THE JOURNAL OF ECONOMIC EDUCATION
2017, VOL. 48, NO. 2, 123–135
http://dx.doi.org/10.1080/00220485.2017.1285737

Routledge
Taylor & Francis Group

FEATURES AND INFORMATION

Using the U.S. *Test of Financial Literacy* in Germany—Adaptation and validation

Manuel Förster[a], Roland Happ[a], and Dimitar Molerov[b]

[a]Department of Law, Management and Economics, Johannes Gutenberg University Mainz, Mainz, Germany;
[b]Department of Education Studies, Humboldt University of Berlin, Berlin, Germany

ABSTRACT

In this article, the authors present the adaptation and validation processes conducted to render the American *Test of Financial Literacy* (TFL) suitable for use in Germany (TFL-G). First, they outline the translation procedure followed and the various cultural adjustments made in line with international standards. Next, they present results from the validation of the TFL-G's content and relations between test scores and external variables, including test takers' prior economic education and interest in economic topics. Preliminary analyses of data gathered from expert interviews and cognitive labs, and the results of the first administration to first-year higher education students ($N = 1{,}108$) indicate that the TFL-G is a valid instrument to assess young adults' understanding of personal finance in Germany. Perspectives for future research are discussed.

KEYWORDS

Adaptation of test instruments; financial literacy; personal finance; *Test of Financial Literacy*

JEL CODES

D14

In recent years, young adults in Germany have had to become increasingly proactive in securing their personal finances due to, for example, government cutbacks in social benefits (Erner, Goedde-Menke, and Oberste 2016). Very few validated test instruments are available to assess knowledge and understanding of personal finance in Germany (Aprea and Wuttke 2016). Further, education institutions in Germany did not participate in the financial literacy section of the Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment study (PISA) (OECD 2014, 57). Consequently, little is known about young adults' understanding of personal finance in Germany (Erner et al. 2016, 96). The lack of valid test instruments to assess such knowledge and understanding in Germany also impedes international comparison.

In this article, we present an adapted and validated version of the American *Test of Financial Literacy* (TFL) (Walstad and Rebeck 2016) for use in Germany (TFL-G). The original TFL was designed to assess students' understanding of personal finance in the United States (Walstad and Rebeck 2017). Studies indicate that personal finance is a construct shaped by the culture of the respective country (see Marchetti et al. 2016, 78; Mason and Wilson 2000, 25). Therefore, mere translation was insufficient to produce a valid test instrument for use in Germany; the items on the TFL must be adapted and validated to ensure that the instrument validly measures students' knowledge and understanding of personal finance in the German context.

Validation of the TFL-G was based on the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME 2014), according to which gathering evidence of the validity of the following is essential: (1) test content, (2) response processes of the test takers, (3) internal structure of the test, (4) relations between test scores and external variables, and (5) consequences of testing. In this article, we focus on test content, response processes, and relations between test scores and external variables. First, we specify the construct of personal finance knowledge and understanding adhered to in the TFL. Second, we give an

CONTACT Manuel Förster manuel.foerster@uni-mainz.de Johannes Gutenberg University Mainz, Department of Law, Management and Economics, Chair of Business Education, Jakob Welder-Weg 9, D-55099 Mainz, Germany.

overview of the translation and adaptation processes, and describe the necessary linguistic and cultural adjustments of the TFL to make the resultant TFL-G suitable for use in Germany. Third, we provide evidence of the validity of the test content, response processes, and relations to personal variables based on data gathered from expert interviews and cognitive interviews, and the results of the test administered to a sample of 1,108 young adults. With regard to relations to personal variables, we compared the grade point average (GPA), relevant prior education, and interests of the test takers to determine whether they matched theoretical assumptions regarding test performance (AERA, APA, and NCME 2014). Lastly, we summarize the results and present avenues for further research.

Construct of knowledge and understanding of personal finance

The TFL is an instrument designed to assess knowledge and understanding of personal finance (Walstad and Rebeck 2017) that young adults need to make responsible and appropriate personal financial decisions in life. This construct generally was adhered to when adapting the TFL. For example, the test items feature realistic scenarios embedded in the social, economic, and political context of the respective nation. For the assessment, we deliberately narrowed down the definition of the construct of financial literacy as it appeared in previous research (see Aprea and Wuttke 2016, 402; Atkinson and Messy 2012, 14). Focus was on the cognitive component of financial literacy, that is, on financial knowledge and understanding. Affective components such as motivation, attitudes, or other cognitive aspects such as metacognition also play a role in making sound financial decisions (Weinert 2001), but are not measured on the TFL. Rather, the aim of the test is to assess knowledge and understanding of financial products and underlying principles because they form the basis of many financial decisions. Accordingly, the test does not measure complex problem solving in the domain of personal finance. Although a situational context is provided in the items on the TFL, the context is not meant to simulate the highly complex environment of young adults and trigger complex cognitive processes. Instead, the items on the test represent simple situations that generally cover relevant domains of financial decision making and have a connection to the everyday environment of young adults. The scenarios are phrased precisely and concisely to cover knowledge and understanding in all content areas adequately and guarantee a satisfactory reliability within a manageable test-taking time of 45 minutes (Walstad and Rebeck 2016).

The TFL covers the six content areas *earning income, buying goods and services, saving, using credit, financial investing*, and *protecting and insuring* (CEE 2013; Walstad and Rebeck 2016). First, we must examine the extent to which the content areas and the corresponding items were adaptable for Germany because realities of financial systems and practices could differ significantly between the United States and Germany.

Translation and adaptation of test content

A preliminary review of the items and data gathered later from expert interviews confirmed that although specific finance-related practices differ between the United States and Germany, most principles of financial reasoning described in the CEE's *National Standards of Financial Literacy* (2013) are similar across the two countries, and the TFL generally is adaptable. The items on the TFL involved with *using credit* and *protecting and insuring* required particular adaptations to be suitable for use in Germany. For example, in the United States, credit cards are a standard method of payment, whereas young people in the target population in Germany pay more often by debit card (Klapper, Lusardi, and Van Oudheusden 2015, 21). Although transactional procedures and related reasoning differ between the two countries, problems of overspending or overdrawing a bank account and related charges and interest are similar in essence. In the area of *protecting and insuring*, general principles of reasoning are comparable, but there are larger cross-cultural differences in the availability, handling of, and attitudes toward insurance and social security plans.

To ensure a high quality translation, the Translation, Review, Adjudication, Pretesting, and Documentation (TRAPD) approach was taken (Harkness 2003), which is standard in the adaptation of international assessments and surveys and in line with the Test Adaption Guidelines (TAGs) (Hambleton 2001). The TRAPD approach stresses team discussion among experts in content, methodology, and translation

(Harkness 2003). The translation team consisted of three test developers and one professional translator, all of whom were German native speakers with more than three years of experience with test adaptations; for example, one of the test developers had more than 30 years of experience. In addition, two graduates of business education with majors in English assisted in the pre-translation preparations (for information on translation teams, see e.g., Arffman 2013, 4; Mohler et al. 2016). The team members also had experience translating and adapting economics tests (for information on adapting the *Test of Economic Literacy*, see Happ et al. 2016; Förster, Zlatkin-Troitschanskaia, and Happ 2015; for adapting the *Test of Understanding College Economics*, see Förster et al. 2015).

The steps we took in the TRAPD process follow. Supervised by the other team members, the two graduates prepared two separate translations of the TFL and annotated all linguistic, cultural, and/or content-related problems they encountered. The initial translations were reviewed individually by the other team members and reconciled in the team in two day-long sessions. Each of the 50 items[1] was discussed openly regarding necessary adaptations and alternatives until consensus was reached, which took twice as much effort as anticipated but enhanced the quality of the items on the German version of the test. Between sessions, issues related to content were resolved, if need be, by consulting the test developers and professionals in the domain to ensure factual correctness, realism of the tasks, and relevance of content. The updated version was proofread and then pretested, first during cognitive interviews with beginning students in higher education and later in the field with a larger sample. All necessary adaptations to content, and modifications that preserved the meaning but altered sentence structures, were documented for reference for later analyses.

Following the TAGs and the rationale on translation laid out in the Cross-Cultural Survey Guidelines (Mohler et al. 2016), the overall translation strategy was to use the ask-the-same-question approach as a basis, which aims to preserve psychometric properties by altering as little as possible, complemented with the ask-a-different-question approach (Mohler et al. 2016). Accordingly, translations closely followed the original meaning and structure of items, sentences, phrases, and terminology, while adaptations were made only if this version was deemed inadequate for the cultural context of Germany. The formal properties of the TFL items, including the short-item stem and four response options, and further elements such as tables were deemed appropriate for assessing basic financial knowledge in Germany. Information within the items was presented in the same sequence. Response options on the TFL-G were presented as closely as possible to the original response options on the TFL in terms of wording, order, and structure because these are known to influence response behavior. If this was not possible, the wording of two to four response options was kept as closely as possible to that in the original response options (Harkness 2003). With regard to style, we used short, simple sentence structures and general to nearly conversational language, but we made the sentences slightly more complex to account for the norms of written language in German. Domain-specific terminology referring to financial products with which students must be familiar did not correspond exactly across languages and countries, and therefore must be modified (e.g., TFL-G item 17 in figure 1). In particular, common abbreviations used in personal finance differed between the United States and Germany, and could not be used interchangeably between items. The entire test contained approximately the same number of abbreviations in order to maintain the same overall level of difficulty (e.g., TFL-G item 22 in figure 1).

Most items on the TFL include narrative devices to increase test takers' engagement. They are presented as brief scenarios of high school students' everyday life and contain focalizing characters with first names that are very common in the United States, including gender-neutral nicknames and first names that are common in minority groups, which should increase the chances that test takers will identify with the characters or will have peers with the same names. In the translation, the same goal was pursued by using frequency lists to identify some of the most common German, Turkish, Polish, and Italian first names among people aged 0–30 years in Germany. For instance, "Juan" on the TFL became "Ali" on the TFL-G, which is equally likely to represent a student with a migration background from a range of countries. Moreover, the names for characters on the TFL-G were chosen in such a way as to avoid cultural or gender stereotyping and suggest diversity; for example, a female electrical engineering student Katharina and a female presumably Middle-Eastern bookkeeper Yasmin were included. The following examples illustrate adaptations to accommodate cultural differences and modifications following linguistic conventions (for more details on relevant test or survey language, see e.g., Harkness 2003) in the TFL-G.

13. **Most states** have **"lemon laws."** These laws protect buyers from A. false advertising. B. credit card fraud. C. defective products.* D. **contaminated produce**.	13. **In Deutschland** schützt das **Gewährleistungsrecht** den Käufer vor A. falscher Werbung. B. Kreditkartenbetrug. C. mangelhaften Produkten.* D. **umweltschädlicher Produktion.**
Modifications to item 13 were necessary due to its legal content. First, legislation regarding the issue of defective products is at the national level in Germany not the state level. Second, in item 13 on the TFL the term "lemon laws" is metaphorical and does not offer any information on the laws. Students must know the term to respond correctly. Although the term "warranty right" [Gewährleistungsrecht] in item 13 on the TFL-G is more explicit, it also does not offer information on the law. The allusion in response option D to the literal meaning of lemons and potential humor could not be retained in the German version of the test. Instead, response option D on the TFL-G reads "environmentally harmful production" and has similar nomological relations within the item.	
17. What is the primary purpose of **an IRA**? A. allows workers to pay taxes with each paycheck B. provides incentives **for people** to save for retirement* C. eliminates the need for savers to collect Social Security D. creates savings **accounts** for current healthcare expenses	17. Was ist das Hauptziel einer **Riesterrente?** A. Sie erlaubt Arbeiternehmern, mit jeder Gehaltsabrechnung Steuern zu zahlen. B. Sie bietet Anreize, für die Rente zu sparen.* C. Sie macht die Einzahlung in die gesetzliche **Renten**versicherung überflüssig. D. Sie schafft **Rücklagen** für laufende Gesundheitsausgaben.
To respond correctly to item 17 on the TFL students must know the meaning of the abbreviation IRA (i.e., individual retirement accounts). There is no equivalent concept in Germany or abbreviation; therefore, another widely known and advertised retirement plan in Germany, the Riesterrente, was chosen, which mostly corresponds with the same response options. Wording of the response options had to be modified: in response B "for people" was omitted, in response C a lexical reference to pensions was made, and in response D "creates savings accounts" was changed to "builds up reserves" for plausibility.	
22. What does a **credit bureau** do? A. makes decisions about credit applications **B. matches banks to applicants who qualify for a loan** C. explains to consumers why they have been denied credit D. provides creditors with reports of consumers' credit-paying histories*	22. Welche Aufgabe hat die **Schufa**? A. Sie trifft Entscheidungen über Kreditanträge. B. Sie bringt **interessierte** Banken und Antragsteller zusammen. C. Sie erklärt Verbrauchern, warum ihr Kreditantrag abgelehnt wurde. D. Sie informiert Kreditgeber über die **Rückzahlung bisheriger Kredite des Kreditnehmers.***
Item 22 is an example of where a common abbreviation was used on the TFL-G but not in the TFL. The "Schufa" (i.e., Schutzgemeinschaft für allgemeine Kreditsicherung [General Credit Protection Agency]) in Germany has similar functions as credit bureaus in the United States and operates nationally. Because the abbreviation is very well known and does not denote credit, response option B had to be shortened for plausibility. In turn, the semantically condensed phrase "credit-paying histories" in response option D on the TFL had to be spelled out on the TFL-G in the phrase "informs creditors about the repayment of previous credits of the credit borrower."	

Figure 1. Sample items from the TFL and TFL-G, adaptations and modifications **in bold**.

The above examples illustrate the scope of the necessary adaptations made. Although differences between the two test versions occurred in 25 of the 50 items, they were considered to be construct-irrelevant, and the items performed well in the pretest. Results of the pretest in October 2015 revealed one problematic linguistic detail in item 33, which had passed the reconciliation step unnoticed but was easily corrected and used in its altered version in April 2016.

Evidence from expert interviews—Validation of test content

To determine whether valid conclusions about test takers' understanding of personal finance in Germany can be drawn from their scores on the TFL-G, the relationship between the test content and the construct was investigated (AERA, APA, and NCME 2014, 14). To ascertain whether the applied concepts from the *National Standards for Financial Literacy* (CEE 2013) upon which the TFL is based also are valid in Germany regarding content, we conducted interviews with 10 experts in various professions: three representatives from the private banking sector and the German Central Bank (Deutsche Bundesbank), two student liaison officers for social and financial issues, one consumer advice center worker, one representative from an insurance company, one debt counselor, and two teachers. During the interviews,

the experts were asked to identify areas of knowledge and understanding they considered crucial for students to be able to make sound personal financial decisions. In the next step, they also evaluated the relevance of the six underlying content areas, as defined by the *Standards* (CEE 2013), for the construct of knowledge and understanding of personal finance for Germany. Next, the experts completed an online questionnaire on which they evaluated the quality and relevance of the items on the TFL and the appropriateness of the scenarios and content for young adults in Germany. The experts were encouraged to make suggestions for improvement. Through interviews with the experts and analysis of their responses to items on an online questionnaire, the relevance of the items on the TFL-G for young adults in Germany was confirmed. The experts rated the relevance of the content areas covered on the test as being good. Although the experts evaluated the content areas related to their own professions as being more relevant (e.g., the representatives from the banking sector rated the content areas *saving* and *using credit* as being particularly important, whereas the consumer advice center worker stressed the importance of the content area *buying goods and services* and hereby especially the topic *consumers' rights*), all in all there was consensus among the experts regarding the content-related relevance of the six content areas and the corresponding questions.

Evidence from cognitive interviews with students—Validation of response processes

The adaptation process and the data gathered from experts led to a preliminary version of the TFL-G with validated content. This version was used to examine whether the item scores and total scores on the TFL-G were achieved by construct-relevant mental response processes. If this was the case, the test scores could be considered indicators of knowledge and understanding of personal finance. If the students succeeded in solving the tasks by merely guessing (see Frary, Cross, and Lowry 1977; Hamdan 1979) or by using their testwiseness (Hess and Neville 1977), interpretation of the findings would be significantly limited. Test scores have great explanatory power if the students solve the tasks using construct-relevant strategies. To examine whether the test takers used the (financial) knowledge and performed mental operations intended by the test developers, the think-aloud method was used both concurrently and retrospectively.[2] These two ways of using the think-aloud method are effective for analyzing the clarity (and/or quality) of items and the (mis)alignment of students' responses and their mental processes (Leighton et al. 2011). Cognitive interviews with seven first-year students were conducted to verify whether the cognitive processes involved in responding to the items were construct-relevant rather than construct-irrelevant test-taking strategies or guessing (AERA, APA, and NCME 2014, 15). Conducting cognitive interviews made it possible to identify items on the test that were not formulated ideally from the perspective of the target group. Furthermore, data gathered from the think-aloud interviews indicated that students used construct-relevant mental processes to respond to the tasks and that the occurrence of construct-relevant processes was significantly positively correlated to correct item responses. Overall, students' responses were based on construct-relevant mental processes; this evidence strengthened the interpretation of the students' scores on the TFL-G as an indicator of their level of financial knowledge and understanding (AERA, APA, and NCME 2014).

During the retrospective phase of the cognitive interviews, various sources of respondents' knowledge of personal finance were revealed. Curricular analyses show that personal finance is unsystematically taught at school in Germany (Frühauf and Retzmann 2016). This fact also is reflected in the sources of knowledge the respondents named during the cognitive labs, the most frequent being personal experience. Other, much less frequently mentioned sources were vocational training in Germany,[3] school, family and friends, and media (newspapers, television, and radio).

Evidence from the pretest—Validation of relations between test scores and external variables

The TFL-G was administered to students at a university in Germany in the winter term 2015–16 and summer term 2016.[4] After omitting 99 students who were not in their first semester, 1,176 beginning students remained in the sample. These students were enrolled in economics, business education, law,

educational sciences, sociology, psychology, and translation programs but had not yet attended a university course. First-year students were chosen for the sample because of the relevance of financial literacy for them: they are entering a new phase of life, and many set up their first own households and must make many financial decisions on their own (e.g., whether or not to take out a student loan). Also, they seem to be suitable because the TFL was developed for students at the end of high school in the United States (Walstad and Rebeck 2016). An additional 61 students were eliminated from the sample because they did not respond to the question regarding age, or they were older than 25, and we assumed these respondents had had the opportunity to gain much more experience making financial decisions. A further seven respondents were omitted because they had more than 30 missing values on the items, suggesting they had completed the questionnaire carelessly.

The age of the remaining 1,108 respondents was between 17 and 25 years, and the average age was 20.02 years. Of the respondents, 59.5 percent were female. In the following, the one-dimensionality of the TFL-G is tested using confirmatory factor analysis (Brown 2006); then, the item statistics for the entire sample are presented. Next, validity evidence from group comparisons is presented to analyze the relations between test scores and external variables that should have an impact on economic knowledge and understanding, as suggested in AERA, APA, and NCME (2014). We assumed that students who had completed vocational training before the beginning of their studies or who had attended a major economics course at school would achieve better test results than their peers who had not obtained vocational training or attended a major economics course. Curricular analyses revealed that knowledge of personal finance can be gained during vocational training (Solga et al. 2014) and in a major economics course at school in Germany (Frühauf and Retzmann 2016). Moreover, the connection between test scores and other important variables was investigated, including interest in financial and economic topics and students' high school GPA. Here, it can be expected that beginning students with greater interest in financial topics (Aprea and Wuttke 2016; Lalonde and Schmidt 2009) and a better school GPA (for information on the influence of school grades on financial knowledge in Germany, see Erner et al. 2016) will achieve better test results. The distribution of the relevant variables in the sample is summarized in table 1.[5]

The original American version of the TFL is considered to be one-dimensional (Walstad and Rebeck 2016), which was to be confirmed for the German version as well. In the confirmatory factor analysis, the data fit for a one-dimensional model with dichotomous items ($Chi^2 = 1{,}475.78$, $df = 902$, RMSEA $= 0.02$, 90% confidence interval of the RMSEA $= [0.22; 0.26]$) was judged to be good. Hence, for the following modeling, a one-dimensional model that does not differentiate between single content areas was assumed to be sufficient.

Table 2 shows the corrected discrimination index (corrected item-total correlation) and the share of correct responses to the 45 and 44 items.[6] In total, the reliability was found to be very good with

Table 1. Descriptive statistics of the sample.

Sample size	Absolute	In percentages
Number of students	1,108	100.00
Major economics course		
No major economics course	902	81.40
Major economics course	199	18.00
Missing values	7	0.60
Commercial vocational training		
No commercial vocational training	980	88.40
Commercial vocational training attended	126	11.40
Missing values	2	0.20
Means		
High school GPA	2.30 (0.57)	
Missing values		18
Interest in financial and economic topics	1.96 (1.04)	
Missing values		6

Note: Standard deviations are in parentheses.

Table 2. Share of correct responses and IRT difficulty and discrimination.

	Analyses without item 33		IRT analyses without item 33		Analyses with item 33
Item no. in both TFL and TFL-G	Share of correct responses	Corrected item-total correlation	IRT difficulty	IRT discrimination index	Corrected item-total correlation
TFL-G 1	.78	.34	−1.44	0.63	.27
TFL-G 2	.59	.29	−0.55	0.45	.21
TFL-G 3	.62	.23	−0.95	0.33	.13
TFL-G 4	.74	.32	−1.37	0.54	.30
TFL-G 5	.71	.37	−1.02	0.64	.30
TFL-G 6	.67	.36	−0.83	0.62	.41
TFL-G 7	.54	.43	−0.17	0.72	.52
TFL-G 8	.71	.27	−1.36	0.44	.33
TFL-G 9	.72	.10	−3.51	0.17	.11
TFL-G 10	.60	.30	−0.61	0.45	.25
TFL-G 11	.93	.32	−2.24	0.85	.11
TFL-G 12	.47	.20	0.23	0.30	.21
TFL-G 13	.59	.24	−0.72	0.34	.25
TFL-G 14	.64	.22	−1.15	0.33	.11
TFL-G 15	.49	.16	0.15	0.23	.13
TFL-G 16	.54	.27	−0.25	0.40	.36
TFL-G 17	.58	.34	−0.41	0.53	.41
TFL-G 18	.32	.17	1.88	0.25	.11
TFL-G 19	.69	.37	−0.97	0.61	.56
TFL-G 20	.64	.35	−0.71	0.57	.37
TFL-G 21	.45	.21	0.45	0.32	.22
TFL-G 22	.61	.36	−0.55	0.58	.40
TFL-G 23	.29	.11	3.32	0.17	.12
TFL-G 24	.62	.30	−0.76	0.45	.36
TFL-G 25	.82	.34	−1.68	0.65	.43
TFL-G 26	.81	.21	−2.59	0.35	.20
TFL-G 27	.59	.45	−0.36	0.79	.62
TFL-G 28	.68	.42	−0.79	0.76	.52
TFL-G 29	.31	.30	1.15	0.47	.41
TFL-G 30	.58	.34	−0.42	0.55	.29
TFL-G 31	.25	.15	2.99	0.23	.23
TFL-G 32	.67	.30	−1.03	0.47	.32
TFL-G 33	.65	—	—	—	.21
TFL-G 34	.42	.08	1.85	0.11	.30
TFL-G 35	.49	.34	0.06	0.54	.38
TFL-G 36	.55	.42	−0.22	0.73	.36
TFL-G 37	.32	.13	2.42	0.19	.09
TFL-G 38	.68	.33	−0.97	0.54	.37
TFL-G 39	.44	.13	0.85	0.17	.14
TFL-G 40	.87	.28	−2.23	0.58	.55
TFL-G 41	.64	.36	−0.70	0.59	.55
TFL-G 42	.85	.29	−2.14	0.55	.47
TFL-G 43	.43	.20	0.61	0.31	.18
TFL-G 44	.57	.19	−0.68	0.29	.25
TFL-G 45	.59	.27	−0.58	0.41	.39
N	*1,108*	*1,108*	*1,108*	*1,108*	*126*
Cronbach's alpha		0.82			0.85

Cronbach's alpha = 0.85 for the 45 items ($N = 126$) and 0.82 for the version without item 33 ($N = 1{,}108$).

Because the sample that completed the TFL-G with the new version of item 33 contained 126 respondents only, the following results are be based on the version without item 33. The revised version of item 33 showed good discrimination and difficulty values[7] (see right column in table 2). All 44 items showed a positive discrimination, with 35 items having a corrected discrimination index of above 0.2. Only items 9 and 34 had a discrimination index of slightly lower than 0.1. Regarding the level of difficulty of the items, 39 questions were in the desired range of a share of correct responses between 0.25 and 0.75. Only six tasks were completed correctly by more than 75 percent of the respondents and therefore generally can be considered easy for first-year students. Especially in the area of empirical educational research,

parameters from Item Response Theory (IRT) are used frequently as item descriptives, so these are added for additional orientation.[8] Low values, in this case, negative values, for IRT difficulty indicate a very easy question. The higher the value for IRT difficulty is, the more difficult the item is. The classic shares of correct responses and the IRT difficulty correlated with (−)0.93, and the correlation of item discrimination of both methods was at 0.95; therefore, the results of both analysis procedures were nearly identical. For further analysis, we used the share of correct responses as an indicator of difficulty because it is easier to interpret than IRT parameters.

Next, the test scores were analyzed in relation to the respondents' opportunities to learn about personal finance. Very few respondents had attended a major economics course at high school. The majority of the respondents had not participated in a course specifically on economics at high school or joined in a minor economics course. It can be assumed that the beginning students who had attended a major economics course at school achieved higher scores than their fellow students. The effect of having completed a major course in economics should only be moderate because curricular analyses show that personal finance and, therefore, principles of individual financial decision-making, are underrepresented in the curricula of economics courses and other courses in Germany. If these topics are touched upon at all in school, focus is on their general relationship to economic concepts and principles, which are related to economic literacy. The cognitive interviews with students supported this fact.

Completion of commercial vocational training constitutes further relevant prior education. When participating in a dual vocational training program, trainees complete an apprenticeship at a company and attend vocational school, where they are taught about various areas of business and economics. Further, they earn money (likely for the first time) while doing an apprenticeship, meaning they have the opportunity to learn first-hand how to handle their personal finances. Hence, it was assumed that the students who had completed vocational training before beginning their studies would achieve better test results than their fellow students who had not obtained vocational training.

As shown in table 3, on most items, students who had completed a major economics course at school achieved better test results than their fellow students who had not attended such a course. On the entire test, a significant difference in the mean of 1.85 correct responses between the two groups was found in favor of the students who had attended a major economics course. This difference in the mean values was not very large (Cohen's $d = 0.27$), as expected, given the generally low share of personal finance topics presented in school curricula in Germany (Frühauf and Retzmann 2016; Erner et al. 2016).

The difference between the subpopulation with and the subpopulation without vocational training was much greater, with Cohen's d of 0.85 indicating a large effect. In line with our expectations, students who had completed vocational training achieved better test results on all items, with the exception of item 43 on the TFL-G, and on average had five and a half more correct responses than their peers. These results confirm the expectation that more opportunities to learn how to handle one's own finances can be found in vocational training programs than in ordinary high school programs. This evidence indicated that the TFL is indeed a suitable instrument to assess practical knowledge needed to handle one's own finances. Furthermore, the results strengthen the findings from the qualitative cognitive interviews, according to which financial knowledge and understanding of financial situations are increasingly based on the respondents' own experiences (Cameron et al. 2014) and much less on formal instruction at school in Germany.

Next, we analyzed the relations between students' test scores and their high school GPA and interest in financial and economic topics. First, the students were divided into four quartiles according to their GPA. Subsequently, the average TFL-G score was calculated for each of the four groups (see table 4).

As expected, the analyses showed a gap between the different groups: the 25 percent of students with the best GPA also had the highest scores on the test (27.57 points). The second quartile followed with 26.75 correct responses to the items. Between the third and fourth quartile, there was almost no difference with 24.97 and 24.91 points, respectively. In the variance analysis, a multi-group comparison with corrected rejection level and confidence intervals was conducted. The upper two quartiles performed significantly better than the two lower quartiles with worse GPAs.

The relation between the TFL-G score and interest in financial and economic topics showed a similar picture, in line with expectations (see table 5). Results of studies indicate that interest in financial

Table 3. Share of correct responses in relation to prior education.

	Completed major economics course	Without major economics course		Completed vocational training	Without vocational training	
	Share of correct responses	Share of correct responses	Diff.	Share of correct responses	Share of correct responses	Diff.
TFL-G 1	.82	.77	0.05	.84	.77	0.07
TFL-G 2	.61	.59	0.02	.65	.58	0.07
TFL-G 3	.69	.60	0.09	.77	.60	0.17
TFL-G 4	.81	.73	0.08	.89	.72	0.17
TFL-G 5	.72	.71	0.01	.78	.70	0.08
TFL-G 6	.72	.66	0.06	.80	.65	0.15
TFL-G 7	.59	.52	0.07	.81	.50	0.31
TFL-G 8	.68	.71	− 0.03	.80	.69	0.11
TFL-G 9	.69	.72	− 0.03	.76	.71	0.05
TFL-G 10	.64	.59	0.05	.75	.58	0.17
TFL-G 11	.91	.93	− 0.02	.94	.93	0.01
TFL-G 12	.47	.48	− 0.01	.56	.46	0.10
TFL-G 13	.74	.56	0.18	.79	.57	0.22
TFL-G 14	.65	.64	0.01	.71	.63	0.08
TFL-G 15	.49	.49	0	.52	.48	0.04
TFL-G 16	.57	.53	0.04	.66	.52	0.14
TFL-G 17	.66	.56	0.1	.83	.54	0.29
TFL-G 18	.39	.31	0.08	.56	.29	0.27
TFL-G 19	.73	.68	0.05	.83	.68	0.15
TFL-G 20	.73	.62	0.11	.79	.62	0.17
TFL-G 21	.45	.45	0	.52	.44	0.08
TFL-G 22	.69	.59	0.10	.82	.58	0.24
TFL-G 23	.34	.28	0.06	.37	.28	0.09
TFL-G 24	.68	.61	0.07	.81	.60	0.21
TFL-G 25	.84	.81	0.03	.95	.80	0.15
TFL-G 26	.79	.81	− 0.02	.86	.80	0.06
TFL-G 27	.64	.58	0.06	.77	.57	0.20
TFL-G 28	.69	.68	0.01	.75	.67	0.08
TFL-G 29	.34	.30	0.04	.50	.29	0.21
TFL-G 30	.58	.58	0	.73	.56	0.17
TFL-G 31	.30	.24	0.06	.34	.24	0.10
TFL-G 32	.72	.66	0.06	.79	.65	0.14
TFL-G 33	.46	.41	0.05	.61	.40	0.21
TFL-G 34	.53	.48	0.05	.60	.48	0.12
TFL-G 35	.63	.53	0.10	.71	.53	0.18
TFL-G 36	.35	.32	0.03	.35	.32	0.03
TFL-G 37	.68	.67	0.01	.76	.66	0.10
TFL-G 38	.50	.43	0.07	.46	.44	0.02
TFL-G 39	.88	.87	0.01	.94	.86	0.08
TFL-G 40	.65	.63	0.02	.72	.63	0.09
TFL-G 41	.86	.84	0.02	.94	.84	0.10
TFL-G 42	.46	.42	0.04	.52	.42	0.10
TFL-G 43	.58	.57	0.01	.56	.58	− 0.02
TFL-G 44	.61	.58	0.03	.71	.57	0.14
N	*199*	*902*		*126*	*980*	
Cronbach's alpha	0.82	0.81		0.83	0.80	
Mean (SD)	27.57 (6.92)	25.72 (6.84)	1.85 p = .00	31.10 (6.59)	25.44 (6.65)	5.66 p = .00
Cohen's d			0.27			0.85

p = *p*-value of the *t*-statistic.

and economic topics influences knowledge of personal finance (Aprea and Wuttke 2016; Lalonde and Schmidt 2009).

Students with greater interest in financial and economic topics scored considerably better on the test. First-year students who indicated that they had much or very much interest in such topics on average scored 28.35 and 29.94 points, respectively, thus 2.5 to almost 7 points more than their peers. In the analysis of variance, the results of the multi-group comparison with corrected rejection levels and confidence intervals indicated that students with some interest scored significantly better than students who

Table 4. Relationship between TFL-G score and high school GPA.

	H	Mean	Standard deviation	Standard error	95% confidence interval for mean value: Lower limit	95% confidence interval for mean value: Upper limit	Minimum	Maximum
Top 25%	305	27.57	6.88	.39	26.80	28.35	7.00	41.00
Upper 25%–50%	280	26.75	6.41	.38	26.00	27.50	10.00	39.00
Lower 25%–50%	255	24.97	6.92	.43	24.12	25.83	7.00	41.00
Lowest 25%	250	24.92	6.97	.44	24.05	25.78	9.00	41.00
Total	1,090	26.14	6.88	.21	25.74	26.55	7.00	41.00

Table 5. Relationship between mean TFL-G score and interest in financial and economic topics.

	H	Mean	Standard deviation	Standard error	95% confidence interval for mean: Lower limit	95% confidence interval for mean: Upper limit	Minimum	Maximum
0 Not at all	106	23.08	6.28	.61	21.87	24.28	9.00	36.00
1 Little	236	24.33	6.23	.41	23.53	25.13	8.00	37.00
2 Somewhat	429	25.82	6.54	.32	25.20	26.44	7.00	41.00
3 Much	263	28.35	7.10	.44	27.49	29.22	7.00	40.00
4 Very much	68	29.94	6.97	.85	28.25	31.63	11.00	41.00
Total	1,102	26.10	6.88	.21	25.69	26.50	7.00	41.00

had little or no interest at all, and significantly worse than students who had much or very much interest in financial and economic topics.

Discussion

In this article, we present the adaptation and validation processes of the TFL for use in Germany. The results from the expert interviews confirmed that the content standards of financial literacy laid out by the CEE (2013) generally are applicable to Germany. Some items on the TFL must be adapted to the German context. Legal, linguistic, and cultural differences were identified and the items adjusted accordingly with the aim of preserving their functions and measurement properties. The TFL-G was judged by experts in each of the assessed content areas to be relevant and realistic for use in Germany. Further analyses of students' response processes showed that students must use construct-relevant thought and response processes to find the correct responses, while guessing and construct-irrelevant response processes led to incorrect responses. In line with theory, the expected differences between sub-groups were confirmed in the empirical study. In summary, the following assumptions were strengthened with empirical evidence:

- The TFL-G assesses components of knowledge and understanding of personal finance that are judged to be relevant in content and realistic for Germany.
- The TFL-G test score offers an indication of the level of students' knowledge and understanding of personal finance.
- The test results confirmed expected differences in knowledge levels depended on relevant personal characteristics of the students, such as prior education and interest.

In this study, we focused on the suitability of the TFL-G for first-year students in higher education in Germany. Whether it also is suitable for reliably and validly assessing the financial knowledge and understanding of high school students remains to be examined in further studies. We expected the sample of first-year higher education students to represent a very positive selection of high school graduates and that the average level of knowledge of high school students should be lower. Analyses of the items showed a few items with a share of correct responses higher than 75 percent, which on average can be expected to be more difficult for high school students. Therefore, the items might be even more appropriate for high school students, who were the original target group of the TFL in the United States, and

the fit of the data to the statistical models might be even better than for the beginning higher education students in this study.

In future research, the TFL-G can be administered to young adults in Germany to determine how they acquire financial knowledge and understanding. The results from the curricular analyses, cognitive interviews with students, and the quantitative study indicate that schooling in Germany, at least for this sample, contributed little to explain students' knowledge and understanding of personal finance. Therefore, important influence factors likely are connected to learning opportunities outside of school, including personal experiences, parental socialization, personal interests, and media reception, which remain to be examined.

Moreover, a follow-up study can take an international perspective and examine the feasibility of comparing the American and German test versions, which could enable cross-national comparisons of test results at a later stage. The necessary adaptations for Germany might present a threat to comparability, which must be determined. Nevertheless, the adaptation approach taken in the development of the TFL-G was geared toward both functions of allowing the creation of a valid target-language assessment instrument and introducing as few changes as possible to the original TFL in order to achieve comparability and preserve psychometric properties of the original items, such as item difficulty. The combined adaptation and validation approach was suitable for creating a high quality German version of the TFL, and while it might appear elaborate, each step contributed to improving or assuring the quality of the test instrument. Thanks to its generality, the adaptation approach should be suitable for creating versions of the test instrument for use in other countries as well. The kind and scope of necessary adaptations will differ depending on the respective legal, linguistic, and cultural context, but the approach is designed to indicate general suitability of the test early on, that is, in the preliminary review of the test and the assessed underlying economic and financial principles. In this regard, the CEE's *National Standards of Financial Literacy* (2013) provided a useful orientation for preliminary review and expert interviews in the adaptation of the TFL-G. Cross-national comparability of the assessments and their psychometric properties remain to be examined with comparable samples, for instance, using confirmatory factor analysis for multi-group comparisons (for cross-national comparability analyses of an economics test, see, e.g., Förster et al. 2015).

Notes

1. In its first version from 2015, the TFL consisted of 50 items (Walstad and Rebeck 2016). After a pretest, the number of items was reduced to 45 in the U.S. version. In Germany, all 50 items were adapted, and the same 45 as in the United States are discussed in this study.
2. In think-aloud interviews, respondents are asked to verbalize their thoughts while working on tasks (in this case, responding to items on the TFL-G). During the item response process, respondents were only prompted to continue talking. Afterward, students were asked more detailed questions in cognitive labs (e.g., from which sources they had acquired their knowledge). These interviews were verbally recorded, transcribed, and qualitatively analyzed. The think-aloud method makes it possible to identify not only the clarity of items but also the cognitive processes taking place, which allows researchers to draw conclusions about the validity of the test instrument (for information on cognitive interviews for economics items, see Brückner and Pellegrino 2016).
3. Germany is well known for its dual vocational training system. One main characteristic of the dual system is the practical education in a company (share of approximately 4/5) combined with a school component (share of approximately 1/5). Usually, the vocational training program lasts two to three and a half years, and trainees are paid a small salary by the company where they do their apprenticeship. Such an experience should have an influence on their knowledge of personal finance because the young adults must handle the money they earn and plan expenses.
4. The survey conducted in summer term 2016 was necessary because one item required modification. The sample in the summer term consisted of 126 students for this article.
5. In the German grading system, 1 is the best and 4 the worst passing grade.
6. Analogous to the test developers' analyses (Walstad and Rebeck 2016), the following analyses relate to the 45 items on the TFL-G.
7. Because the revised item 33 was used in the pretest in summer term 2016 only, the difficulty of this item refers to 126 respondents only.
8. For a description of an IRT model for economic tests, see Walstad and Robson (1997). The calculations of the confirmatory factor model and item response modeling were made using the Mplus (version 7.3) software. See Muthén and

Muthén (1998–2015) for more details. The remaining modeling was conducted with the SPSS (version 22) software (IBM 2013).

References

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). 2014. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Aprea, C., and E. Wuttke. 2016. Financial literacy of adolescents and young adults: Setting the course for a competence-oriented assessment instrument. In *International handbook of financial literacy*, ed. C. Aprea, E. Wuttke, K. Breuer, N. Keng Koh, P. Davies, B. Greimel-Fuhrmann, and J. Lopus, 397–414. Singapore: Springer Science-Business Media.

Arffman, I. 2013. Problems and issues in translating international educational achievement tests. *Educational Measurement: Issues and Practice* 32 (2): 2–14.

Atkinson, A., and F. Messy. 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. doi:10.1787/5k9csfs90fr4-en.

Brown, T. A. 2006. *Confirmatory factor analysis for applied research (Methodology in the social sciences)*. New York: Guilford Press.

Brückner, S., and J. Pellegrino. 2016. Analyzing mental operations of higher education students using a sociocognitive approach. *Journal of Educational Measurement* 53 (3): 293–312.

Cameron, M., R. Calderwood, A. Cox, S. Lim, and M. Yamaoka. 2014. Factors associated with financial literacy among high school students in New Zealand. *International Review of Economics Education* 16 (Part A): 12–21.

Council for Economic Education (CEE). 2013. *National standards for financial literacy*. New York: CEE.

Erner, C., M. Goedde-Menke, and M. Oberste. 2016. Financial literacy of high school students: Evidence from Germany. *Journal of Economic Education* 47 (2): 95–105.

Förster, M., O. Zlatkin-Troitschanskaia, S. Brückner, R. Happ, R. K. Hambleton, W. B. Walstad, T. Asano, and M. Yamaoka. 2015. Validating test score interpretations by cross-national comparison: Comparing the results of students from Japan and Germany on an American test of economic knowledge in higher education. *Zeitschrift für Psychologie* 223 (1): 14–23.

Förster, M., O. Zlatkin-Troitschanskaia, and R. Happ. 2015. Adapting and validating the "*Test of economic literacy*" to assess the prior economic knowledge of first-year students in business and economic studies in Germany. Discussion Paper. Annual Meeting of the American Economic Association, January 3–5. Boston: AEA.

Frary, R. B., L. H. Cross, and S. R. Lowry. 1977. Random guessing, correction for guessing, and reliability of multiple-choice test scores. *Journal of Experimental Education* 46 (1): 11–15.

Frühauf, F., and T. Retzmann. 2016. Financial literacy in Germany. In *International handbook of financial literacy*, ed. C. Aprea, E. Wuttke, K. Breuer, N. K. Koh, P. Davies, B. Greimel-Fuhrmann, and J. S. Lopus, 263–76. Singapore: Springer Science-Business Media.

Hambleton, R. K. 2001. The next generation of the ITC test translation and adaption guidelines. *European Journal of Psychological Assessment* 17 (3): 164–72.

Hamdan, M. A. 1979. On the correction for guessing on a multiple-choice examination. *Journal of Experimental Education* 48 (1): 29–31.

Happ, R., F. Förster, O. Zlatkin-Troitschanskaia, and V. Carstensen. 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. doi:10.1177/2047173416646597.

Harkness, J. 2003. Questionnaire translation. In *Cross-cultural survey methods*, ed. J. Harkness, F. van de Vijver, and P. Mohler, 35–56. Hoboken, NJ: Wiley.

Hess, A. K., and D. Neville. 1977. Testwiseness: Some evidence for the effect of personality testing on subsequent test results. *Journal of Personality Assessment* 41 (2): 170–77.

IBM Corp. 2013. *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.

Klapper, L., A. Lusardi, and P. van Oudheusden. 2015. Financial literacy around the world: Insights from the Standard & Poor's ratings services global financial literacy survey. New York: S&P Global. http://media.mhfi.com/documents/2015-Finlit_paper_17_F3_SINGLES.pdf (accessed December 2, 2016).

Lalonde, K., and A. Schmidt. 2009. Credit cards and student interest: A financial literacy survey of college students. *Higher Education Journal* 10 (1): 1–14.

Leighton, J. P., C. Heffernan, M. K. Cor, R. J. Gokiert, and Y. Cui. 2011. An experimental test of student verbal reports and teacher evaluations as a source of validity evidence for test development. *Applied Measurement in Education* 24 (4): 324–48.

Marchetti, A., I. Castelli, D. Massaro, and A. Valle. 2016. Combining development and education: Why do decision-making and social norms matter for financial education? In *International handbook of financial literacy*, ed. C. Aprea, E. Wuttke, K. Breuer, N. K. Koh, P. Davies, B. Greimel-Fuhrmann, and J. S. Lopus, 69–81. Singapore: Springer Science-Business Media.

Mason, C. L., and R. M. S. Wilson. 2000. *Conceptualising financial literacy*. Loughborough: Business School, Loughborough University.

Mohler, P., B. Dorer, J. de Jong, and M. Hu. 2016. *Translation: Overview. Guidelines for best practice in cross-cultural surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.

Muthén, L. K., and B. O. Muthén. 1998–2015. *Mplus user's guide*. 7th ed. Los Angeles, CA: Muthén & Muthén.

Organisation for Economic Co-operation and Development (OECD). 2014. *PISA 2012 results: Students and money. Financial literacy skills for the 21st century*. (Vol. VI). PISA: OECD Publishing.

Solga, H., P. Protsch, C. Ebner, and C. Brzinsky-Fay. 2014. The German vocational education and training system: Its institutional configuration, strengths, and challenges. WZB Discussion Paper SP I 2014–502. Berlin: WZB.

Walstad, W. B., and K. Rebeck. 2016. *Test of financial literacy: Examiner's manual*. New York: Council for Economic Education.

Walstad, W. B., and K. Rebeck. 2017. The *Test of financial literacy*: development and measurement characteristics. *Journal of Economic Education* 48 (2): 113–122.

Walstad, W. B., and D. Robson. 1997. Differential item functioning and male-female differences on multiple-choice tests in economics. *Journal of Economic Education* 28 (2): 155–71.

Weinert, F. E. 2001. Concept of competence: A conceptual clarification. In *Defining and selecting key competencies*, ed. D. S. Rychen and L. H. Salganik, 45–65. Seattle: Hogrefe & Huber.