Journal of Family and Economic Issues
https://doi.org/10.1007/s10834-026-10097-5

ORIGINAL PAPER

An Overlooked Population: Why Linguistic Background Matters in Personal Finance Survey Design

Katrina Borowiec[1] · Miranda Reiter[2] · Ashley B. LeBaron-Black[3] · Kurt A. Schindler[4]

Received: 10 April 2024 / Accepted: 13 November 2024

Abstract

Linguistic diversity in the United States is growing. While language is often a focal area of consideration when designing and validating international surveys, the diversity of language within one individual country may be overlooked. Indeed, despite the diversity of languages used in the United States, most surveys are still administered in English. Depending on English language proficiency, question reading levels and question complexity may impact the responses provided. This may be especially true for personal finance surveys that incorporate complex terminology. Using a sample of 978 adults living in the United States, this exploratory quantitative study provides insights on the relationships among primary language used at home, the reading difficulty of survey items, and survey responses. The analyses include descriptive statistics on the reading difficulty of select personal finance scales; comparisons of item- and score-level response variance between individuals who primarily use English at home (N = 826) and those primarily using other languages (N = 152) both generally and in relation to reading difficulty; and comparisons of response styles or tendencies by primary language, reading difficulty, and their interaction. While the findings were mixed, the results suggest that reading difficulty appears to play some role in how participants respond to personal finance surveys and that this relationship differs by language group. We recommend that cognitive interviews with linguistically diverse participants are needed to understand what cognitive processes may lead to observed differences in item response by primary language and what changes to surveys are needed to better measure linguistically diverse participants' experiences.

Keywords Linguistic diversity · Reading difficulty · Response styles · Language in survey development · Language proficiency

✉ Katrina Borowiec
katrina.borowiec@bc.edu

Miranda Reiter
mreiter@ttu.edu

Ashley B. LeBaron-Black
ashley_lebaron-black@byu.edu

Kurt A. Schindler
Kurt.schindler@upr.edu

[1] Boston College Lynch School of Education and Human Development, 140 Commonwealth Avenue Chestnut Hill, Newton, MA 02467, USA

[2] School of Financial Planning, Texas Tech University, 1301 Akron Ave, Box 41210, Lubbock, TX 79409, USA

[3] School of Family Life, Brigham Young University, 2065 JFSB, Provo, UT 84602, USA

[4] College of Business Administration, University of Puerto Rico, Rio Piedras Campus, 601 Ave. Universidad, San Juan 00925, USA

Introduction

Responding to survey questions involves complex cognitive processes. Upon being asked a survey question, individuals must comprehend the question and instructions; retrieve the information needed to respond; use their judgment to evaluate their memory and draw inferences required to answer the question; and finally, provide a response to the question by mapping their judgment onto the appropriate response category (Tourangeau et al., 2000). Tourangeau and colleagues (2000) caution that respondents vary in their willingness and ability to extend the effort and time needed to accurately complete these cognitive tasks. Further, completing a survey in a non-native or secondary language can exacerbate these issues by increasing cognitive demand, since the language used to administer a survey can impact how participants perceive survey questions and the time they need to respond to each question (Peytcheva, 2018).

Published online: 19 June 2026

Indeed, researchers have identified differences in responses by language, including social desirability effects, differences in response consistency, and other behaviors. For example, Brown (2015) noted that primarily Spanish-speaking Hispanic respondents may be more susceptible to responding in socially desirable ways – i.e., providing a response that they believe would make them appear "better" or more favorable. Moreover, Reid (1990) found lower correlations among item responses among English language learners than native English speakers, suggesting less consistency in responses. In another study, Wenz et al. (2021) found that survey results from non-native English speakers resulted in lower-quality data. More specifically, these participants were more likely to change answers across waves, select the primary response choice, and had higher non-response rates when asked to self-complete the survey. In addition, they were more likely to engage in "straightlining" behavior – i.e., providing the same response option to all items within the same grid. One possible explanation for these behaviors is that respondents who take surveys in a foreign language may not completely understand the questions, and it may feel onerous for them to complete all the survey items (Jacobsen, 2022). This language mismatch may produce a new and often unacknowledged source of item nonresponse, potentially leading to biased results (Čehovin et al., 2023).

Other interesting intersections have been identified between language and survey response patterns. These response patterns, also known as "response styles" or "response sets," are patterns of behavior that influence how individuals respond to survey questions (Cronbach, 1946). Response styles include using the midpoint; acquiescent response style or a tendency to agree; disacquiescent response style or a tendency to disagree; and extreme response style or a tendency toward using the endpoints (Chen et al., 1995; Cronbach, 1946; Harzing, 2006; Si & Cullen, 1998). Harzing (2006) found differences in the use of the midpoint and endpoints of scales between participants answering the same survey in their native language (non-English) compared to those answering it in English. Specifically, those completing the English questionnaire used the midpoint more frequently than the endpoints, whereas they used the endpoints more often when the survey was in their native language.

While language is often a focal area of consideration when designing and validating international surveys (Johnson et al., 2018; Smith, 2018), the diversity of language within one individual country may be overlooked. Data from the United States Census indicates that between 1980 and 2019, the proportion of people primarily speaking a language other than English in their home increased from 10 to 20% (Dietrich & Hernandez, 2022b). Among those whose primary language spoken at home was not English, the top languages were Spanish or Spanish Creole (61.6%), Chinese (5.2%), Tagalog (2.6%), Vietnamese (2.3%), and Arabic (1.9%) in 2019 (Dietrich & Hernandez, 2022b).

Those primarily speaking a language other than English at home vary in terms of their English proficiency (Dietrich & Hernandez, 2022a). Language development is thought to occur across four domains that combined encompass expressive and interpretive communication: listening, speaking, reading, and writing (WIDA, 2020). While highly related, these domains are distinct and require unique skills (Powers, 2010). Among those who primarily speak a language other than English at home, 81% of this group speak English "well" or "very well," while 13% speak English "not well" and 6% speak English "not at all" (Dietrich & Hernandez, 2022a). However, when completing an online survey, reading and comprehension skills are arguably the most critical domain of language development. Yet, 2017 data from the international Survey of Adult Skills indicate that 17.5% of adults in the United States scored in the lowest literacy proficiency level (OECD, n.d.). This finding suggests that literacy skills may interfere with survey understanding for over one in six survey respondents.

The purpose of this research study is to examine the relationship between respondents' language background, the reading difficulty of survey items, and item response within the domain of personal finance-related surveys. To our knowledge, this has not yet been investigated within personal finance. Language may have heightened relevance in this domain, given the domain-specific vocabulary knowledge required to accurately respond to the survey items. Using data collected by the National Endowment for Financial Education (NEFE) to facilitate the validation of personal finance scales across diverse populations, this study aims to provide an example of why participants' language background may be an important consideration in scale validation research. Thus, this exploratory correlational study addresses three research questions:

1. What is the reading difficulty of measures within the personal finance domain?
2. Are there differences in the variability of personal finance survey responses at the item level and score level depending on whether individuals primarily use English at home, and are these differences exacerbated by reading difficulty?
3. Are there differences in response styles for personal finance measures depending on whether individuals primarily use English at home or another language, item reading difficulty, and their interaction?

We hypothesized that, depending on the primary language used at home, there would be differences in the variability of responses at the item level and score level (H1) and differences in response styles (H2). Furthermore, for surveys administered in English, we hypothesized that reading difficulty would have a stronger association with item response for individuals whose primary language used at home was *not* English (H3).

Importantly, we recognize that using a language other than English at home does not mean that survey respondents cannot also be highly proficient in English. For example, in the United States, children of Korean immigrant parents may use Korean at home but be highly skilled in English through their schooling. Nevertheless, individuals who primarily use English at home likely use English more often, on average, than those who primarily use other languages at home. Higher engagement with the English language may make it easier to respond to survey items in English. Furthermore, personal finance surveys primarily pertain to participants' home life, and thus, participants who use languages other than English at home may be more familiar with financial concepts in the other language, irrespective of their general reading proficiency.

Methodology

Data Collection

The data set used for the current study comes from a larger survey research study on measures of financial knowledge, attitudes, and behaviors sponsored by the National Endowment for Financial Education (NEFE). Data were collected online from individuals in the United States by Qualtrics Panel Services between late October 2022 and mid-January 2023. The survey was available in English and Spanish. However, only five participants completed the survey in Spanish. On average, participants took 21.82 min (SD = 25.24) to complete the survey.[1] Participants also received a reward from Qualtrics for their time, in accordance with Qualtrics's standard proprietary incentive structure.

Participants

Overall, 993 participants had complete data on the survey questions related to key measures. However, given the research focus, the 10 participants who did not provide information on their primary language used at home were excluded from the current analysis, as were the five participants who completed the survey in Spanish. Of the 978 remaining participants, 826 reported that their primary language used at home was English, while 152 reported 24 other languages. The languages reported by at least five participants were Spanish (n = 30), American Sign Language (n = 26), Armenian (n = 19), Arabic (n = 18), Bengali, Panjabi, Nepali, or other Indic languages (n = 11), Afrikaans (n = 9), and Cantonese, Mandarin, or other Chinese languages (n = 9).

Measures

The present study included an initial analysis of 11 personal finance scales:

1. Brief money management scale (Ksendzova et al., 2017)
2. Domain-specific risk-taking scale (Blais & Weber, 2006)
3. Executive personal finance (Spinella et al., 2007)
4. Financial decision-making ability scale (Warmath, 2022)
5. Financial self-efficacy scale (Warmath, 2022)
6. Financial skill scale (Consumer Financial Protection Bureau, 2018)
7. Financial well-being scale (Netemeyer et al., 2018)
8. Material values scale (Richins, 2004)
9. Personal savings orientation scale (Dholakia et al., 2016)
10. Propensity to overspend on credit cards scale (Sotiropoulos & d'Astous, 2013)
11. Propensity to plan for money long-term (Lynch et al., 2010).

These 11 scales were selected because they were included in a larger survey research study on measures of financial knowledge, attitudes, and behaviors sponsored by the National Endowment for Financial Education (NEFE). The NEFE study sought to examine the measurement properties of widely used personal finance measures across diverse populations. More details on each scale can be found in their corresponding papers.

Reading difficulty for each scale and item was assessed using the Flesch Reading Ease and Flesch-Kincaid Grade Level (Flesch, 1948, 1979; Kincaid et al., 1975). Flesch Reading Ease is reported on a scale of 0 to 100, where higher scores indicate that the text is easier to read. Flesch (1979) notes that a score of at least 60 is needed to constitute "Plain English" text. Flesch-Kincaid Grade Levels correspond to

[1] The response time data should be interpreted with caution, since respondents can leave the survey and return to it later. Prior to computing the mean, three observations were excluded given their very high response times that did not seem reliable (i.e., 670 min, 2944 min, and 14,192 min). The mean response time prior to removing these three outliers was 39.96 min (SD = 463.68).

specific school grade levels – for example, a Flesch-Kincaid Grade Level of 5 corresponds to a reading level associated with fifth grade in the United States, where students are typically 10–11 years old (Colorado Department of Education, n.d.). These statistics consider the average sentence length (ASL) and average syllables per word (ASW). Furthermore, these statistics can be calculated automatically in Microsoft Word, and the following formulas are used by the software (Kincaid et al., 1975; Microsoft, n.d.):

$$\text{Flesch Reading Ease} = 206.835 - (1.015 \times ASL) - (84.6 \times ASW)$$

$$\text{Flesch} - \text{Kincaid Grade Level} = (.39 \times ASL) + (11.8 \times ASW) - 15.59$$

To further put these scores into context, Flesch (1979) analyzed various texts and provided several examples of types of readings falling into certain categories. For instance, a random sample of texts in the *New York Times* had a reading ease of 39, corresponding to college-level reading. Flesch also found that the typical consumer magazine ad had a reading ease of 82, corresponding to sixth grade, in which students in the United States are usually between 11 and 12 years old (Colorado Department of Education, n.d.).

In relation to our study, Flesch Reading Ease and Flesch-Kincaid Grade Level values are not designed to measure reading challenges encountered specifically by non-native English speakers. Rather, historically these scales were used to ensure that training materials for Navy personnel were developed at an appropriate reading level (Kincaid et al., 1975). Currently, in seeking Institutional Review Board approval for a research project, researchers are commonly required to provide a readability index score for research consent forms to ensure that participants can understand the risks and benefits associated with a research study (Paasche-Orlow et al., 2003). Therefore, while we are particularly interested in how reading difficulty may complicate the survey response process for those whose primary language used at home is not English, reading difficulty—as operationalized in our study—is relevant to all respondents.

Data Analysis

The overarching goal of the data analysis was to understand whether there were differences in the variance of item responses by primary language used at home and whether these differences were exacerbated by item reading difficulty. This broad aim was addressed through multiple steps. Data were analyzed using SPSS 29 (IBM Corp., 2022), HLM 8 (Scientific Software International, Inc., 2021), and Stata 18 (StataCorp LLC, 2023).

First, to answer Research Question 1, we summarized data on the Flesch Reading Ease and Flesch-Kincaid Grade Level (Flesch, 1948, 1979; Kincaid et al., 1975) values for all 11 scales. We used this data to select scales of varying reading difficulty for further analysis in Research Questions 2 and 3.

Second, Research Question 2 was addressed by conducting Levene's test of homogeneity of variance. This test was used to examine differences in variance at the item level and scale level in responses between individuals whose primary language used at home was English and those primarily using other languages at home. Additionally, independent means t-tests were also computed on each item and scale to provide additional context on differences in scores between groups. The total scale scores represented a raw sum of responses. Next, two separate regression models were computed for each of three groups: (1) respondents whose primary language used at home was English, (2) respondents whose primary language used at home was a language other than English, and (3) all respondents. In these models, the mean and group standard deviation for each item, respectively, was regressed on the Flesch-Kincaid grade level associated with the item. We used grade level rather than reading ease for this analysis, since grade level is more directly interpretable.

Third, the purpose of Research Question 3 was to examine differences in "response styles" (1) by primary language used at home, comparing those who primarily use English at home to all other languages, (2) by item reading difficulty, and (3) by the interaction of these person and item characteristics. Four types of response styles were examined for each participant on each scale: (1) acquiescence (i.e., tendency to agree or endorse), (2) disacquiescence (i.e., tendency to disagree or not endorse), (3) use of midpoints, and (4) use of endpoints. Response styles were analyzed using bivariate and multivariate analyses. The bivariate analyses were used to examine aggregated response styles across a given scale, while the multivariate analyses were used to model individual item responses in relation to particular response styles.

Specifically, for the bivariate analyses, each of the five scales received a score for each response style, with the score corresponding to the number of times the participant responded in a manner aligning with the response style. Hence, the total possible score for each response tendency corresponded to the total number of items in the scale. Independent sample t-tests were used to examine differences in response style scores for each scale by primary language used at home.

Next, for the multivariate analyses, we used separate cross-classified random effects logistic regression models to predict each of the four response styles. These models are multilevel models in which survey item responses (level one) are crossed within the persons and items (both at level

two) examined in this study (Hoffman, 2015; Raudenbush et al., 2019). Therefore, while the independent samples t-tests examined response style scores in the aggregate across each of the five scales, these multilevel models examine individual item responses. While both techniques provide insightful information, the advantage of the multilevel approach is that we can model responses on individual items as a function of person characteristics, item characteristics, and their interactions, while controlling for other covariates. These models controlled for age, income, gender, educational attainment, and race/ethnicity, as well as the item response options available (e.g., strongly disagree to strongly agree vs. never to always).

Moreover, to add further context to our findings, we examined differences in demographic characteristics, survey response times, and financial experiences, depending on the primary language used at home. These financial experiences include the number of credit cards the participant had, if any, and whether the participant had various types of financial accounts or products (e.g., checking, retirement).

When conducting bivariate analyses, the Benjamini–Hochberg procedure was also used to correct for multiple comparisons in families of related tests by primary language (Benjamini & Hochberg, 1995; Donald, 2024): (1) demographic comparisons, (2) financial experiences, (3) personal finance score variances, (4) personal finance score means, (5) personal finance item variances, (6) personal finance item means, (7) response styles, and (8) the simple linear regression models. The Benjamini–Hochberg procedure is designed to reduce type 1 error rates in statistical testing. As the number of statistical tests increases, the chance of falsely rejecting a true null hypothesis increases. In brief, the procedure first rank orders *p*-values obtained from related tests from low to high. Next, rather than directly comparing the observed p-value to alpha (often set to .05), each *p*-value is compared against an adjusted threshold that accounts for the total number of tests in a family of related tests, the pre-set alpha level set by the researcher, and the rank order of the *p*-value. See Donald (2024) for more details on this procedure. We explicitly mention when this procedure results in a change in which the tests are statistically significant.

Finally, indicators of effect size are included to allow for further interpretation of statistically significant findings. Phi and Cramer's V were provided for tests of association between two categorical variables, with Phi being used for two variables each with two categories (e.g., Variable 1 has two categories, "Yes" and "No," and Variable 2 has two categories, "Has obtained a bachelor's degree" and "Has not obtained a bachelor's degree") and Cramer's V used when at least one variable had more than two categories. Rea and Parker (2014) provide the following lower thresholds for weak, moderate, relatively strong, strong, and very strong effect sizes for Phi and Cramer's V: maximum values of .10, .20, .40, .60, and .80, respectively. Cohen's *d* was provided for independent means t-tests using the following thresholds for small, medium, and large effect sizes: an absolute value of .20, .50, and .80, respectively (Cohen, 1992). For the linear regression models with one predictor, Cohen (1988) suggests that an R-squared of .010 corresponds to a small effect, .059 to a medium effect, and .138 to a large effect. For the multilevel models, the odds ratios provide an indicator of the strength of the relationship between the independent variables and the outcome (Chen et al., 2010). Finally, to our knowledge, there is no well-recognized indicator of effect size for Levene's test.

Results

Preliminary Comparisons on Demographic, Survey Response Times, and Financial Experiences

Table 1 provides information about the following demographic characteristics: sex, education, race/ethnicity, age, and income. Data are provided for participants overall (N = 978) and separately for participants who primarily use English at home (N = 826) and those who primarily use other languages at home (N = 152). Overall, 66% of participants reported that their sex assigned at birth was female; 29% reported male; 1% reported "other"; and 5% did not provide information about their sex assigned at birth. Participants who primarily use languages other than English were less likely to report their sex at birth (14% had unknown data compared to 3% who primarily speak English) ($p = .000$). In both primary language groups, a higher proportion of participants who provided data identified as female than male.

Overall, 29% of participants had a bachelor's degree or higher. A higher proportion of those primarily using languages other than English at home had a bachelor's degree or higher (33% vs. 29%) ($p = .000$). However, those who primarily used languages other than English were more likely to have less than a high school degree (13% vs. 4%) ($p = .000$), reflecting considerable variation in educational background among those who primarily use languages other than English at home.

With respect to race/ethnicity, overall, 19% of participants identified as Hispanic, 41% as non-Hispanic White, 21% as non-Hispanic Black, 7% as non-Hispanic Asian, and 8% as non-Hispanic other race, while 5% had an unknown racial/ethnic identity. Notably, a higher proportion of those primarily using languages other than English identified as Hispanic (35% vs. 16%) or did not provide information about their racial/ethnic background (14% vs. 3%) ($p = .000$). Conversely, a lower proportion of those primarily

Table 1 Select demographic characteristics of participants overall and by primary language used at home[a]

	All languages		English		Not English		Significance testing	
	N	%	N	%	N	%	Sig	Phi/Cramer's V
Total	978	100%	826	100%	152	100%		
*Sex**							0.000	0.244
Female	641	66%	571	69%	70	46%		
Male	279	29%	224	27%	55	36%		
Other	10	1%	4	0%	6	4%		
Unknown	48	5%	27	3%	21	14%		
*Education**							0.000	0.18
Less than high school	56	6%	37	4%	19	13%		
High school or equivalent	294	30%	245	30%	49	32%		
Some college credit/no degree	211	22%	192	23%	19	13%		
Associate's degree	131	13%	116	14%	15	10%		
Bachelor's degree	167	17%	146	18%	21	14%		
Graduate/professional degree	119	12%	90	11%	29	19%		
*Race/ethnicity**							0.000	0.31
Hispanic	182	19%	129	16%	53	35%		
Non-Hispanic White	403	41%	384	46%	19	13%		
Non-Hispanic Black	201	21%	174	21%	27	18%		
Non-Hispanic Asian	65	7%	52	6%	13	9%		
Non-Hispanic Other Race	79	8%	60	7%	19	13%		
Unknown	48	5%	27	3%	21	14%		

	Mean	SD	Mean	SD	Mean	SD	Sig	Cohen's d (Not English-English)
Age*	40.32	17.16	41.42	17.38	34.36	14.58	0.000	-0.42
Income*	$58,310.33	$ 53,012.71	$57,239.71	$51,182.22	$64,128.29	$ 61,892.38	0.197	0.13

[a]Bold values in the "Sig" column correspond to p-values that are statistically significant at the alpha .05 level of lower. SD=Standard deviation. The mean age and income are estimated using the midpoint of the age and income brackets reported by respondents. For example, a reported age of 20 to 24 years old was first converted to 22 before computing the mean. Cohen's d corresponds to the "Not English" group minus the "English" group. *Indicates that the differences in demographics remained significant after using the Benjamini and Hochberg (1995) procedure to control for the five related tests

Table 2 Overall survey response time in minutes by language group [a]

	All languages	English	Not English	Sig	Cohen's d (Not English-English)
N	975	824	151		
Mean	21.82	23.15	14.60	**0.000**	-0.341
Standard Deviation	25.24	26.46	15.24		
Median	16.12	17.00	9.13		
Minimum	3.22	4.67	3.22		
Maximum	271.33	271.33	118.45		

[a]Prior to the independent samples t-test, three observations with response times above 300 min were excluded, corresponding to 670 min, 2944 min, and 14,192 min. Two out of three of these responses were for those whose primarily language used at home was English

Bold values indicates to p-values that are statistically significant atthe alpha .05 level or lower.

using languages other than English at home identified as non-Hispanic White (13% vs. 46%) (p=.000).

The average age overall was 40.32 years old (SD=17.16), but the age was significantly lower for those primarily using languages other than English (M=34.36, SD=14.58) than those primarily using English (M=41.42, SD=17.38) (p=.000). Finally, the mean income for all participants was $58,310.33 (SD=$53,012.71), and there were no significant differences by primary language.

Table 2 displays descriptive statistics for the response time data. As discussed in the Methodology section, participants spent an average of 21.82 min (SD=25.24) completing the full survey. Individuals who primarily use English (M=23.15, SD=26.46) spent 8.55 more minutes, on average, completing the survey than those using other languages (M=14.60, SD=15.24) (p=.000).

Table 3 provides information about participants' financial experiences overall and by primary language used at home. Overall, 71% of participants reported that they or a member of their household had a checking account; 58% reported a savings or money market account, or certificate of deposit; 38% reported retirement investments through their employer (e.g., pension, Thrift savings, 401(k), 403(b)); 28% reported

Table 3 Select financial experiences for participants overall and by primary language used at home [a]

	All languages		English		Not English		Significance testing	
	N	%	N	%	N	%	Sig	Cramer's V
Total	978	100%	826	100%	152	100%		
Individual respondent or a member of the household has:								
*Checking account**							**0.000**	0.23
Yes	695	71%	622	75%	73	48%		
No	118	12%	92	11%	26	17%		
Don't know/prefer not to disclose	165	17%	112	14%	53	35%		
*Savings, money market, or certificate of deposit**							**0.000**	0.17
Yes	566	58%	505	61%	61	40%		
No	239	24%	193	23%	46	30%		
Don't know/prefer not to disclose	173	18%	128	15%	45	30%		
*Retirement investments through employer (e.g., pension, Thrift savings, 401(k), 403(b))**							**0.000**	0.13
Yes	374	38%	327	40%	47	31%		
No	411	42%	354	43%	57	38%		
Don't know/prefer not to disclose	193	20%	145	18%	48	32%		
*Retirement investments not through your employer (e.g., IRA, Keogh, SEP)**							**0.000**	0.15
Yes	275	28%	235	28%	40	26%		
No	516	53%	454	55%	62	41%		
Don't know/prefer not to disclose	187	19%	137	17%	50	33%		
*Non-retirement investments**							**0.003**	0.11
Yes	246	25%	210	25%	36	24%		
No	529	54%	460	56%	69	45%		
Don't know/prefer not to disclose	203	21%	156	19%	47	31%		
*College savings plan (529)**							**0.000**	0.19
Yes	125	13%	101	12%	24	16%		
No	642	66%	572	69%	70	46%		
Don't know/prefer not to disclose	211	22%	153	19%	58	38%		
*Number of credit cards**							**0.001**	0.15
No credit cards	195	20%	175	21%	20	13%		
1	203	21%	177	21%	26	17%		
2 to 3	293	30%	246	30%	47	31%		
4 to 8	130	13%	106	13%	24	16%		
9 or more	104	11%	74	9%	30	20%		
Don't know/prefer not to disclose	53	5%	48	6%	5	3%		

[a]*Indicates that the differences in financial experiences remained significant after using the Benjamini and Hochberg (1995) procedure to control for the seven related tests

Bold value indicates to p-values that are statistically significant atthe alpha

retirement investments not through their employer (e.g., IRA, Keogh, SEP); 25% reported non-retirement investments; and 13% reported a college savings plan (i.e., 529 plan). For each account or investment type, there were statistically significant differences in the distribution of participants responding "yes," "no" or "don't know"/ "prefer not to disclose" by primary language used at home ($p = .003$ or lower in all cases). Notably, participants who primarily used a language other than English were more likely to say "don't know" or "prefer not to disclose" for all for forms of financial engagement. For example, when asked about checking accounts, 35% of those who primarily used a language other than English at home reported "don't know" or "prefer not to disclose" compared to only 14% of those primarily using English. Additionally, overall, 20% of participants had no credit cards; 21% had one credit card; 54% had two or more credit cards; and 5% did not know or want to disclose how many credit cards they owned. There were statistically significant differences in the reported number of credit cards a participant had by primary language ($p = .001$), such that individuals who primarily used languages other than English were less likely to have no credit cards (13% vs. 21%) and more likely to have nine or more credit cards (20% vs. 9%).

Reading Difficulty of Personal Finance Scales

To address Research Question 1, the Flesch Reading Ease and Flesch-Kincaid Grade Levels were calculated for the 11 scales explored in this study (Table 4). The reading difficulty is based on the item/question text only and does not include the question stem or response options. Using Flesch's (1979) standards, the Flesch Reading Ease was on average 73.7 for all scales, corresponding to a "fairly easy" reading level, and ranged from 53.8 on the Brief Money Management Scale (BMMS) to 95.9 on the Propensity to Overspend on Credit Cards (POCC) Scale. Using the Flesch Reading Ease standards, these scales range from "fairly difficult" to "very easy" to read. In relation, the Flesch-Kincaid Grade Level was 5.9 on average, corresponding to approximately a sixth-grade reading level in the United States. The scales ranged from a grade level of 2.8 for the POCC to 7.8 for the BMMS.

We focused on a subset of scales for the remaining analyses. First, we selected the four-item POCC and the 14-item BMMS, since they corresponded to the scales with the lowest and highest grade levels, respectively. BMMS typically includes four subscales: Cash, Credit, Insurance, and Save. While we would have liked to include all four subscales, the secondary data set used in the present study excluded the Save subscale, and thus, the three remaining subscales were analyzed separately rather than as a full scale. The Flesch Reading Ease values for the BMMS Cash, Credit, and Insurance subscales were 35.6, 86.3, and 27.4, respectively, while the Flesch-Kincaid Grade Levels were 10.1, 3.6, and 11.4, respectively. Thus, the Credit subscale is notably much easier to read than the other subscales. As a reminder, the Flesch Reading Ease and Flesch-Kincaid Grade Levels are not specific to this sample, because they are calculated based on the question text only.

Table 4 Flesch reading ease and Flesch-Kincaid grade level for each scale[a]

	Flesch Reading ease	Flesch-Kincaid grade level
Brief money management scale (BMMS)	53.8	7.8
Financial decision-making ability (FDMA)	63.6	7.5
Financial self-efficacy (FSE)	68.8	6.9
Propensity to plan for money long-term (PMLT)	76.2	6.9
Personal savings orientation (PSO)	69.0	6.7
Financial skill (FS)	69.9	6.1
Domain-specific risk-taking (DOSPERT) scale	71.5	6.0
Financial well-being (FWB)	75.4	5.4
Material values scale (MVS) (15-item version)	79.7	5.0
Executive personal finance (EPF)	87.3	3.4
Propensity to overspend on credit cards (POCC)	95.9	2.8

[a] For BMMS, only three (Cash, Credit, and Insurance) of the four subscales were included, since one (Save) was not included in the survey instrument used in the study. Flesch Reading Ease ranges from 0 to 100, with higher values corresponding to easier reading levels

Second, we selected the eight-item Financial Self-Efficacy Scale (FSE) because its grade level was in the middle range, and the scale was of interest to multiple members of the research team. Also, the FSE Scale used a "Not at all true" to "Completely true" scale which, offered a contrast to the POCC which, used "Strongly disagree" to "Strongly agree" and the BMMS, which used "Never" to "Always." This better ensured that our findings were not dependent on specific response categories.

Variability in Responses by Primary Language and in Relation to Reading Difficulty

Research Question 2 was addressed through a series of steps. First, in terms of scoring, the BMMS: Cash, BMMS: Credit, and BMMS: Insurance subscales each had a five-point scale ranging from 1 = Never to 5 = Always. Since the BMMS: Cash Subscale had four items, possible values for total scores ranged from 4 to 20. In comparison, the other two subscales had five items, and thus, possible values for total scores ranged from 5 to 25. Higher scores on each of these subscales correspond to greater engagement in the behavior, meaning management of cash flow, healthy credit, and appropriate insurance coverage, respectively.

The POCC had a seven-point scale ranging from 1 = Strongly disagree to 7 = Strongly agree. Since there were four items, the possible values for total scores ranged from 4 to 28. Higher scores correspond to higher propensities to overspend on credit cards.

The FSE had a five-point scale ranging from 1 = Not at all true to 5 = Completely true. Since there were eight items, the possible values for FSE total scores ranged from 8 to 40. Higher scores correspond to higher financial self-efficacy.

Levene's Homogeneity of Variance Test indicated that the variance in total scores did not differ significantly by primary language (see Table 5). Independent means t-tests indicated that individuals whose primary language used at home was not English scored higher, on average, on the BMMS: Credit (M = 14.69, SD = 5.49 vs. M = 12.51, SD = 5.31) (p = .000) and the BMMS: Insurance (M = 15.59, SD = 5.13 vs. M = 14.51, SD = 5.31) (p = .021) subscales, while they scored lower on the FSE scale (M = 25.40, SD = 7.52 vs. M = 27.00, SD = 6.44) (p = .006). These findings suggest that those primarily using languages other than English at home more often engaged in behavioral money management practices related to managing their credit and insurance coverage but also reported lower financial self-efficacy than participants who primarily used English at home. Please refer to Table 5.

Table 5 Scale means and standard deviations—sorted by Flesch-Kincaid grade level[a]

Scale	Flesch Kincaid Grade Level (Sorted Low to High)	Independent means test				Levene's homogeneity of variance test		
		Mean for English	Mean for Not English	Sig	Cohen's d (Not English- English)	SD English	SD Not English	Sig
POCC	2.8	15.16	16.09	0.106	0.14	6.66	6.51	0.303
BMMS: Credit^	3.6	12.51	14.69	**0.000**	0.41	5.31	5.49	0.562
FSE^	6.9	27.00	25.40	**0.006**	-0.24	6.44	7.52	0.075
BMMS: Cash	10.1	12.98	13.00	0.952	0.01	3.83	4.20	0.082
BMMS: Insurance^	11.4	14.51	15.59	**0.021**	0.20	5.31	5.13	0.290

[a]*Indicates that the differences in variance remained significant after the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995). ^Indicates that the differences in means remained significant after the Benjamini–Hochberg procedure

Bold values in the "Sig" column correspond to p-values that are statistically significant atthe alpha .05 level or lower.SD=Standard deviation

Table 6 Item means and standard deviations – sorted by Flesch-Kincaid grade level[a]

Scale and item number	Flesch- Kincaid grade level (Sorted Low to High)	Independent means test				Levene's homogeneity of variance test		
		Mean for English	Mean for not English	Sig	Cohen's d (Not English—English)	SD English	SD Not English	Sig
BMMS: Credit 5^	0.5	2.61	3.01	**0.000**	0.32	1.26	1.33	0.652
POCC 1	1.0	3.88	4.07	0.263	0.10	1.89	1.82	0.417
FSE 1^	1.8	3.37	3.11	**0.004**	-0.26	1.01	1.18	0.115
BMMS: Credit 4^	2.3	2.42	3.00	**0.000**	0.44	1.30	1.43	0.125
BMMS: Credit 3^	2.4	2.28	2.86	**0.000**	0.44	1.32	1.37	0.767
POCC 4	2.5	3.93	4.00	0.652	0.04	1.85	1.86	0.985
POCC 3	3.6	3.78	3.97	0.254	0.10	1.90	1.82	0.198
FSE 3^	3.8	3.51	3.15	**0.000**	-0.33	1.06	1.23	0.082
POCC 2^	4.4	3.57	4.06	**0.003**	0.26	1.89	1.82	0.058
BMMS: Credit 2^	5.6	2.62	2.97	**0.004**	0.26	1.36	1.36	0.235
BMMS: Cash 1	6.4	3.29	3.37	0.466	0.06	1.20	1.25	0.258
FSE 2	6.7	3.32	3.13	**0.039**	-0.18	1.03	1.20	0.074
FSE 4*	6.7	3.33	3.18	0.158	-0.14	1.01	1.21	**0.002**
BMMS: Credit 1^	7.0	2.57	2.84	**0.020**	0.20	1.35	1.30	**0.020**
FSE 5	7.5	3.35	3.19	0.151	-0.14	1.07	1.26	**0.014**
FSE 7*^	7.5	3.47	3.18	**0.009**	-0.27	1.03	1.25	**0.001**
BMMS: Insurance 5^	9.5	2.84	3.19	**0.003**	0.28	1.24	1.33	**0.030**
BMMS: Cash 3	9.5	3.14	3.11	0.772	-0.03	1.33	1.30	0.369
BMMS: Insurance 3*^	10.3	2.71	3.01	**0.014**	0.21	1.42	1.34	**0.008**
FSE 6	10.5	3.39	3.28	0.218	-0.11	1.06	1.12	0.468
BMMS: Insurance 2	11.1	3.25	3.28	0.800	0.02	1.42	1.29	**0.025**
BMMS: Cash 2	11.1	3.23	3.2	0.791	-0.03	1.13	1.26	**0.013**
FSE 8*	11.5	3.27	3.19	0.454	-0.07	1.07	1.22	**0.009**
BMMS: Insurance 4*	12.6	2.78	2.99	0.082	0.14	1.47	1.34	**0.001**
BMMS: Insurance 1	14.0	2.93	3.12	0.106	0.13	1.47	1.34	**0.047**
BMMS: Cash 4	14.2	3.32	3.33	0.957	0.00	1.21	1.20	0.774

[a]*Indicates that the differences in variance remained significant after the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995). ^Indicates that the differences in means remained significant after the Benjamini–Hochberg procedure

Bold values in the "Sig" column correspond to p-values that are statistically significant atthe alpha .05 level or lower. SD=Standard deviation.

Next, Table 6 displays means and standard deviations for each item by primary language used at home. There were differences in the variances for 11 out of the 26 items. None of the items on the POCC Scale, which had the lowest Flesch-Kincaid Grade Level, had significant differences.

However, after using the Benjamini–Hochberg procedure to adjust for multiple comparisons and reduce the likelihood of type 1 error (Benjamini & Hochberg, 1995; Donald, 2024), five of the 26 items remained significant: FSE 4, FSE 7, FSE 8, BMMS: Insurance 3, and BMMS: Insurance 4.

For the three FSE items, the variance was higher for those whose primary language was not English, while the variance was lower for this group on the two BMMS: Insurance items. Moreover, the items that remained significant after the adjustment had a Flesch-Kincaid Grade Level of 6.7 or higher.

As can be seen in Table 6, there were also differences in mean item responses for 12 out of the 26 items: BMMS: Credit 1, BMMS: Credit 2, BMMS: Credit 3, BMMS: Credit 4, BMMS: Credit 5, BMMS: Insurance 3, BMMS: Insurance 5, FSE 1, FSE 2, FSE 3, FSE 7, and POCC 2. The difference in means remained significant for all but one item (FSE 2) after the Benjamini–Hochberg adjustment (Benjamini & Hochberg, 1995; Donald, 2024). Overall, the means were higher on BMMS items for those primarily using languages other than English at home, but lower for FSE items.

In the final stage of answering Research Question 2, item-level means and standard deviations were separately regressed on Flesch-Kincaid Grade Levels for participants overall, participants who primarily used English at home, and participants who primarily used languages other than English at home. The means and standard deviations were calculated separately for each group.[2] Table 7 shows that the Flesch-Kincaid Grade Level was a significant predictor of aggregated item means for those whose primary language used at home was not English ($p = .012$). Overall, the Flesch-Kincaid Grade Level of the item explained 24% of the variation in the item means for those whose primary language used at home was not English. Additionally, for every one grade level increase in the Flesch-Kincaid Grade Level, the item mean was predicted to increase by 0.016 points. This corresponds to a standardized regression coefficient of 0.486. However, there was no significant association for those whose primary language used at home was English and for all languages when combined. Table 7 also shows that there was no association between Flesch-Kincaid Grade Level and item-level standard deviations for participants overall or for any specific language group.

The relationship between primary language used at home, item-level Flesch-Kincaid Grade Level, and item-level mean is also depicted in a scatterplot (Fig. 1). From the scatterplot, one can observe that item means are higher for those whose primary language is not English than for those whose primary language is English, except at the highest grade levels when the gap narrows. Also, while the regression line for those whose primary language is English appears steeper, which might suggest a stronger relationship, one can see that the observations are more tightly clustered around the regression line for those whose primary language is not English. Thus, there is less error in the prediction of item means from Flesch-Kincaid Grade Level for those whose primary language is not English.

Response Styles by Primary Language

Research Question 3 examined differences in four "response styles" by primary language used at home: (1) acquiescence, (2) disacquiescence, (3) use of the midpoint, and (4) use of the endpoints. Table 8 displays which response options correspond to each response style for each scale. Furthermore, for each response style, the total possible score corresponded to the total number of items in the scale.

Response styles were first examined at the scale level. Table 9 displays the mean prevalence of each response style for each scale by primary language used at home. There were statistically significant differences for five of the 20

Table 7 Linear regression model of aggregated item means and standard deviations on Flesch-Kincaid grade level by primary language used at home[a]

Population		Aggregated item mean					Aggregated item standard deviation				
		b	SE for *b*	Beta	Sig	R-squared	*b*	SE for *b*	Beta	Sig	R-squared
All Languages	(Constant)	2.849	0.120		0.000	0.113	1.197	0.053		**0.000**	0.028
	Flesch-Kincaid Grade	0.026	0.015	0.336	0.093		0.005	0.007	0.167	0.414	
English	(Constant)	2.822	0.135		0.000	0.103	1.174	0.059		**0.000**	0.039
	Flesch-Kincaid Grade	0.028	0.017	0.322	0.109		0.007	0.007	0.198	0.332	
Not English	(Constant)	2.994	0.047		0.000	0.236	1.270	0.029		**0.000**	0.000
	Flesch-Kincaid Grade	0.016	0.006	0.486	**0.012***		0.000	0.004	-0.017	0.936	

[a]*Indicates that the Flesch-Kincaid Grade Level remained a significant predictor after the Benjamini–Hochberg procedure controlled for the three-related regression models (Benjamini & Hochberg, 1995)

[2] Prior to conducting this pooled analysis across scales, the seven-point Propensity to Overspend on Credit Cards (POCC) scale was re-scaled to a five-point scale, such that values of 1 on the original seven-point scale were recoded to 1 on the re-scaled five-point scales; values of 2 were recoded to 1.67; values of 3 were recoded to 2.33; values of 4 were recoded to 3.00; values of 5 were recoded to 3.67; values of 6 were recoded to 4.33; and values of 7 were recoded to 5.00. Since the item means and item-level standard deviations were the dependent variable in the simple linear regression models, it was important that all items had the same range of 1 to 5. Otherwise, item-level means and standard deviations could be confounded with item reading difficulty because the seven-point POCC scale had the easiest to read items overall.

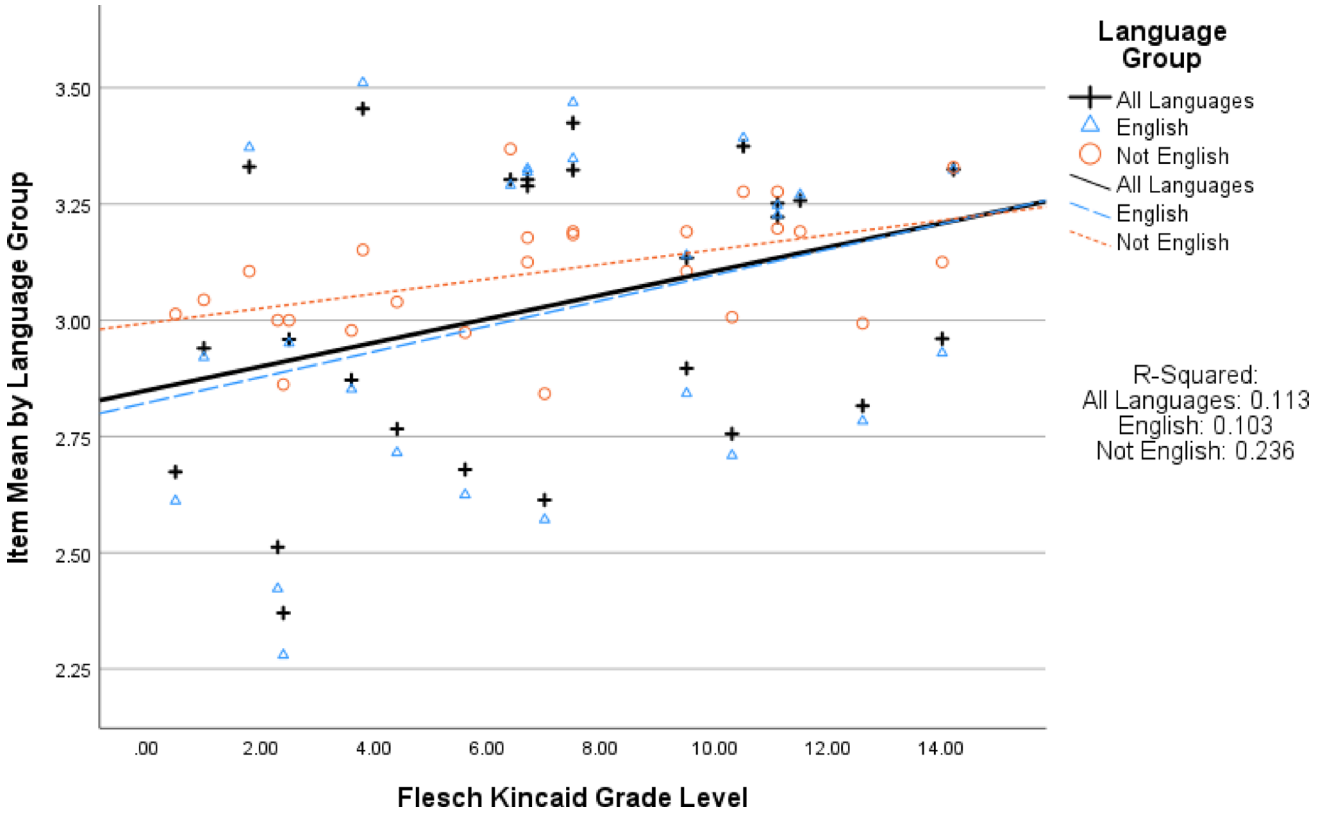

Fig. 1 Scatterplot of item means by language group regressed on Flesch-Kincaid grade level

Table 8 Response options corresponding to various response styles for each scale

	BMMS (Cash, Credit, and Insurance subscales)	FSE	POCC
Acquiescence (i.e., tendency to agree or endorse)	Often Always	Very true Completely true	Somewhat agree Agree Strongly agree
Disacquiescence (i.e., tendency to disagree or not endorse)	Rarely Never	Not very true Not at all true	Somewhat disagree Disagree Strongly disagree
Use of the midpoint	Sometimes	Somewhat true	Neither agree nor disagree
Use of the endpoints	Never Always	Not at all true Completely true	Strongly disagree Strongly agree

tests (i.e., four response styles by five scales). Notably, there were no response style differences observed for the BMMS: Cash and POCC scales.

After using the Benjamini–Hochberg procedure to adjust for multiple comparisons (Benjamini & Hochberg, 1995; Donald, 2024), only three of the 20 response styles had statistically significant differences. Specifically, acquiescence was lower on the BMMS: Credit Subscale for those whose primary language was English (M = 1.21, SD = 1.53) compared to those with other primary languages (M = 1.80, SD = 1.75) (p = .000). In comparison, disacquiescence was higher on the BMMS: Credit Subscale for those whose primary language was English (M = 2.51, SD = 1.92) compared to those with other primary languages (M = 1.94, SD = 1.75) (p = .000). Yet, disacquiescence was lower on the FSE Scale for those whose primary language was English (M = 1.37, SD = 2.00) compared to those with other primary languages (M = 2.19, SD = 2.45) (p = .000).

Response styles were then examined at the item level. Table 10 displays the results of the four cross-classified random effects logistic models, in which person and item characteristics were used to predict item response styles. First, in the model predicting respondents' acquiescence, there was a statistically significant main effect for participants' primary language used at home (Odds Ratio [OR] = 0.66, p = .004), but not for the item Flesch-Kincaid Grade level (OR = 1.02, p = .207). There was also a statistically significant interaction between these two variables for acquiescence (OR = 1.04, p = .000). Additionally, there were statistically significant main effects for primary language (OR = 1.54, p = .003) and item grade level (OR = 1.03, p = .042) and an interaction effect (OR = 0.97, p = .015) for the model predicting

Table 9 Mean use of response styles by primary language used at home[a]

	English		Not English		Significance test	
Scale	Mean	SD	Mean	SD	Sig	Cohen's d (Not English—English)
BMMS: Cash (out of 4)						
Acquiescence	1.73	1.47	1.62	1.54	0.388	-0.08
Disacquiescence	0.99	1.26	1.08	1.29	0.447	0.07
Use of midpoint	1.27	1.26	1.30	1.23	0.801	0.02
Use of endpoints	1.16	1.34	1.28	1.48	0.376	0.08
BMMS: Credit (out of 5)						
Acquiescence^	1.21	1.53	1.80	1.75	**0.000**	0.38
Disacquiescence^	2.51	1.92	1.94	1.75	**0.000**	-0.30
Use of midpoint	1.28	1.48	1.26	1.32	0.831	-0.02
Use of endpoints	2.10	1.85	1.82	1.80	0.082	-0.15
BMMS: Insurance (out of 5)						
Acquiescence	1.77	1.65	1.99	1.76	0.128	0.13
Disacquiescence	1.88	1.77	1.59	1.63	0.059	-0.17
Use of midpoint	1.35	1.46	1.42	1.51	0.610	0.05
Use of endpoints	2.16	1.82	1.78	1.83	**0.018**	-0.21
FSE (out of 8)						
Acquiescence	3.53	2.94	3.08	2.85	0.084	-0.15
Disacquiescence^	1.37	2.00	2.19	2.45	**0.000**	0.40
Use of midpoint	3.11	2.57	2.73	2.46	0.096	-0.15
Use of endpoints	1.65	2.39	2.20	2.65	**0.011**	0.23
POCC (out of 4)						
Acquiescence	1.51	1.60	1.68	1.62	0.228	0.11
Disacquiescence	1.56	1.63	1.45	1.54	0.445	-0.07
Use of midpoint	0.93	1.33	0.87	1.24	0.599	-0.05
Use of endpoints	1.07	1.53	0.93	1.49	0.300	-0.09

[a]^Indicates that the differences in means remained significant after the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995)

respondents' use of the midpoint, but not for respondents' use of the endpoints and disacquiescence.

Interaction plots can aid in the interpretation of interactions. However, marginal means were not available for the crossed-effects models in HLM 8. To facilitate interpretation, estimates of marginal means were obtained for these models in Stata 18 using a multilevel model in which responses are nested within persons and only fixed effects are obtained for item characteristics. These plots should be viewed as approximations used to illustrate relationships rather than precise estimates of marginal means.[3] Using the margin plot displayed in Fig. 2, those who primarily use languages other than English are more likely to acquiesce when responding to an item, although the gap between language groups decreases as reading grade level increases. Similarly, using the margins plot displayed in Fig. 3, the mean predicted probability of using the midpoint is higher for those who primarily use English at home. For those who primarily use languages other than English at home, the mean predicted probability of using the midpoint increases as the grade level increases.

[3] Comparisons of the results obtained using each type of model indicated that the regression coefficients were close.

Summary of Results

Three research hypotheses were stated in the introduction:

- H1: Depending on the primary language used at home, there would be differences in the variability of responses at the item level and score level.
- H2: Depending on the primary language used at home, there would be differences in response styles.
- H3: For surveys administered in English, reading difficulty would have a stronger association with item response for individuals whose primary language used at home was *not* English.

All three research hypotheses were partially supported. For H1, Levene's tests indicated that there were no differences in the variance in responses at the score level by primary language but there were differences for five of the 26 items.

For H2, the bivariate analyses indicated that there were no differences in use of the midpoint or endpoints of the five scales by primary language. However, there were differences in the acquiescence for the BMMS Credit scale and differences in disacquiescence for the BMMS Credit scale and Financial Self-Efficacy scale. The multivariate analyses

Table 10 Cross-classified random effects logistic regression models predicting response styles[a]

	Acquiescent Response				Disacquiescent Response			
	Coef	SE	OR	Sig	Coef	SE	OR	Sig
Fixed effects								
Intercept	-1.19	0.30	0.30	**0.000**	-0.29	0.31	0.75	0.360
Person characteristics								
Age	0.00	0.00	1.00	0.120	0.00	0.00	1.00	0.221
Income	0.000003	0.00	1.00	**0.000**	0.00	0.00	1.00	0.514
Sex (ref: Male)								
Female	-0.09	0.10	0.91	0.340	0.01	0.09	1.01	0.897
Other/unknown sex	-0.71	0.44	0.49	0.108	0.78	0.42	2.18	0.060
Education (ref: Less than high school)								
High school	0.27	0.20	1.31	0.170	0.01	0.19	1.02	0.936
Some college credit, no degree	0.35	0.21	1.41	0.093	-0.14	0.19	0.87	0.485
Associate's degree	0.57	0.22	1.77	**0.009**	-0.23	0.21	0.80	0.269
Bachelor's degree	0.62	0.21	1.86	**0.004**	-0.41	0.20	0.66	**0.039**
Graduate/professional degree	0.72	0.23	2.06	**0.002**	-0.48	0.22	0.62	**0.028**
Race/ethnicity (ref: Non-Hispanic White)								
Hispanic	0.08	0.13	1.08	0.538	-0.28	0.12	0.76	**0.023**
Non-Hispanic Black	0.42	0.12	1.53	**0.000**	-0.22	0.12	0.80	0.060
Non-Hispanic Asian	-0.15	0.18	0.86	0.411	-0.27	0.18	0.76	0.126
Non-Hispanic Other Race	0.05	0.17	1.05	0.759	-0.35	0.16	0.70	**0.030**
Unknown Race/ethnicity	0.32	0.48	1.38	0.501	-0.67	0.45	0.51	0.136
Primary Language Used at Home: English (ref: Not English)	-0.42	0.15	0.66	**0.004**	0.08	0.14	1.09	0.555
Item characteristics								
Flesch-Kincaid Grade	0.02	0.02	1.02	0.207	-0.05	0.02	0.96	0.053
Response Format (ref: Propensity to Use Credit Cards)								
Financial Self-Efficacy	0.04	0.18	1.04	0.843	-0.98	0.24	0.37	**0.000**
Brief Money Management	-0.52	0.18	0.59	**0.009**	0.21	0.24	1.24	0.380
Person by item characteristics								
Primary language used at home: English by grade	0.03	0.01	1.04	**0.000**	-0.01	0.01	0.99	0.363
	Variance	Sig			Variance			Sig
Random effects								
Person-variance	1.45	**0.000**			1.34			**0.000**
Item-variance	0.07	**0.000**			0.13			**0.000**
	Use of midpoint				Use of endpoint			
	Coef	SE	OR	Sig	Coef	SE	OR	Sig
Fixed effects								
Intercept	-1.39	0.27	0.25	**0.000**	-1.34	0.36	0.26	**0.000**
Person characteristics								
Age	-0.01	0.00	0.99	**0.012**	0.01	0.00	1.01	**0.026**
Income	-0.000003	0.00	1.00	**0.001**	0.00	0.00	1.00	0.219
Sex (ref: Male)								
Female	0.07	0.09	1.08	0.428	0.01	0.12	1.01	0.937
Other/unknown sex	0.06	0.42	1.06	0.885	0.16	0.55	1.18	0.766
Education (ref: Less than high school)								
High school	-0.18	0.19	0.84	0.343	0.04	0.25	1.04	0.866
Some college credit, no degree	-0.10	0.19	0.91	0.607	-0.23	0.26	0.80	0.369
Associate's degree	-0.20	0.21	0.82	0.331	-0.04	0.27	0.97	0.896
Bachelor's degree	-0.09	0.20	0.91	0.651	-0.34	0.26	0.71	0.200
Graduate/professional degree	-0.20	0.22	0.82	0.357	0.12	0.28	1.13	0.678
Race/ethnicity (ref: Non-Hispanic White)								
Hispanic	0.23	0.12	1.26	0.055	-0.50	0.16	0.61	**0.002**
Non-Hispanic Black	-0.22	0.12	0.81	0.064	-0.16	0.15	0.85	0.296
Non-Hispanic Asian	0.40	0.17	1.49	**0.023**	-0.65	0.23	0.52	**0.005**
Not Hispanic Other Race	0.29	0.16	1.34	0.068	-0.27	0.21	0.76	0.199
Unknown Race/ethnicity	0.31	0.45	1.37	0.493	-0.45	0.59	0.64	0.447

Table 10 (continued)

	Use of midpoint				Use of endpoint			
	Coef	SE	OR	Sig	Coef	SE	OR	Sig
Primary Language Used at Home: English (ref: Not English)	0.43	0.14	1.54	**0.003**	-0.14	0.18	0.87	0.421
Item characteristics								
Flesch-Kincaid grade	0.03	0.01	1.03	**0.042**	-0.01	0.02	0.99	0.536
Response format (ref: Propensity to Use Credit Cards)								
Financial self-efficacy	0.85	0.14	2.33	**0.000**	-0.32	0.21	0.73	0.152
Brief money management	0.27	0.13	1.30	0.058	0.79	0.21	2.20	**0.000**
Person by item characteristics								
Primary language used at home: English by grade	-0.03	0.01	0.97	**0.015**	0.00	0.01	1.00	0.672
	Variance	Sig			Variance			Sig
Random effects								
Person-variance	1.33	**0.000**			2.42			**0.000**
Item-variance	0.04	**0.000**			0.10			**0.000**

[a]See the footnote on Table 1 for details on how age and income were measured. The Propensity to Use Credit Cards response format is a seven-point scale from "Strongly Disagree" to "Strongly Agree." The Financial Self-Efficacy response format is a five-point scale from "Not at All True" to "Completely True." The Brief Money Management response format is a five-point scale ranging from "Never" to "Always."

Bold value indicates to p-values that are statistically significantat the alpha .05 level or lower

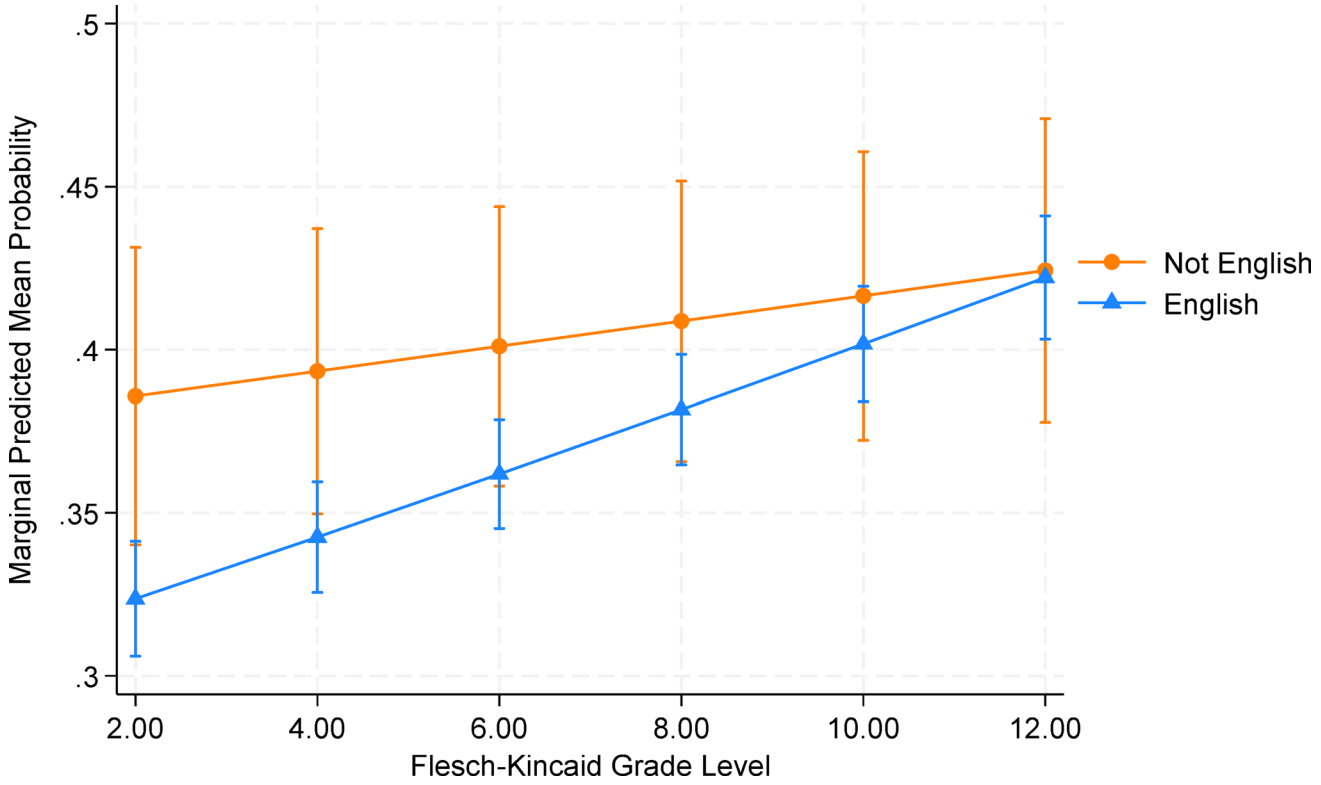

Fig. 2 Marginal predicted mean probabilities (with 95% confidence intervals) of acquiescence response style by language group and Flesch-Kincaid Grade level [a]. [a]*Note*. These margins were estimated using a multilevel model in which responses are nested within persons and only fixed effects were obtained for item characteristics. Since the crossed random effects logistic model could not be used to obtain marginal means, these plots should be viewed as approximations used to illustrate relationships rather than precise estimates of marginal means

found significant main effects for primary language used at home with respect to respondents' use of the midpoint and acquiescence.

Finally, for H3, there was no relationship between item-level standard deviation and reading difficulty for either language group analyzed, but there was a positive relationship between reading difficulty and item means for those whose primary language used was not English. Higher means are suggestive of acquiescence. Additionally, the multivariate analyses found a significant interaction effect for primary language and item grade level when examining respondents'

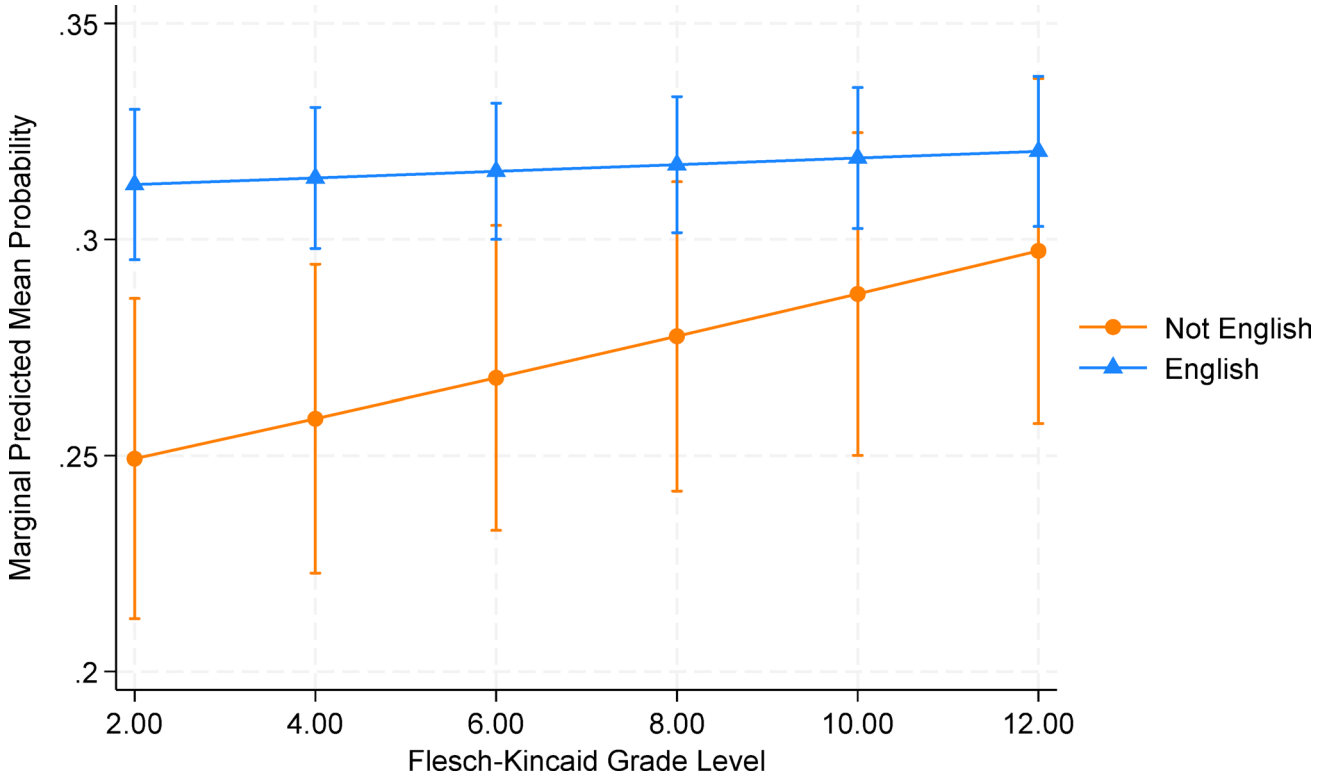

Fig. 3 Marginal predicted mean probabilities (with 95% confidence intervals) of using the midpoint by language Group and Flesch-Kincaid grade level.[a]*Note*. See note for Fig. 2

use of the midpoint and acquiescence, but not for use of the endpoint and disacquiescence.

Discussion

This exploratory study provides insights on the relationships among primary language used at home, the reading difficulty of survey items, and survey responses. Our insights are organized around four main themes: (1) Reading difficulty and survey response, (2) Response style differences by primary language, (3) Need for cognitive interviews with linguistically diverse groups, and (4) Limitations and opportunities for further research.

Reading Difficulty and Survey Responses

Reading difficulty appears to play some role in survey responses. Indeed, other researchers have emphasized the importance of language competencies in survey design. For example, Wenz and colleagues (2021) noted that when contacting potential non-native English speakers to participate in an English-language survey, one of the first major considerations is measuring competence in the English language. This is particularly important for the purposes of informed consent. According to UCLA (2023) guidelines, which follow federal regulations, researchers must provide translated consent materials and questionnaires for non-English speaking populations. Many universities require that bilingual speakers translate these survey materials (Duke University Health System, 2021; UCLA, 2023; Yale, 2015). While these policies may differ slightly for research that is considered to have minimal risk, researchers must take language competency seriously to meet IRB standards.

In the current study, we examined reading difficulty at the item level and total scale level. Oftentimes, for instruments primarily administered in one language, researchers may select scales based on their target construct and the age of their population without consideration of language. Moreover, available reliability and validity evidence for the scale may exclude information about language. After the Benjamini–Hochberg adjustment, we found no differences in total score variance by primary language, but differences on five of the 26 items were observed. Thus, disaggregating the data from the total score level down to the item level revealed differences that were not apparent at the total score level.

Furthermore, one of our most interesting findings occurred when item-level means were regressed on Flesch-Kincaid Grade Levels, with separate models for all participants, participants whose primary language used was English and participants whose primary language was not

English. We found that grade level explained 24% of the variance in item-level means for those whose primary language was not English, compared to 10% for those who primarily used English at home. Additionally, grade level was a significant predictor in the model only for those primarily using languages other than English. As reading difficulty increased, item-level means were predicted to increase. This finding may imply that as items become more difficult to read, participants exhibit more acquiescent tendencies, which is generally consistent with findings from the item-level response style analysis.

Response Style by Primary Language

Response styles or tendencies (Cronbach, 1946) have been explored in cross-cultural research (e.g., Harzing, 2006). As discussed earlier, Harzing (2006) found that participants completing a survey in their native language (non-English) used the endpoints more frequently than the midpoints, while those completing the survey in their non-native language (English) used the midpoint more than the endpoints.

However, overall, the current study provides mixed evidence of response style differences by primary language. For the bivariate analyses, out of the 20 response style differences tested (i.e., four response styles by five scales), there were three significant differences after adjusting for multiple comparisons via the Benjamini–Hochberg procedure. In two cases, the response style differences related to disacquiescence. For the BMMS: Credit Subscale, those who spoke English as their primary language at home were more likely to engage in disacquiescence, whereas they were less likely for the FSE Scale. One potentially important difference between the BMMS: Credit and FSE scales is that they use different response options. While the BMMS: Credit Subscale uses a five-point "Never" to "Always" scale, the FSE Scale uses a five-point "Not at all true" to "Completely true" scale. In comparison, primary language did not significantly predict disacquiescence in the multivariate analyses.

In the bivariate analyses, one response style difference was also found for acquiescence. Specifically, those whose primary language was not English engaged in more acquiescence for the BMMS: Credit Subscale. Regarding acquiescence, the multivariate analysis also found a significant main effect for primary language and an interaction effect between primary language and reading difficulty. Those whose primary language was not English had a higher probability of acquiescing and both language groups had increasingly similar probabilities of acquiescence as reading difficulty increased.

Furthermore, while there were no differences in the use of the midpoint in the bivariate analysis, the multivariate modeling revealed significant main and interaction effects for primary language and reading difficulty. Notably, those who primarily used English were more likely to use the midpoint, and as reading difficulty increased, use of the midpoint increased for those whose primary language was not English. In relation, both the bivariate and multivariate analyses indicated that there were no differences in the tendency to use the endpoint by primary language groups.

The bivariate analyses also indicated that there were no response style differences for the POCC Scale. The POCC scale differed from the others in potentially meaningful ways that could be explored in further research: it had (1) the lowest Flesch-Kincaid Grade Level, (2) used seven response options instead of five, and (3) used a "strongly disagree" to "strongly agree" scale.

Cognitive Interviews with Linguistically Diverse Participants

While the results discussed thus far provide interesting insights, they also raise many questions. For example, why do aggregated item means increase as the Flesch-Kincaid Grade Level increases for those whose primary language used at home is not English? Moreover, why were response style differences observed in some scales but not others?

We believe that cognitive interviews with potential participants of various English proficiencies and diverse linguistic backgrounds are needed to understand what cognitive processes may lead to these observed differences and what changes to surveys are needed to better measure linguistically diverse participants' experiences. Given the complexity of responding to survey items (Tourangeau et al., 2000), cognitive interviewing is often recommended during measurement development and validation. This may be especially significant when survey translation and interpreting options, considered important tools in multilingual survey development, are limited.

Cognitive interviewing is an established, widely-used method for exploring the target population's understanding of measures—and, therefore, testing content validity (Beatty & Willis, 2007; Boateng et al., 2018; Drennan, 2003). Cognitive interviewing is frequently conducted by inviting a member of the target population to read and answer the items aloud, one at a time, while verbally explaining their understanding of the meaning of the items and their reasoning for their answers. In other words, as participants think aloud, researchers can ascertain whether the items are being understood and answered as intended, whether any of the items are inadvertently confusing or difficult to answer, etc. and make adjustments to the measure as needed (see for example LeBaron-Black et al., 2022 and Nguyen, 2019). The interviewer may ask follow-up questions such as "Can you tell me in your own words what that question

was asking?" (Beatty & Willis, 2007, p. 290) or "You took a little while to answer that question—what were you thinking about?" (Beatty & Willis, 2007, p. 300) to aid in this process.

One can imagine how the cognitive interviewing process could lead to insights about how participants with varying levels of English proficiency are thinking, not only about the item content, but the response options and the question stems. The cognitive interviews could also identify instances where cultural references or slang are not understood by respondents (Pettit, 2015). From this process, items, response options, and/or question stems could be modified (see Dörnyei & Taguchi, 2010), or in some cases, the researcher may need to search for a new measure entirely. Pettit (2015) recommends keeping questions short. In relation, Trzcinski (1981) recommends involving members from linguistically diverse language backgrounds in the survey development process to reduce misunderstandings.

To use cognitive interviews to improve surveys for linguistically diverse populations, we recommend that researchers consider adopting the following procedures. First, linguistic diversity could be included as a criterion when selecting participants for cognitive interviews, just like researchers would ensure diversity with respect to gender, race/ethnicity, socio-economic status, and other characteristics. Second, cognitive interviewers could consider removing or revising items that were challenging to read to ensure high-quality data across linguistic backgrounds and literacy levels. Third, prior to the cognitive interviews, items could be screened for reading difficulty using the Flesch indicators described in this paper (or other measures), and more difficult questions could be removed or revised.

Limitations and Future Research Directions

While this exploratory research study provides insight into an underexplored area of personal finance measurement, there are several limitations of the current study that should be noted. Importantly, these limitations also highlight areas for further research.

First, the current study analyzed all individuals whose primary languages used at home were not English as one group (N = 152), including American Sign Language. This decision was primarily made due to sample size limitations. The largest primary language reported by the non-English group was Spanish, with only 30 participants. Since the nature of the differences we observed may change depending on the specific primary language used, we recommend that future research studies examine specific languages by targeting specific linguistic populations in the sampling design. For instance, Reid (1990) found that when asked to respond to a 5-point Likert scale, English language learners whose primary language was Japanese tended to avoid using the endpoints of the scale (i.e., "strongly agree" or "strongly disagree" in the study) and instead tended to use the mean of the scale. In fact, we might have found mixed evidence for response style differences in the present study because of the varying language backgrounds and cultures represented within the "non-English" group.

Second, given the sample size limitations and the study's exploratory nature, the present study provides limited insight on within language group heterogeneity. In addition to exploring specific languages other than English, future researchers should also consider an intersectional understanding of language that incorporates culture, immigrant background, education, and socioeconomic status. For example, in their analysis of native and non-native English speakers, Wenz et al. (2021) found differences in data quality indicators depending on whether participants were born within the country or immigrated to the country later. There was less evidence of primacy effects (i.e., selecting the first response provided) and higher completion rates for the self-completion questions for respondents who were born within their current country than those who immigrated later. The authors conclude, "Being born in the UK likely has the impact of improving English language even among respondents [who] self-identified as having a different first language" (Wenz et al., 2021, p. 87). In a related study, Kleiner et al. (2015) found that foreign-born survey respondents were more likely to use the "don't know" response. Kleiner et al. (2015) posit that linguistic and cultural "proximity" may impact response quality as foreign-born individuals who immigrate from countries that are further away, in terms of linguistic differences and cultural differences, may give lower-quality answers. Combined, results from Wenz et al. and Kleiner et al. suggest that it is not just language proficiency that may relate to personal finance survey responses, but cultural influences. Beyond these studies, we also recommend exploring other factors such as education and socioeconomic background.

Third, when administering surveys that use complex and technical vocabulary, such as the vocabulary found in personal finance research, it may be advisable to include questions about language proficiency. Researchers should explore approaches to measuring language proficiency within the survey instrument. In addition to incorporating standard survey questions, researchers might draw from educational researchers and practitioners and incorporate multimodal forms of communication (Kress & van Leeuwen, 2006). For example, next to specific questions, there could be an audio option to have the item read aloud or visual cues to assist the survey participant.

Conclusion

Linguistic diversity within the United States is increasing (Dietrich & Hernandez, 2022b), but is unclear to what extent language proficiency and background are considerations when designing English-based surveys in the United States. The current study provides mixed results for the relationships among primary language, item reading difficulty, and item response. However, the findings suggest that not only are reading difficulty and primary language used at home related to the acquiescence and tendency to use the midpoint response styles but also that the relationship between reading difficulty and select item response styles differ depending on language background. However, the cognitive process behind this relationship needs further exploration through cognitive interviews with linguistically diverse participants. Finally, we hope this study highlights language as an important consideration in survey development and will spark further research in this area across diverse language backgrounds and proficiencies.

Acknowledgements We thank Morgen Nations for her research assistance and the anonymous reviewers for their feedback.

Funding National Endowment for Financial Education.

Data Availability Data for the study were provided by the National Endowment for Financial Education (NEFE). Readers may contact NEFE for more information.

Declarations

Conflict of interest Support for this project was provided to the authors by the National Endowment for Financial Education in the forms of data for analysis and an honorarium for expert panel participation.

References

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 287–311. https://doi.org/10.1093/poq/nfm006

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series b: Statistical Methodology, 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blais, A. R., & Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making, 1*(1), 33–47. https://doi.org/10.1017/S1930297500000334

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Qui˜nonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, Article Article 149. https://doi.org/10.3389/fpubh.2018.00149

Brown, A. (2015). *The unique challenges of surveying U.S. Latinos.* Pew Research Center. https://doi.org/https://www.pewresearch.org/methods/2015/11/12/the-unique-challenges-of-surveying-u-s-latinos/

Čehovin, G., Bosnjak, M., & Lozar Manfreda, K. (2023). Item nonresponse in web versus other survey modes: A systematic review and meta-analysis. *Social Science Computer Review, 41*(3), 926–945. https://doi.org/10.1177/08944393211056229

Chen, C., Lee, S. Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science, 6*(3), 170–175. https://doi.org/10.1111/j.1467-9280.1995.tb00327.x

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics, 39*(4), 860–864.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Colorado Department of Education. (n.d.). *Grade to age chart*. Retrieved April 2, 2024, from https://doi.org/https://www.cde.state.co.us/cdereval/grade-age-chart

Consumer Financial Protection Bureau (CFPB). (2018). *Measuring financial skill*. https://doi.org/https://www.consumerfinance.gov/data-research/research-reports/measuring-financial-skill/

IBM Corp. (2022). *IBM SPSS Statistics for Windows (Version 29.0)* [Computer software].

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6*(4), 475–494. https://doi.org/10.1177/001316444600600405

Dholakia, U., Tam, L., Yoon, S., & Wong, N. (2016). The ant and the grasshopper: Understanding personal saving orientation of consumers. *Journal of Consumer Research, 43*(1), 134–155. https://doi.org/10.1093/jcr/ucw004

Dietrich, S., & Hernandez, E. (2022b). *Nearly 68 million people spoke a language other than English at home in 2019*. https://doi.org/https://www.census.gov/library/stories/2022/12/languages-we-speak-in-united-states.html

Dietrich, S., & Hernandez, E. (2022a). *Language use in the United States: 2019*: *American Community Survey Reports.* https://doi.org/https://www.census.gov/content/dam/Census/library/publications/2022/acs/acs-50.pdf

Donald, J. H. (2024). *6.1: Multiple comparisons*. Retrieved January 8, 2024 from https://doi.org/https://stats.libretexts.org/Bookshelves/Applied_Statistics/Biological_Statistics_(McDonald)/06%3A_Multiple_Tests/6.01%3A_Multiple_Comparisons

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Taylor & Francis.

Drennan, J. (2003). Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing, 42*(1), 57–63. https://doi.org/10.1046/j.1365-2648.2003.02579.x

Duke University Health System. (2021). *Research involving non-English speaking research participants*. https://doi.org/https://irb.duhs.duke.edu/sites/default/files/2022-04/non_english-speaking_subjects_translation_policy_7-7-2021.pdf

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221–233. https://doi.org/10.1037/h0057532

Flesch, R. (1979). *How to write plain English*. Harper & Row.

Harzing, A. W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management, 6*(2), 243–266. https://doi.org/10.1177/1470595806066332

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge.

Jacobsen, J. (2022). If they don't understand the question, they don't answer. Language mismatch in face-to-face interviews. *Journal of Official Statistics, 38*(2), 453–484. https://doi.org/10.2478/jos-2022-0022

Johnson, T. P., Pennell, B.-E, Stoop, I. A. L., & Dorer. B. (2018). The promise and challenge of 3MC research. In T. P. Johnson, B.-E. Pennell, I. I. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methodology: Multinational, multiregional, and multicultural contexts (3MC)* (pp. 52-60). John Wiley & Song

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel*. Institute for Simulation and Training. https://doi.org/https://stars.library.ucf.edu/istlibrary/56

Kleiner, B., Lipps, O., & Ferrez, E. (2015). Language ability and motivation among foreigners in survey responding. *Journal of Survey Statistics and Methodology, 3*(3), 339–360. https://doi.org/10.1093/jssam/smv015

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.

Ksendzova, M., Donnelly, G. E., & Howell, R. T. (2017). A brief money management scale and its associations with personality, financial health, and hypothetical debt repayment. *Journal of Financial Counseling and Planning, 28*(1), 62–75. https://doi.org/10.1891/1052-3073.28.1.62

LeBaron-Black, A. B., Curran, M. A., Hill, E. J., Freeh, M. E., Toomey, R. B., & Speirs, K. E. (2022). Parent financial socialization scale: Development and preliminary validation. *Journal of Family Psychology, 36*(6), 943–953. https://doi.org/10.1037/fam0000927

Lynch, J. G., Jr., Netemeyer, R. G., Spiller, S. A., & Zammit, A. (2010). A generalizable scale of propensity to plan: The long and the short of planning for time and for money. *Journal of Consumer Research, 37*(1), 108–128. https://doi.org/10.1086/649907

Microsoft. (n.d.). *Get your document's readability and level statistics*. Retrieved March 31, 2024 from https://doi.org/https://support.microsoft.com/en-us/office/get-your-document-s-readability-and-level-statistics-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2

Netemeyer, R. G., Warmath, D., Fernandes, D., & Lynch, J. G., Jr. (2018). How am I ng? Perceived financial well-being, its potential antecedents, and its relation to overall well-being. *Journal of Consumer Research, 45*(1), 68–89. https://doi.org/10.1093/jcr/ucx109

Nguyen, H. T. (2019). Development and validation of a women's financial self-efficacy scale. *Journal of Financial Counseling and Planning, 30*(1), 142–154. https://doi.org/10.1891/1052-3073.30.1.142

OECD. (n.d.). *United States: Adult skills (Survey of Adult Skills, PIAAC, 2017)*. Retrieved March 27, 2024, from https://doi.org/http://gpseducation.oecd.org

Paasche-Orlow, M. K., Taylor, H. A., & Brancati, F. L. (2003). Readability standards for informed-consent forms as compared with actual readability. *New England Journal of Medicine, 348*(8), 721–726. https://doi.org/10.1056/NEJMsa021212

Pettit, A. (2015). *How non-native English speakers respond to English surveys* [Powerpoint Slides]. Peanut Labs. https://doi.org/https://www.slideshare.net/anniepettit/how-nonnative-english-speakers-respond-to-english-surveys

Peytcheva, E. (2018). Can the language of survey administration influence respondents' answers? In T. P. Johnson, B. Pennell, I. Stoop, & B. Dorer. (Eds.), *Advances in Comparative Survey Methods* (pp. 325–340). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118884997.ch15

Powers, D. E. (2010). *The case for a comprehensive, four-skills assessment of English language proficiency*. Educational Testing Service. https://citeseerx.ist.psu.edu/document?repid=rep1type=pdfdoi=a875ae03ee025060d0b8996f11ae9d2c42131593

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2019). *HLM 8: Hierarchical linear and nonlinear modeling*. https://ssicentral.com/wp-content/uploads/2020/07/HLM_Guide.pdf

Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide* (4th ed.). John Wiley & Sons.

Reid, J. (1990). The dirty laundry of ESL survey research. *TESOL Quarterly, 24*(2), 323–338.

Richins, M. L. (2004). The material values scale: Measurement properties and development of a short form. *Journal of Consumer Research, 31*(1), 209–219. https://doi.org/10.1086/383436

Scientific Software International, Inc. (2021). HLM *(Version 8)* [Computer software].

Si, S. X., & Cullen, J. B. (1998). Response categories and potential cultural bias: Effects of an explicit middle point in cross-cultural surveys. *The International Journal of Organizational Analysis, 6*(3), 218–230. https://doi.org/10.1108/eb028885

Smith, T. W. (2018). Improving multinational, multiregional, and multicultural (3MC) comparability using the Total Survey Error (TSE) paradigm. In T. P. Johnson, B.-E. Pennell, I. I. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methodology: Multinational, multiregional, and multicultural contexts* (3MC) (pp. 61-91). John Wiley & Song

Sotiropoulos, V., & d'Astous, A. (2013). Attitudinal, self-efficacy, and social norms determinants of young consumers' propensity to overspend on credit cards. *Journal of Consumer Policy, 36*, 179–196.

Spinella, M., Yang, B., & Lester, D. (2007). Development of the executive personal finance scale. *International Journal of Neuroscience, 117*(3), 301–313. https://doi.org/10.1080/00207450500534043

StataCorp LLC. (2023). *Stata (Version 18)* [Computer software].

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Trzcinski, E. (1981). *A review of the problem of limited English competence of survey respondents* (Report No. HEW-100–79–0127). University of Michigan. https://cdn.isr.umich.edu/pubFiles/historicPublications/Limitedenglishcompetence_report_4481.PDF

UCLA. (2023). *Guidance and procedure: Research involving non-English speaking research participants*.https://ora.research.ucla.edu/OHRPP/Documents/Policy/9/NonEnglish_Participants.pdf

Warmath, D. (2022). *Financial decision making*. Report to the Australian Securities and Investment Commission.

Wenz, A., Al Baghal, T., & Gaia, A. (2021). Language proficiency among respondents: Implications for data quality in a longitudinal face-to-face survey. *Journal of Survey Statistics and Methodology, 9*(1), 73–93. https://doi.org/10.1093/jssam/smz045

WIDA. (2020). *WIDA English language development standards framework, 2020 edition: Kindergarten–grade 12*. Board of Regents of the University of Wisconsin System. https://wida.wisc.edu/sites/default/files/resource/WIDA-ELD-Standards-Framework-2020.pdf

Yale University Institutional Review Boards. (2015). *Guidance on inclusion of non-English speaking participants in human*

research. https://your.yale.edu/sites/default/files/200gd2nonenglishspeakers_5-6-15.pdf