Empirical Research in
Vocational Education and Training

RESEARCH

Open Access

Measuring financial literacy with a Situational Judgement Test: do some groups really perform worse or is it the measuring instrument?

Eveline Wuttke[1], Christin Siegfried[1*] and Carmela Aprea[2]

*Correspondence:
siegfried@em.uni-frankfurt.de
[1] Goethe University Frankfurt,
Theodor-W.-Adorno-Platz 4,
60629 Frankfurt, Germany
Full list of author information
is available at the end of the
article

Abstract

Due to current trends in society and economy, financial literacy is often considered as an important twenty-first century skill. However, regardless of the postulated relevance, studies suggest that financial illiteracy seems to be a widespread phenomenon in the population of many nations. Some studies also show that some groups perform particularly poorly (e.g. women, persons with migration background and/or low level of education). These differences are often attributed to different individual characteristics such as abilities, dispositions or socialisation patterns. However, available research also suggests that even after controlling for them, a quite large portion of the performance differences between the various groups of test-takers remains unexplained. One explanation for performance gaps in financial literacy might be that differences in test scores could also be evoked by the test instruments itself and may thus, at least in part, be interpreted as testing bias. In this paper, we present a newly developed Situational Judgement Test, which is focused on financial competence. For this test, we examine whether differences between groups are attributable to individual differences or due to a test bias. To analyse a possible test bias, we tested one facet of financial literacy (with three factors: control of one's financial situation, budgeting and handling of money) related to everyday money management for measuring invariance for different groups. If measuring invariance could be assumed, we analysed group differences by calculating t-tests. Results show that two factors of the test show measurement invariance for all groups considered (gender, migration and educational background, opportunities to learn). Group comparisons are thus possible and potential differences are not due to a test bias. For one factor, we can only assume measurement invariance for the group with/without migration background and with/without opportunities to learn in financial topics. When we look at group differences, we find that in contrast to the findings of many previous studies, the analysis of the mean differences does not show any systematic deficits in financial literacy for specific groups.

Keywords: Financial literacy, Situational judgement test, Measuring invariance, Gender, Migration background, Educational background, Opportunities to learn

Introduction

The current societal and economic landscape is characterized by a growing degree of complexity, increasingly risky and globalised marketplaces as well as a high diversity of available financial products. In addition, a wide-ranging transfer of risk has occurred from governments and employers to employees and consumers (e.g., reduced state-supported pensions and health-care benefits in many countries). This imposes onto individuals the responsibility to care for their own financial security in case of, for example, illness, unemployment or retirement. Furthermore, if individuals use the services of financial intermediaries/advisors, they need to understand what is being offered to them (Aprea et al. 2016). Against the background of these requirements, the issue of finance-related knowledge, skills and attitudes—usually termed as financial literacy—is increasingly attracting the attention of politicians and scientists. A high level of financial literacy is considered as being conditional for sound financial decisions as well as one protective factor to avoid over-indebtedness and to provide for illness and old age in order to secure personal financial prosperity (e.g., Braunstein and Welch 2002; Lusardi and Mitchell 2014). Besides its influence at the micro level (individual life), financial literacy is also considered important when it comes to macro level concerns such as financial stability (e.g., Mitchell and Lusardi 2015).

Due to the relevance of the topic, numerous national and international surveys and empirical studies (e.g., Allianz 2017; OECD 2017) have been conducted in recent years, showing that financial illiteracy seems to be a widespread phenomenon in the population of many nations. These studies also show that some groups often perform particularly poorly. This is primarily the case for women as well as for persons with migration background and/or a low level of education (e.g., Bucher-Koenen et al. 2017; Happ and Förster 2019). These differences are often attributed to different individual dispositions, such as interest in financial issues (e.g., Brown and Graf 2013; Lührmann et al. 2015) or differences in socialisation patterns and learning opportunities (e.g., Rinaldi 2017; Rudeloff 2019). However, despite the fact that all these aspects are plausible and show to have a certain explanatory power, available research (e.g., Fonseca et al. 2012; Greimel-Fuhrmann and Silgoner 2018; Rudeloff et al. 2019) also suggests that even after considering them, a quite large portion of the performance differences between the various groups of test-takers remains unexplained. Consequently, further research on alternative explanations is required. One explanation for performance gaps in financial literacy might be that differences in test scores could also be evoked by the test instruments itself and may thus, at least in part, be interpreted as testing bias of (conventional) financial literacy measurements. Those measurements typically take the form of knowledge oriented multiple-choice questions, as for example the widely used "big three" or "big five" financial literacy questions developed by Lusardi and Mitchell (2011).

However, there is a clear research gap on the question of test bias in measuring financial literacy. This is quite surprising, not only because conventional financial literacy measurements have been criticised for quite a long time (e.g., Remund 2010; Huston 2010), but also because testing biases have been discussed intensively in psychological assessment research under the notion of "test fairness" for many years (e.g., Melikyan et al. 2019). Moreover, bias-induced disparities of test performance have been found in related domains, such as mathematics and economics (e.g., Asarta et al. 2014; Reardon

et al. 2018; for a recent review see also Siegfried and Wuttke 2019). If unaccounted for, testing biases may result in inappropriate diagnosis, treatment, placement, or denial of services/positions (e.g., Dilworth-Anderson et al. 2008).

The reasons for choosing a Situational Judgement Test (SJT) for our studies have been explained in detail elsewhere (Wuttke and Aprea 2018), therefore we will only outline them briefly here. In many test situations, knowledge tests are used to evaluate financial literacy of test takers. With these results, it is then tried to predict behaviour in later real life situations. The problem with this approach is that there usually is a gap between knowing and doing and that test results that represent knowledge do not necessarily predict later behaviour (e.g., Fernandes et al. 2014; Tang et al. 2015; Kaiser and Menkhoff 2017). SJTs, however, represent a type of psychological test that presents test takers with realistic, hypothetical situations or scenarios and asks them to identify the most appropriate response or rank the responses in the order they feel is most suitable (e.g. Kahmann 2014; McDaniel and Nguyen 2001; Whetzel and McDaniel 2009). Later behaviour in real situations can be inferred from these decisions. It is assumed that SJTs measure the participants' procedural context-specific knowledge and situational decision-making ability (Kahmann 2014, p. 49). Detailed information on the different forms of SJT and the test development can be found in Wuttke and Aprea (2018). Since we use this new test form (at least "new" in the area of financial literacy) in our studies, we will examine in this paper whether this type of testing can reduce or eliminate differences often found in conventional tests with disadvantages for women, people with migration background, those with lower educational qualification and those with less opportunities to learn in financial topics.

We proceed as follows: In chapter two, we describe the state of research on financial literacy and especially focus on factors that differentiate performance in financial literacy (gender, migration background, educational background, opportunities to learn). We then outline the research questions and the study design (Sect. 3), and present results of the study (Sect. 4). These results as well as the limitations of the study are then discussed, and conclusions with regard to further steps are drawn (Sect. 5).

Different results in financial literacy for specific groups

Studies that measure financial literacy of (young) adults indicate that a large group of these seem to have a considerable lack of knowledge in finance related topics and considerable difficulties in making proper financial decisions (e.g., Gramatki 2017; Ergün 2017; Lusardi and Mitchell 2014; Happ et al. 2018; Rudeloff et al. 2019; Strömbäck et al. 2017). Although different financial literacy tests are used in these studies and thus partly different test formats besides the classic multiple-choice items (e.g., Gramatki 2017; Happ and Förster 2019; Lusardi et al. 2014; Rudeloff 2019; Strömbäck 2017) or true–false items (Tang et al. 2015) are used, results of the studies show rather unanimously that financial literacy appears to differ in terms of various socio-demographic factors and educational background.

Gender

Regarding socio-demographic data, results of most studies indicate a *gender effect* and generally men perform better than women in financial literacy tests (Chen and Volpe

Wuttke *et al. Empirical Res Voc Ed Train*   (2020) 12:18

Page 4 of 21

2002; Lusardi and Mitchell 2011, 2014; Hung et al. 2009; Atkinson and Messey 2012; Woodyard and Robb 2012; Agnew et al. 2013; Bucher-Koenen et al. 2014; Lusardi et al. 2014; Schuhen and Schürkmann 2014; Agnew and Harrison 2015; Almenberg and Dreber 2015; Filipial and Walle 2015; Bannier and Neubert 2016; Ergün 2017; Gramatki 2017; Hasler and Lusardi 2017; Killins 2017; OECD 2017; Strömbäck et al. 2017; Förster et al. 2018; Greimel-Fuhrmann and Silgoner 2018; Happ et al. 2018; Preston and Wright 2019). While most studies find these differences when the construct is modelled one-dimensionally[1] (Rudeloff et al. 2019), there are some studies that model financial literacy multi-dimensionally and find differences in partial facets in favor of women. It is interesting to note that this gender gap seems to persist across different age groups (Lusardi and Mitchell 2014; Happ et al. 2018; Rudeloff et al. 2019) and even a better educational background of women is not always able to overcome the gender gap (Mahdavi and Horton 2012).

Only a few studies show—at least partially—no gender differences (e.g., Hill and Asarta 2016; Walstad et al. 2010; OECD 2017; Strömbäck et al. 2017) or point to advantages for women compared to men in some facets of financial literacy, especially if other factors are considered simultaneously. Förster et al. (2018) show in a sample of over 1000 young adults that women perform significantly worse than men in banking, everyday money management and insurance, but that this effect disappears when controlling for interest and media use. Schürkmann (2017) tested students between 14 and 17 years of age in six areas of financial literacy and found that men only significantly outperformed women in the area of debt. In a study by Rudeloff et al. (2019), female participants perform better in money and payments and insurance, while male subjects perform better in savings and monetary policy. There are no differences in the facet loans. Some PISA results show a country specific gender gap. Women score lower than men only in Italy. In Australia, Lithuania, Poland, Slovakia, and Spain, on the contrary, girls perform significantly better (OECD 2017). The 2018 PISA study shows no systematic differences in favour of male participants (OECD 2018).

Migration background

Another factor that differentiates between the performance of test-takers in financial literacy tests is their *migration background* (e.g., Gramatki 2017; Happ et al. 2018; Rudeloff et al. 2019; Happ and Förster 2019). This effect is explained by the fact that immigrants often have a poorer economic background and parents who work in lower-skilled jobs or who do not speak the test language at home. Some studies, which not only look at a global specification of migration background but rather take into account in which generation the migration has taken place, find that the strongest negative effect is recorded for the first-generation immigrants. And the effect decreases continuously with the second- and third- generation immigrants (Gramatki 2017). A definite limitation is that different definitions of the construct financial literacy are used across the studies and the

---

[1] One-dimensionality of a construct is given, if there are no further differentiations within the construct. Multidimensionality is assumed if theoretically and/or empirically differentiable substructures can be assumed within a comprehensive construct (e.g. financial literacy). In our test we distinguish the dimensions "financial literacy relevant for individual decisions" vs. "financial literacy relevant for societal decisions". Within these dimensions we model different facets, e.g. in the individual dimension the facet "planning and managing financial decisions of everyday life". Moreover, each of the facets contains several factors.

Wuttke *et al. Empirical Res Voc Ed Train*   (2020) 12:18

Page 5 of 21

operationalisation of migration background varies. Furthermore, the studies are from different countries. Therefore, results are not fully comparable. Nonetheless the studies show, that migration background plays a significant role and can be systemized as follows: The *language spoken at home* and/or the *country of origin* has an effect on the test scores in financial literacy. While Ali et al. (2016) find a positive effect of a language other than the national one spoken at home, most other studies report significant negative effects on test scores in financial literacy if the language spoken at home is not the national one and/or if the participants themselves or their parents have a different country of origin than the country of residence (Driva et al. 2016; Gramatki 2017; Happ and Förster 2019; OECD 2014, 2017; Worthington 2006). Brown and Graf (2013) and Cameron et al. (2014) operationalize migration background as mother tongue of the participants. Both studies find that native speakers perform significantly better in financial literacy tests. Chen and Volpe (2002) can show that there is a small, yet not significant negative effect of migration background (nationality and race of participants) on the test scores. Khan et al. (2019) report that immigrants score lower than natives in financial literacy tests.

Educational background

A further influencing factor on financial literacy is the educational background. Young adults with a *higher level of education* such as a Master's or PHD degree, appear to have higher financial literacy (e.g. Lusardi and Mitchell 2014; Gramatki 2017; Ergün 2017). One study indicates that, if only a Bachelor's degree has been obtained, this has a negative effect on financial literacy (Ergün 2017). Other studies that do not include the degree in their explanatory models, but use the *number of attended school years*, achieve quite inconsistent results. Some of them refer to negative effects (Kaiser and Menkhoff 2017; Happ et al. 2018) others to positive effects of years at school (Gramatki 2017; Strömbäck et al. 2017). This is understandable as the extent of school attendance alone, without knowledge of the contents covered, is not very meaningful.

Opportunities to learn

There is some evidence that there might be an influence of *learning opportunities in finance-related topics* on the test scores. Studies that include this aspect usually focus either on formal learning opportunities such as attended courses in school or university or on informal learning opportunities such as discussions with parents, television reports on the topic, newspaper articles etc. Some studies also ask how respondents—independent of curricular offers—inform themselves about finance-related topics (informal learning opportunities, e.g. by reading newspapers, consulting counselling services, asking parents etc.). In this respect, studies generally point to a positive impact of learning opportunities on financial literacy (Rudeloff et al. 2019; Kaiser and Menkhoff 2017). Rudeloff et al. (2019) can show furthermore, that male and female participants profit differently from learning opportunities. Although it may seem trivial that students who had more learning opportunities perform better than those who had fewer opportunities, we will nevertheless test for measuring-invariance for this variable. The reason is that—if we can show that the test works structurally similar for both groups (more or fewer learning opportunities)—this provides a good basis for using the test before and

after an intervention and thus for reliably measuring knowledge gains. As a limitation, it must be said that this is only possible for the analysis of the effects of formal learning opportunities, only there can a clear distinction be made between intervention and control group. In the case of informal learning opportunities, there are clear limitations because it is not possible to form distinguishable groups.

In summary, previous studies suggest to some extent that gender, migration background, educational background and learning opportunities can cause differences in the level of financial literacy. However, it is unclear whether these differences are actually group differences or whether the different results are due to the test.

Methods

Research questions

The study addresses two research questions:

(1) Are there similar group differences in our test as in previous studies?
(2) Can possible differences actually be interpreted as different abilities in different groups or might they be the result of a test bias?

In order to answer these questions, we proceeded as follows: We examine whether we can assume measuring invariance for the groups, and whether a mean value comparison of the groups (female vs. male test takers, persons with or without a migration background, persons with a more or less pronounced educational background and persons with more vs. less previous opportunities to learn in financial topics) is thus possible. If this is the case, we will analyze whether there are group differences and how pronounced these differences are.

Sample

Data collection took place in 2016/2017. Tests with too many missing were removed from the sample. The resulting sample is N = 206.

149 participants of the sample have no migration background (51 = participants with migration background, 6 = no answer). We operationalise migration background via the mother tongue of the participants. Using only the information whether the parents are born in another country than Germany is not appropriate, since studies point to the fact, that the language spoken at home is more predictive (Happ et al. 2018). Regarding educational background, the sample can be divided into two groups: participants have either an academic background (university students, N = 105) or a vocational background (students in full time vocational schools or in dual vocational education, N = 101). With regard to previous opportunities to learn (OTL) in finance-related topics the test persons were asked to what extent financial topics were or were not addressed at school or during vocational education and training (0 = no addressing of financial literacy content, 1 = addressing of financial literacy content). Thus, a distinction is made between persons who have either had such OTL in financial topics during general and/or vocational education and training and those who have not yet had such OTL in their school career

Wuttke *et al. Empirical Res Voc Ed Train* (2020) 12:18

Page 7 of 21

(OTL in finance-related topics, N = 98, no OTL in finance-related topics, N = 102, 6 = no answer).

Instruments

To measure financial literacy, we use a SJT, which is based on a competence-oriented approach of financial literacy. In our test we distinguish the *dimensions* "financial literacy relevant for individual decisions" vs. "financial literacy relevant for societal decisions". Within these dimensions we model different *facets*, e.g. in the individual dimension the facet "planning and managing financial decisions of everyday life". Moreover, each of the facets contains several factors such as "saving money and building assets", "borrowing money" or "comparing and contracting insurances". In this paper, we particularly focus on the competence facet "planning and managing financial matters of everyday life" (for details on the basic assumptions and elaborations of the competence-oriented approach cf. Aprea and Wuttke 2016 as well as Leumann et al. 2016). The test for this facet consists of 22 items developed in a previous study (Wuttke and Aprea 2018). It comprises three factors that explain 39% of variance:

(1) Overview/control of one's own financial situation (9 items, max. 36 points, α = 0.754)
(2) Budgeting (6 items, max. 24 points, α = 0.573)[2]
(3) Handling of money (7 items, max. 28 points, α = 0.691)

Furthermore, we collected demographical data such as age, gender, migration background, educational background and the extent of (formal) OTL in finance-related topics. Figure 1 shows an example situation of the test.

Data analysis

Since the answering of the research questions presupposes an equivalence of the construct measurement in all groups, the measurement models for the groups to be investigated must at first be estimated and then simultaneously checked whether they are identical (or comparable) in all groups with regard to the factor loadings, the intercepts and, if applicable, the error terms of the indicators used (see Table 1).

The statistical analyses required for this purpose are carried out in AMOS statistical packages (Arbuckle 2016). For the measurement invariance check, a step-by-step approach is taken, starting with the least restrictive form of measurement invariance (configural measurement invariance) and gradually making the models more restrictive. The extent to which the restriction can be assumed is tested by means of the χ2 difference test. In addition, based on the rule of thumb according to Chen (2007), it will be

---

[2] In line with many other SJT the reliability of this scale is quite low. Catano, Brochu and Lamerson (2012) present results from a meta-analysis and report an average internal consistency of .46. They explain such a relatively low outcome with the fact that even people with a similar result in a construct may act differently in concrete situations. Moreover, a SJT represents a behaviour-based simulation of a criterion behaviour that is not a "pure" construct (Muck, 2013). Since many constructs require multiple skills, the search for unidimensional factors can be limiting. To act reasonably in financial decision situations, people need many dispositions, including financial knowledge, mathematical knowledge and the ability to delay gratification. According to Catano and colleagues (2012), applying reliability estimates other than internal consistency (e.g. retest reliability in a longitudinal perspective) may provide additional insights in this regard.

Wuttke *et al. Empirical Res Voc Ed Train*   (2020) 12:18

Page 8 of 21

You are in your first year of vocational education. You still live with your parents. In the following table, your monthly expenses and income sources are listed. You have no savings because you recently bought a fancy motor scooter.

Income		Expenses	
Account balance at the end of the previous month	€ 20	Gas for the scooter	€ 34
Monthly wage	€ 520	Monthly insurance for the scooter	€ 6
		Food and drink	€ 100
		Monthly flat rate for your smartphone	€ 30
		Parties, cinema, etc.	€ 100
Sum	€ 540	Sum	€ 270

Now you want to buy a new smartphone. You find an offer for one that costs € 530.
Please judge the following situations and decide how possible every decision is and how you would react in this situation.

(1 = very unlikely, 2 = unlikely, 3 = neither likely nor unlikely, 4 = likely, 5 = very likely)

	1	2	3	4	5
I have enough money and buy the smartphone immediately.					
I have a monthly surplus of € 270 and wait for one month to buy the smartphone.					
I notice that I do not have enough money. However, I really want to have the smartphone now. I am sure I will find an opportunity to borrow money. I buy the smartphone.					
I don't buy the smartphone now, but wait until I have enough savings to afford it.					

Fig. 1 Situation 1 with four answer options (items)

considered, if the CFI decreases by less than 0.02 units and the RMSEA increases by less than 0.015 units. The following models are examined.

1. Configural measurement invariance is the least restrictive form of measurement invariance and assumes an equivalent factor structure for the subgroups studied. This means that the same model with the same parameters is estimated in each subgroup but allows the factor loadings to assume different values. In the presence of this invariance, it can accordingly be assumed that the loading patterns of the same manifest variables on an identical latent variable in both subgroups do not differ significantly from each other.

Table 1 Overview of the sample

		N
Type of school or university	University, Business/Business Education	28
	University, Educational Science	32
	University, study programme not specified	45
	Full-time vocational schools, dual vocational education (drafting, carpentry, specialty in removal services)	101
Age	16–25 years, mean: 20.6 years	
Gender	108 = female, 93 = male, 5 = no answer	
Native language German	74%,	
		206

2. Metric measurement invariance (also called weak invariance) is more restrictive compared to configural invariance, because in addition the non-standardized factor loadings of the manifest variables are equated for the assumed groups. This means that not only the loading patterns, but also the factor loadings are tested for their equivalence. If metric invariance can be assumed for a measurement model, it is expected that the examined latent construct has the same meaning for the sub-groups.

3. Scalar measurement invariance (also called strong measurement invariance or tau equivalence) builds on the metric invariance by testing the additional assumption that the intercepts (regression constants) of the manifest variables are identical across the subgroups, i.e. invariant. If this assumption is confirmed, it can be assumed that there are no item-specific differences in difficulty between the subgroups and that the expression in the latent variable, i.e. potential differences in mean values, can be compared between the groups.

4. Strict invariance (also called invariance of the measurement errors) is present if, in addition to the scalar invariance, the equality of the measurement error variances over the examined subgroups can be assumed. If this most restrictive form of measurement invariance is not fulfilled, this points to potential differences in reliability between the sub-groups (Temme and Hildebrandt 2008).

Results

The mean values of the three factors of financial literacy for the respective subgroups are shown in Table 2. At first glance they indicate that persons with a *migration background* (*Control*: M = 29.25; *Budgeting*: M = 16.73; *Handling of money*: M = 17.24) seem to perform worse in the financial literacy test than persons without a migration background (*Control*: M = 30.50; *Budgeting*: M = 17.59; *Handling of money:* M = 18.74). In addition, respondents who attend university and thus have a *higher educational background* (high school diploma) seem to perform better in most factors of the here considered facet of financial literacy (*Control*: M = 30.37; *Budgeting*: M = 18.68; *Handling of money*: M = 17.42) than respondents who are enrolled in vocational education (*Control*: M = 29.99; *Budgeting*: M = 16.00; *Handling of money*: M = 19.22). However, the first unexpected results are evident for *gender*: women seem to perform better in all three factors of financial literacy (*Control*: M = 31.31; *Budgeting*: M = 17.91; *Handling of money*: M = 19.29) than men (*Control*: M = 29.07; *Budgeting*: M = 17.01; *Handling of money*: M = 17.40). Furthermore, the results indicate that a domain-related prior education (OTL in finance-related topics; *Control*: M = 30.20 vs. 30.18; *Budgeting*: M = 16.86 vs. 17.98; *Handling of money*: M = 17.82 vs. 19.01) does not result in higher scores in the knowledge test.

As a prerequisite for comparing the mean values between the different groups regarding the individual characteristics of the participants (gender, migration and educational background, learning opportunities), there must be at least metric invariance for the measurement models of the three factors of the considered facet of financial literacy.

**Table 2 Descriptive test results in the three factors of the financial literacy facet, separated by groups**

Group variable	N	Control M (SD)	N	Budgeting M (SD)	N	Handling of money M (SD)
Gender						
Male	81	29.07 (5.33)	83	17.01 (6.42)	80	17.40 (4.84)
Female	103	31.31 (4.58)	105	17.91 (5.72)	102	19.29 (3.70)
Migration background						
With	139	29.25 (6.15)	141	16.73 (6.02)	139	17.24 (4.42)
Without	44	30.50 (4.97)	45	17.59 (6.14)	42	18.74 (4.35)
Educational background						
Academic	100	30.37 (4.98)	102	18.68 (5.54)	99	17.42 (5.02)
Vocational	86	29.99 (5.61)	87	16.00 (6.43)	85	19.22 (3.53)
Opportunities to learn (OTL) in finance-related topics						
Attended	92	30.20 (5.50)	95	16.86 (6.14)	92	17.82 (4.49)
Not attended	91	30.18 (5.07)	91	17.98 (6.09)	89	19.01 (4.22)

A first look at the measurement models of the three factors for the entire data set shows that the assumed measurement model (see Fig. 2) satisfies the model criteria (*Control*: $\chi^2 = 39.124$, p $= 0.062$, df $= 27$, CFI $= 0.96$; RMSEA $= 0.047$, Pclose $= 0.537$; *Budgeting*: $\chi^2 = 24.711$, p $= 0.01$, df $= 11$, CFI $= 0.96$; RMSEA $= 0.078$, Pclose $= 0.119$; *Handling of money*: $\chi^2 = 17.189$, p $= 0.046$, df $= 9$, CFI $= 0.96$; RMSEA $= 0.066$, Pclose $= 0.248$, Aichholzer 2017).

Tables 3, 4, 5 and 6 show the successive tests of measurement invariance across the groups (gender, migration background, educational background and OTL in finance-related topics).

Gender

Within the framework of the configural invariance model for the three factors of the considered facet of financial literacy, all factor loadings, intercepts and error terms were freely estimated across both genders (see Table 3). The fit statistics first show that the model fit can be assumed for all three configural invariance models (*Control*: $\chi^2 = 61.07$, p $= 0.098$, df $= 48$, CFI $= 0.95$; RMSEA $= 0.037$, Pclose $= 0.778$; *Budgeting*: $\chi^2 = 41.103$, p $= 0.008$, df $= 22$, CFI $= 0.94$; RMSEA $= 0.066$, Pclose $= 0.185$; *Handling of money*: $\chi^2 = 36.555$, p $= 0.006$, df $= 18$, CFI $= 0.89$; RMSEA $= 0.072$, Pclose $= 0.131$) even if the CFI in particular is too low for the factor *Handling of money*. It can thus be assumed that the three factors are conceptualized in a similar way in both groups. With regard to the equality restriction on the factor loadings, which were set within the framework of metric measurement invariance, results depend on the factor, however. While for the factors *Control* and *Budgeting* the model fit does not decline significantly (ΔCFI $\leq$ |.02|, ΔRMSEA $\leq$ .015) and it can therefore be deduced that the unit of measurement of the two scales is identical for female and male test participants, a decline in model fit is found for the factor *Handling of money* with this model restriction ($\chi^2 = 47.516$, p $= 0.003$, df $= 24$, CFI $= 0.86$; RMSEA $= 0.07$, Pclose $= 0.122$). Only by releasing the factor loadings of item 6 (this was done on the

Fig. 2 Measurement model of the three facets of financial literacy
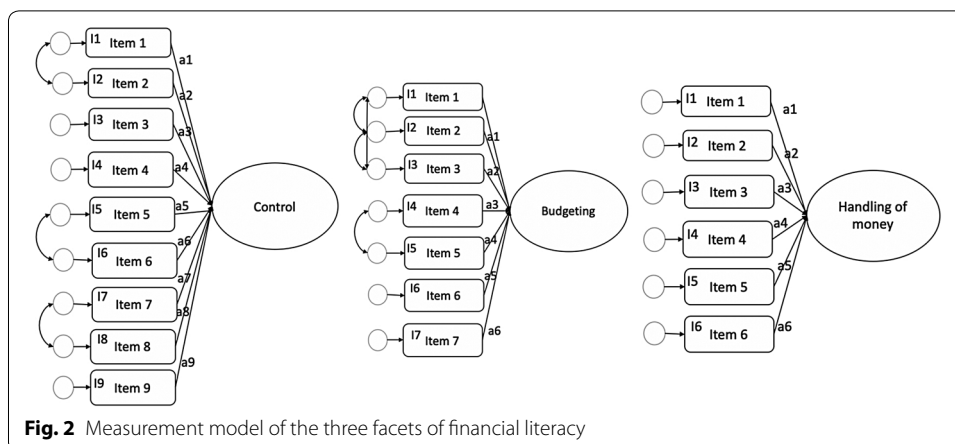
basis of the size of the modification index) a similar model fit as for the configural measurement invariance can be achieved ($\chi^2 = 43.184$, p $= 0.007$, df $= 23$, CFI $= 0.89$; RMSEA $= 0.066$, Pclose $= 0.175$).

For the test of strong measuring invariance not only the factor loadings but also the intercepts of the individual items used for the latent scales were equated for male and female subjects. The decline of the model fits (ΔCFI $> |.02|$) illustrates that this model restriction cannot be assumed for any of the factors. Here too, however, it is worth investigating whether the strong invariance can be achieved at least partially. For this purpose, the intercept of item 7 for the factor *Control*, the intercept of the item 6 for the factor *Budgeting* and the intercept of item 5 for the factor *Handling of money* was freely estimated, while the other factor loadings were restricted to equality over the two assumed subgroups. The test statistics of the model comparison show that partial measurement invariance for the factors *Control* ($\chi^2 = 86.701$, p $= 0.002$, df $= 65$, CFI $= 0.91$; RMSEA $= 0.041$, Pclose $= 0.736$) and *Budgeting* ($\chi^2 = 59.521$, p $= 0.006$, df $= 35$, CFI $= 0.92$; RMSEA $= 0.059$, Pclose $= 0.259$) can be achieved with this approach. For the factor *Handling of money* the model fit is still not given ($\chi^2 = 52.35$, p $= 0.002$, df $= 27$, CFI $= 0.85$; RMSEA $= 0.069$, Pclose $= 0.128$). For the factor *Budgeting* it can be shown that even a strict measurement invariance can be assumed ($\chi^2 = 70.489$, p $= 0.004$, df $= 42$, CFI $= 0.91$; RMSEA $= 0.058$, Pclose $= 0.265$). Against this background, only the factors *Control* and *Budgeting* provide the prerequisites for a meaningful comparison of the mean values of these scales.

When we analyse group differences by calculating t-tests between the groups (male/female) they only show one significant difference, namely for the factor *Control to the advantage of female subjects* (*Control*: t(182) $= -$ 3,058, p $= 0.003$; *Budgeting*: t(186) $= -$ 1,018, p $= 0.310$).

Migration background

Within the framework of the configural invariance model for the three factors of the considered facet of financial literacy all factor loadings, intercepts and error terms were freely estimated across the two groups of subjects with and without migration background (see Table 4). The fit statistics show that the model fit can be assumed for

Wuttke *et al. Empirical Res Voc Ed Train* (2020) 12:18

Page 12 of 21

Table 3 Measurement invariance for the subgroup gender

Model	Control ΔDF	ΔCMIN	p	Budgeting ΔDF	ΔCMIN	p	Handling ΔDF	ΔCMIN	p
Configural	χ² = 61.07, p = .098, df = 48, CFI = .95; RMSEA = .037, Pclose = .778			χ² = 41.103, p = .008, df = 22, CFI = .94; RMSEA = .066, Pclose = .185			χ² = 36.555, p = .006, df = 18, CFI = .89; RMSEA = .072, Pclose = .131		
Metric	9; χ² = 72.422, p = .082, df = 57, CFI = .94; RMSEA = .037, Pclose = .801	11.352	0.252	7; χ² = 48.483, p = .013, df = 29, CFI = .94; RMSEA = .058, Pclose = .298	7.379	0.39	6; χ² = 47.516, p = .003, df = 24, CFI = .86; RMSEA = .07, Pclose = .122	10.961	0.09
Partial metric	—	—	—	—	—	—	5; χ² = 43.184, p = .007, df = 23, CFI = .89; RMSEA = .066, Pclose = .175[c]	6,629	0.25
Scalar	18; χ² = 90.36, p = .025, df = 66, CFI = .90; RMSEA = .043, Pclose = .686	29.29	0.045	14; χ² = 63.041, p = .004, df = 36, CFI = .91; RMSEA = .061, Pclose = .214	21.937	0.08	10; χ² = 54.828, p = .002, df = 28, CFI = .84; RMSEA = .069, Pclose = .116	18.273	0.051
Partial scalar	17; χ² = 86.701, p = .002, df = 65, CFI = .91; RMSEA = .041, Pclose = .736[a]	25.631	0.081	13; χ² = 59.521, p = .006, df = 35, CFI = .92; RMSEA = .059, Pclose = .259[b]	18.418	0.142	9; χ² = 52.35, p = .002, df = 27, CFI = .85; RMSEA = .069, Pclose = .128[d]	15.795	0.071
Strict	26; χ² = 114.448, p = .002, df = 74, CFI = .91; RMSEA = .052, Pclose = .398	53.378	0.001	20; χ² = 70.489, p = .004, df = 42, CFI = .91; RMSEA = .058, Pclose = .265	29.385	0.08	15; χ² = 82.075, p < .001, df = 33, CFI = .72; RMSEA = .086, Pclose = .007	45.52	<.001

[a] Intercepts are freely estimated for Item 7

[b] Intercepts are freely estimated for Item 5

[c] Factor loadings are freely estimated for Items 6

[d] Intercepts are freely estimated for Item 6

Wuttke *et al. Empirical Res Voc Ed Train*   (2020) 12:18

Page 13 of 21

Table 4 Measurement invariance for the subgroup migration background

Model	Control			Budgeting			Handling		
	ΔDF	ΔCMIN	p	ΔDF	ΔCMIN	p	ΔDF	ΔCMIN	p
Configural	$\chi^2 = 73.91$, p = .01, df = 48, CFI = .92; RMSEA = .052, Pclose = .415	–	–	$\chi^2 = 45.016$, p = .003, df = 22, CFI = .93; RMSEA = .073, Pclose = .104	–	–	$\chi^2 = 25.3$, p = .117, df = 18, CFI = .96; RMSEA = .045, Pclose = .537	–	–
Metric	9 $\chi^2 = 81.451$, p = .018, df = 57, CFI = .92; RMSEA = .047, Pclose = .578	7.538	0.581	7 $\chi^2 = 52.405$, p = .005, df = 29, CFI = .93; RMSEA = .064, Pclose = .193	7.389	0.389	6 $\chi^2 = 31.954$, p = .128, df = 24, CFI = .96; RMSEA = .041, Pclose = .629	6.654	0.354
Partial metric	– –			– –			– –		
Scalar	18 $\chi^2 = 89.176$, p = .03, df = 66, CFI = .93; RMSEA = .042, Pclose = .71	15.263	0.644	14 $\chi^2 = 59.748$, p = .008, df = 36, CFI = .93; RMSEA = .058, Pclose = .292	14.732	0.397	12 $\chi^2 = 39.726$, p = .11, df = 30, CFI = .95; RMSEA = .04, Pclose = .659	14.426	0.274
Partial scalar	– –			– –			– –		
Strict	27 $\chi^2 = 123.733$, p < .001, df = 75, CFI = .85; RMSEA = .057, Pclose = .242	49.819	0.005	21 $\chi^2 = 76.912$, p = .0012, df = 43, CFI = .90; RMSEA = .063, Pclose = .165	3,896	0.06	18 $\chi^2 = 69.846$, p = .001, df = 36, CFI = .82; RMSEA = .069, Pclose = .097	44.546	< .001

configural, weak and strong invariance models, whereas the strict invariance can only be accepted for *Budgeting* since the model fit does not decline significantly (ΔCFI $\leq$ |.02|, ΔRMSEA $\leq$ 0.015). This means that the three factors in both groups are conceptualized in a similar way, the factor loadings can be assumed to be similar and also the items can be assumed to be similarly difficult for both groups. The p-values also show here that the increase in model restriction in relation to the degrees of freedom gained is not statistically significant.

The results of the t-tests indicate that for the factors *Control* and *Budgeting* there are no significant group differences based on the migration background (*Control*: t(181) = − 1. 367, p = 0.17; *Budgeting*: t(184) = 0.817, p = 0.42), but that there is a significant group difference for the factor *Handling of money* (t(179) = − 1.955, p = 0.05) *to the advantage of those subjects who have no migration background.*

Educational background

For the two subgroups academic vs. vocational background in a first step the assumption of configural measurement invariance was tested for all three factors of the considered facet of financial literacy (see Table 5). The fit statistics show that the model fit is only suitable for the factors *Control* (χ^2 = 68.585, p = 0.14, df = 57, CFI = 0.96; RMSEA = 0.032, Pclose = 0.881) and *Budgeting* (χ^2 = 48.616, p = 0.013, df = 29, CFI = 0.94; RMSEA = 0.058, Pclose = 0.306). This means that these two factors are conceptualized in a similar way in both groups. However, for the factor *Handling of money* not even the conceptualization of the model seems acceptable (χ^2 = 41.696, p = 0.001, df = 18, CFI = 0.87; RMSEA = 0.08, Pclose = 0.058). The examination of metric and scalar measuring invariance assumption shows that this can be confirmed for the factor *Control*, since the model fit does not significantly decline with increasing restriction of the models (ΔCFI $\leq$ |.02|, ΔRMSEA $\leq$ 0.015). For the factor *Budgeting* a weak measurement invariance can be assumed (χ^2 = 48.616, p = 0.013, df = 29, CFI = 0.94; RMSEA = 0.058, Pclose = . 306), however, for the scalar invariance, Intercepts of item 1 and Item 5 must be additionally estimated freely in order to achieve the model fit (χ^2 = 58.594, p = 0.005, df = 34, CFI = 0.92; RMSEA = 0.06, Pclose = 0.253), so that only a partial strong invariance can be found here. The further equation of the error terms of the two groups for the factors *Control* and *Budgeting* produces a significant decline in the model fit, so that no strict measurement invariance can be assumed.

Against the background, that a (partial) strong measurement invariance could be achieved for the two factors *Control* and *Budgeting*, the comparison of the mean values of test persons with an academic vs. a vocational background is permissible. T-tests then show that the test takers *with an academic background perform significantly better in the factor Budgeting than those with a vocational background* (*Control*: t(184) = − 0.491, p = 0.62; *Budgeting*: t(187) = − 3.073, p = 0.002).

OTL in finance-related topics

With reference to the comparison of the measurement models between subjects with OTL in finance-related topics and subjects without OTL in finance-related topics, the presence of configural, weak, strong and strict measurement invariance was tested

Wuttke *et al. Empirical Res Voc Ed Train*   (2020) 12:18

Page 15 of 21

Table 5 Measurement invariance for the subgroup educational background

Model	Control			Budgeting			Handling		
	Δ DF	Δ CMIN	p	ΔDF	Δ CMIN	p	Δ DF	Δ CMIN	P
Configural	χ^2 = 59.442, p = .124, df = 48, CFI = .96; RMSEA = .034, Pclose = .825			χ^2 = 39.888, p = .011, df = 22, CFI = .94; RMSEA = .063, Pclose = .226			χ^2 = 41.696, p = .001, df = 18, CFI = .87; RMSEA = .08, Pclose = .058		
Metric	9 χ^2 = 68.585, p = .14, df = 57, CFI = .96; RMSEA = .032, Pclose = .881	9.143	0.424	7 χ^2 = 48.616, p = .013, df = 29, CFI = .94; RMSEA = .058, Pclose = .306	8.728	0.273			
Partial metric	– –	–	–	– –	–	–			
Scalar	18 χ^2 = 81.107, p = .1, df = 66, CFI = .95; RMSEA = .033, Pclose = .876	21.665	0.247	14 χ^2 = 72.947, p < .001, df = 36, CFI = .88; RMSEA = .071, Pclose = .071	33.059	0.003			
Partial scalar	– –	–	–	12 χ^2 = 58.594, p = .005, df = 34, CFI = .92; RMSEA = .06, Pclose = .253[a]	18.706	.096			
Strict	27 χ^2 = 105.149, p = .012, df = 75, CFI = .90; RMSEA = .044, Pclose = .666	45.706	0.014	19 χ^2 = 83.865, p < .001, df = 41, CFI = .86; RMSEA = .072, Pclose = .054	43.977	.001			

[a] Intercepts are freely estimated for Item 1 and Item 5

(see Table 6). The fit statistics show that the model fit can initially be assumed for the first two factors *Control* and *Budgeting* for configural, weak, strong and strict invariance, since the model fit does not decline significantly with increasing restriction (ΔCFI $\leq$ |.02|, ΔRMSEA ≤ 0.015). This means that these two factors are conceptualized in a similar way in both groups, the loadings of the parameters can be assumed to be similar and the items can also be assumed to be similarly difficult for both groups. The p-values also show here that the increase in model restriction in relation to the degrees of freedom gained is not statistically significant. Against this background, the conditions for a meaningful comparison between the mean values of the two groups are given for these two factors.

For the third factor *Handling of money*, however, initially only a configural measurement invariance could be assumed ($\chi^2 = 28.393$, p $= 0.056$, df $= 18$, CFI $= 0.95$; RMSEA $= 0.054$, Pclose $= 0.393$), since the model fit declines with the determination of the factor loadings ($\chi^2 = 45.642$, p $= 0.005$, df $= 24$, CFI $= 0.89$; RMSEA $= 0.067$, Pclose $= 0.157$). Accordingly, it is helpful at this point to look at the model fit for the case of partial determination of the factor loadings. For this purpose, the factor loading for item 1 is freely estimated, which leads to a good model fit ($\chi^2 = 43.466$, p $= 0.04$, df $= 29$, CFI $= 0.93$; RMSEA $= 0.05$, Pclose $= 0.465$) and also allows a mean value comparison between the subgroups for the factor *Handling of money*. However, for the third factor no strict measurement invariance can be assumed ($\chi^2 = 69.537$, p < 0.001, df $= 35$, CFI $= 0.83$; RMSEA $= 0.071$, Pclose $= 0.081$). The investigation of potential mean differences for persons with OTL in finance-related topics and persons without these OTL does not point to significant differences (*Control*: t(181) $= 0.025$, p $= 0.98$; *Budgeting*: t(184) $= -1.242$, p $= 0.216$; *Handling money*: t(179) $= -1.846$, p $= 0.067$).

Discussion

In this study, we presented a newly developed SJT for measuring financial literacy in a competence oriented way. This type of format was chosen because of its closeness to related behavior in real life situations. With regard to this test, we asked the questions if (1) the test demonstrates similar group differences (i.e. female vs. male test takers, persons with or without a migration background, persons with a more or less pronounced educational background and persons with vs. without previous opportunities to learn in financial topics) as in many other studies, and (2) if possible differences can actually be interpreted as different abilities in different groups or might rather be the result of a test bias. To answer these questions, we examined whether measuring invariance for the groups can be assumed, and a mean value comparison of the groups is thus possible. If this was the case, we analyzed whether there are group differences and how pronounced these differences are.

With regard to gender differences, the results can be summarized as follows: Only for the factors *Control* and *Budgeting* does the test fulfil the prerequisites for a meaningful comparison of the mean values. In this context, the t-test only shows a significant difference for the factor *Control* to the advantage of female subjects (*Control*: t(182) $= -3.058$, p $= 0.003$; *Budgeting*: t(186) $= -1.018$, p $= 0.310$). The differences with disadvantages for female participants reported in many studies (see chapter two

Wuttke *et al. Empirical Res Voc Ed Train*   (2020) 12:18

Page 17 of 21

Table 6 Measurement invariance for the subgroup with OTL in finance-related topics

Model	Control			Budgeting			Handling		
	ΔDF	ΔCMIN	p	ΔDF	ΔCMIN	p	ΔDF	ΔCMIN	p
Configural	χ^2=64.115, p=.06, df=48, CFI=.95; RMSEA=.041, Pclose=.695			χ^2=45.615, p=.002, df=22, CFI=.93; RMSEA=.073, Pclose=.10			χ^2=28.393, p=.056, df=18, CFI=.95; RMSEA=.054, Pclose=.393		
Metric	9	12.825	0.171	7	5.336	0.619	6	17.249	0.008
	χ^2=76.94, p=.04, df=57, CFI=.93; RMSEA=.042, Pclose=.695			χ^2=50.95, p=.007, df=29, CFI=.93; RMSEA=.062, Pclose=.227			χ^2=45.642, p=.005, df=24, CFI=.89; RMSEA=.067, Pclose=.157		
Partial metric	–	–	–	–	–	–	5	7.647	0.177
	–			–			χ^2=36.04, p=.04, df=23, CFI=.94; RMSEA=.054, Pclose=.399[a]		
Scalar	18	18.749	0.407	14	8,121	0,883	11	15.073	0.179
	χ^2=82.864, p=.078, df=66, CFI=.94; RMSEA=.036, Pclose=.836			χ^2=53.736, p=.029, df=36, CFI=.95; RMSEA=.05, Pclose=.475			χ^2=43.466, p=.04, df=29, CFI=.93; RMSEA=.05, Pclose=.465		
Partial scalar	–	–	–	–	–	–	–	–	–
	–			–			–		
Strict	27	27.636	0.43	21	10.18	0.976	17	41.143	0.001
	χ^2=91.751, p=.092, df=75, CFI=.94; RMSEA=.034, Pclose=.886			χ^2=55.795, p=.091, df=43, CFI=.96; RMSEA=.039, Pclose=.73			χ^2=69.537, p<.001, df=35, CFI=.83; RMSEA=.071, Pclose=.081		

[a] Factor loading are freely estimated for Item 1

of this paper) are not found in our study. This is of course only true for the scales that allow a comparison.

The following result can be summarized for the migration background: For all three factors, the prerequisite for testing differences in mean values between the subgroups is given. The results of the t-tests indicate that for the factors *Control* and *Budgeting* there are no significant group differences regarding the migration background (*Control*: t(181) = − 1.367, p = 0.17; *Budgeting*: t(184) = 0,817, p = 0.42), but this is the case for the factor *Handling of money* (t(179) = − 1.955, p = 0.05). Results show an advantage for those subjects who do not have a migration background. Again, it can be stated that the disadvantages for participants with a migration background that are reported in many studies cannot be found in our study.

As far as the educational background is concerned, we can summarize that a (partial) strong invariance in measurement can be achieved for the two factors *Control* and *Budgeting*. For these factors, a comparison of the mean values of the scales between test persons with a vocational vs. an academic educational background is permissible. T-tests show that participants with an academic background perform significantly better in *Budgeting* than those who are enrolled in the vocational school system (*Control*: t(184) = − 0.491, p = 0.62; *Budgeting*: t(187) = − 3.073, p < 0.01). This is in line with the majority of studies reported above and confirms that the educational background plays an important role with regard to the extent of financial literacy. However, it is unexpected that in the factore *Handling of money* of financial literacy young adults with an academic background perform worse than those with a vocational background. Even if this cannot be tested inferentially due to the lack of invariance in measurement, the mean difference of ∇M = 1.80 is noticeable. This can possibly be explained by the fact that while young adults in vocational education directly experience the handling of money through their salary, students only occasionally receive regular income through e.g. mini-jobs, which they have to manage.

With reference to the comparison of the measurement models between participant with OTL in finance-related topics and those without OTL in finance-related topics, the model fit can initially be assumed for the first two factors *Control* and *Budgeting* for configural, weak, strong and strict invariance. Against this background, the conditions for a meaningful comparison between the mean values of the two groups are given for these two factors.

For the third factor *Handling of money*, however, initially only a configural measurement invariance could be assumed. When the factor loading for item 1 is freely estimated, this leads to a good model fit which also allows a mean value comparison between the subgroups for the factor *Handling of money*. The investigation of potential mean differences for persons with OTL in finance-related topics and persons without these OTL does not point to significant differences (*Control*: t(181) = 0.025, p = 0.98; *Budgeting*: t(184) = − 1.242, p = 0.216; *Handling money*: t(179) = − 1.846, p = 0.067).

All results considered we find that in contrast to the findings of many previous studies, the analysis of the mean differences does not show any systematic deficits in financial literacy for specific groups.

With regard to the quality of the test, the following result can be summarized: The analysis of the measurement invariance shows that the developed test for the factors *Control of*

Wuttke *et al. Empirical Res Voc Ed Train*   (2020) 12:18

Page 19 of 21

one's own financial situation and *Budgeting* shows measurement invariance for all groups considered, group comparisons are thus possible and potential differences are not due to a test bias. For the factor *Handling of money*, we can only assume measurement invariance with regard to learning opportunities.

Conclusion and further studies

Given the above-mentioned results, it seems that the SJT format offers a viable way to ensure test fairness in financial literacy assessment. However, we are aware that our study has some limitations and therefore interpretations need to be cautious. One of these limitations is the small sample size. Moreover, other tests that have produced specific group differences have not been evaluated in comparison. Given these limitations, we cannot prove the assumption that differences in previous studies may be caused by a test bias rather than by different abilities or interests of the groups considered. What we can show is, however, that the test we have developed does not pose this problem. Since we cannot assume measurement invariance for all factors, parts of the test have to be revised. This is the case for the factor *Handling of money*. For this purpose, a study using the method of thinking aloud is planned, which should give us information on how to revise the items of the facet. In addition, those items that had to be freely estimated need to be revised. This concerns the items 1 and 5 from the *Budgeting* factor and item 7 from the *Control* factor. A think aloud study would be helpful here as well, in order to discover how these items can be changed and adapted accordingly. After revision, the test will be analyzed again. Finally, the present study has not yet been able to test all facets of the underlying construct (Aprea and Wuttke 2016; Leumann et al. 2016).

Author details
[1] Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60629 Frankfurt, Germany. [2] University of Mannheim, L 4, 1, 68161 Mannheim, Germany.

References:
Aprea C, Wuttke E (2016) Financial literacy of adolescent and young adults: setting the course for a competence-oriented assessment approach. In: Aprea C, Wuttke E, Breuer K, Keng NK, Da-vies P, Greimel-Fuhrmann B, Lopus J (eds) International Handbook of Financial Literacy. Springer, Singapore, pp 397–414
Aprea C, Wuttke E, Breuer K, Keng NK, Davies P, Greimel-Fuhrmann B, Lopus J (2016) Finan-cial literacy in the twenty-first century: An introduction to the International Handbook of Finan-cial Literacy. In: Aprea C, Wuttke E, Breuer K, Keng NK, Davies P, Greimel-Fuhrmann B, Lopus J (eds) International Handbook of Financial Literacy. Springer, Singapore, pp 1–4
Aichholzer J (2017) Einführung in lineare Strukturgleichungsmodelle mit Stata [Introduction to linear structural equation models with Stata]. Springer, Berlin
Arbuckle JL (2016) IBM SPSS AMOS 24 user´s guide. SPSS Inc., Chicago

Agnew S, Harrison N (2015) Financial literacy and student attitudes to debt: a cross national study examining the influence of gender on personal finance concepts. J Retail Consum Serv 25:122–129. https://doi.org/10.1016/j.jretconser.2015.04.006

Ali P, Anderson M, McRae C, Ramsay I (2016) The financial literacy of young people: socio-economic status, language background and the rural-urban chasm. Aust Int J Rural Educ 26(1):54–66

Allianz (2017) When will the penny drop? Money, financial literacy and risk in the digital age. https://gflec.org/initiatives/money-finlit-risk

Almenberg J, Dreber A (2015) Gender, stock market participation and financial literacy. Econ Lett 137:140–142

Asarta C, Butters RB, Thompson E (2014) The gender question in economic education: is it the teacher or the test? Perspect Econ Educ Res 9:1–19

Atkinson A, Messy F (2012) Measuring financial literacy: results of the OECD/International network on financial education (INFE) pilot study. OECD Working Papers on Finance, Insurance and Private Pensions, No. 15. Paris: OECD Publishing.

Bannier C, Neubert M (2016) Gender differences in financial risk taking: the role of financial literacy and risk tolerance. Econ Lett 145:130–135

Braunstein S, Welch C (2002) Financial Literacy: an overview of practice, research and policy. Federal Reserve Bulletin 88:445–457

Brown M, Graf, R (2013) Financial literacy and retirement planning in Switzerland. Numeracy, 6(2), Art. 6. https://scholarcommons.usf.edu/numeracy/vol6/iss2/art6.

Bucher-Koenen T, Lamla B (2014) The long shadow of socialism: on east-west german differences in financial literacy. Munich Center for Economics of Aging (meA) Discussion Papers 282–2014. München: Max-Planck-Institut für Sozialrecht und Sozialpolitik. Abgerufen von https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2497391

Bucher-Koenen T, Lusardi A, Alessie R, van Rooij M (2017) How financially literate are women? an overview and new insights. J Consumer Affairs 51(2):255–238

Cameron M, 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

Catano VM, Brochu A, Lamerson CD (2012) Assessing the reliability of situational judgment tests used in high-stakes situations. Int J Selection Asses 20:333–346

Chen H, Volpe RP (2002) Gender differences in personal financial literacy among college students. Fin Serv Rev 11:289–307

Chen FF (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Model 14:464–504

Dilworth-Anderson P, Hendrie HC, Manly JJ, Khachaturian AS, Fazio S (2008) Diagnosis and assessment of Alzheimer's Disease in diverse populations. Alzheimer's Dementia 4(4):305–309

Driva A, Lührmann M, Winter J (2016) Gender differences and stereotypes in financial literacy: off to an early start. Econ Lett 146:143–146

Ergün K (2017) Financial literacy among university students: a study in eight European countries. Int J Consumer Stud 42:2–15. https://doi.org/10.1111/ijcs.12408

Fernandes D, Lynch JG, Netemeyer RG (2014) Financial literacy, financial education, and downstream financial behaviors. Manage Sci 60(8):1861–1883. https://doi.org/10.1287/mnsc.2013.1849

Filipiak U, Walle JM (2015) The financial literacy gender gap: a question of nature or nurture? (No. 176). Courant Research Centre: Poverty, Equity and Growth-Discussion Papers.

Fonseca R, Mullen K, Zamarro G, Zussumopoulos J (2012) What explains the Gender Gap in Financial Literacy? the role of household decision making. J Consumer Affairs 46(1):90–106

Förster M, Happ R, Maur A (2018) The relationship among gender, interest in financial topics and understanding of personal finance. Empirische Pädagogik 32(3/4):292–308

Gramaţki I (2017) A comparison of financial literacy between native and immigrant school students. Educ Econ 25(3):304–322. https://doi.org/10.1080/09645292.2016.1266301

Greimel-Fuhrmann B, Silgoner M (2018) Analyzing the Gender Gap in Financial Literacy. Int J Infonomics (IJI) 11(3):1779–1787

Happ R, Förster M (2019) The relationship between migration background and knowledge and understanding of personal finance of young adults in Germany. Int Rev Econ Educ 30(1):1–14

Happ R, Förster M, Grein M, Bültmann A (2018) The importance of controlling for language skills when assessing the correlation between young adults' knowledge and understanding of personal finance and migration background. Empirische Pädagogik 32(3/4):329–348

Hasler A, Lusardi A (2017) The gender gap in financial literacy: a global perspective. The George Washington University School of Business, Global Financial Literacy Excellence Centre

Hill AT, Asarta CJ, (2016) Gender and student achievement in personal finance: evidence from keys to financial success. In: Aprea C, Wuttke E, Breuer K, Keng NK, Davies P, Greimel-Fuhrmann B, Lopus J (ed), International Handbook of Financial Literacy. Springer, Singapore, pp. 545–567. https://doi.org/https://doi.org/10.1007/978-981-10-0360-8_35.

Hung AA, Parker AM, Yoong JK (2009) Defining and Measuring Financial Literacy – RAND Working Paper Series WR-708. Santa Monica: RAND Corporation.

Huston SJ (2010) Measuring financial literacy. J Consum Aff 44(2):296–316

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. https://doi.org/10.1093/wber/lhx018

Kahmann J (2014) Entwicklung und Validierung eines Situational Judgement Tests (SJT) zur Erfassung sozialer Kompetenzen von Studienplatzbewerbern und –interessenten der Human- und Zahnmedizin [Development and validation of a situational judgement test for assessing social competences of university applicants in human and dental medicine]. Doctoral dissertation, Ruprecht-Karls-Universität, Heidelberg.

Khan M, Ferrer I, Lee Y, Rothwell D (2019) Understanding the financial knowledge of immigrants in Canada. In Strangers in New Homeland, Winnipeg

Killins RN (2017) The financial literacy of Generation Y and the influence that personality traits have on financial knowledge: evidence from Canada. Fin Serv Rev 26(2):143–165

Leumann S, Heumann M, Syed F, Aprea C (2016) Towards a comprehensive financial literacy framework: voices from stakeholders in european vocational education and training. In: Wuttke E, Seifried J, Schumann S (eds) Economic competence and financial literacy of young adults status and challenges. Opladen, Barbara Budrich, pp 19–39

Lührmann M, Serra-Garcia M, Winter J (2015) Teaching teenagers in finance: does it work? J Bank Finance 54:160–174

Lusardi A, Mitchell OS (2011) Financial literacy around the world: an overview. J Pension Econ Fin 10(4):497–508

Lusardi A, Mitchell O, Curto V (2014) Financial literacy and financial sophistication in the older population. J Pension Econ Fin 13(4):347–366. https://doi.org/10.1017/S1474747214000031

Lusardi A, Mitchell OS (2014) The economic importance of financial literacy: theory and evidence. J Econ Literature 52(1):5–44

Mahdavi M, Horton N (2012) Financial knowledge among educated women: room for improvement. J Consumer Affairs 48(2):403–417

McDaniel MA, Hartmann NS, Whetzel DL, Grubb WL (2007) Situational judgment tests, response instructions, and validity: a meta-analysis. Pers Psychol 60:63–91

Melikyan ZA, Agranovich AV, Puente AE (2019) Fairness in psychological testing. In: Goldstein G, Allen DN, Deluca J (eds) Handbook of Psychological Assessment, 4th edn. Academic Press, Cambridge, pp 551–564

Mitchell OS, Lusardi A (2015) Financial literacy and economic outcomes: evidence and policy implications. J Retirement 3(1):107–114

Muck PM (2013) Evidenzbasierte Entwicklung von Situational Judgment Tests: Konzeptionelle Überlegungen und empirische Befunde [Evidence-based development of situational judgement tests: conceptual considerations and empirical findings]. Zeitschrift für Arbeits- und Organisationspsychologie 57:185–205

OECD (2014) "What Are Tertiary Students Choosing to Study?" Education Indicators in Focus, no. 19.

OECD (2017) G20/OECD INFE report on adult financial literacy in G20 countries. https://www.oecd.org/daf/fin/financial-education/G20-OECD-INFE-report-adult-financial-literacy-in-G20-countries.pdf

OECD (2018) Education at a Glance 2018: OECD Indicators. OECD Publishing, Paris. https://doi.org/10.1787/eag-2018-en

Preston AC, Wright RE (2019) Understanding the gender gap in financial literacy: evidence from Australia. Econ Record 95(S1):1–29

Reardon SF, Kalogrides D, Fahle EM, Podolsky A, Zárate RC (2018) The relationship between test item format and gender achievement gaps on Math and ELA Tests in 4th and 8th Grade. Educ Res 47(5):284–294

Remund DL (2010) Financial literacy explicated: the case for a clearer definition in an increasingly complex economy. J Consumer Affairs 44(2):276–295

Rinaldi E (2017) Gender differences in financial literacy in Italy. Exploratory Explan 84:89–95

Rudeloff M (2019) Der Einfluss informeller Lerngelegenheiten auf die Finanzkompetenz von Lernenden am Ende der Sekundarstufe I [The impact of informal learning opportunities on the financial literacy of students at the end of lower secondary education]. Springer, Wiesbaden

Rudeloff M, Brahm T, Pumptow M (2019) Does gender matter for the use of learning opportunities? Potential explanation for the gender gap in financial literacy. Citizenship Soc Econ Educ 18(3):128–142

Schuhen M, Schürkmann S (2014) Construct validity of financial literacy. Int Rev Econ Educ 16:1–11

Schürkmann S (2017) FILS – Financial Literacy Study Validierung und Analyse einer schülerorientierten financial literacy mittels der Methodiken des Strukturgleichungsmodells und des Rasch-Modells [FILS - Financial Literacy Study Validation and analysis of student-oriented financial literacy using the methods of the structural equation model and the Rasch model]. De Gruyter Oldenbourg.

Siegfried C, Wuttke E (2019) Are multiple choice items unfair? And if so, for whom? Citizenship Soc Econ Educ 18(3):123–127

Strömbäck C, Lind T, Skagerlund K, Västfjäll D, Tinghög G (2017) Does self-control predict financial behavior and financial well-being? J Behav Exper Fin 14:30–38

Tang N, Baker A, Peter P (2015) Investigating the disconnect between financial knowledge and behavior: the role of parental influence and psychological characteristics in responsible financial behaviors among young adults. J Consumer Affairs 2:376–406

Temme D, Hildebrandt L (2008) Gruppenvergleiche bei hypothetischen Konstrukten – Die Prüfung der Übereinstimmung von Messmodellen mit der Strukturgleichungsmethodik [Group comparisons for hypothetical constructs - Testing the agreement of measurement models with the structural equation methodology]. In Schriftenreihe Economic Risk SFB 649 Papers, Discusson Paper 2008–042.

Walstad WB, Rebeck K, MacDonald RA (2010) The effects of financial education on the financial knowledge of high school students. J Consum Aff 44(2):336–357

Whetzel DL, McDaniel MA (2009) Situational judgment tests: an overview of current research. Hum Resour Manag Rev 19:188–202

Woodyard A, Robb C (2012) Financial Knowledge and the Gender Gap. J Fin Ther 3(1):16. https://doi.org/10.4148/jft.v3i1.1453

Worthington AC (2006) Predicting financial literacy in Australia. Financial Services Rev 15(1):59–79

Wuttke E, Aprea C (2018) A situational judgement approach for measuring young adults' financial literacy. Empirische Pädagogik: EP 32:272–292

Publisher's Note