.34)
<i>bank.exp</i>	3.616 (3.53)	1.832 (2.47)	2.476 (2.98)	2.481 (2.89)	0.918 (3.86)
<i>time</i>	3.281*** (0.36)	3.145*** (0.31)	2.973*** (0.34)	2.797*** (0.35)	2.531*** (0.40)
<i>ownsalary</i>	-1.460 (5.53)	-4.177 (4.70)	0.013 (5.19)	-0.760 (5.33)	-5.552 (6.79)
<i>constant</i>	398.905*** (10.06)	443.843*** (7.98)	456.631*** (8.33)	469.040*** (7.81)	497.531*** (9.43)

Notes: Results from simultaneous quantile regression. Covariates are drawn from 20 multiple imputations; dependent variable Economic Competence is multiple imputed from latent regression (Plausible Values). Number of observations is N=1687. Standard errors in parentheses are bootstrapped with 1,000 replications. * p<0.05, ** p<0.01, *** p<0.001.

Table A6: 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_WS	-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_owns salary	-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	.