Journal of Financial Literacy and Wellbeing (2025), 1–20
doi:10.1017/flw.2024.11

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Consumer financial well-being: Does scale choice alter the measure?

Patrick Heck, Caroline Ratcliffe and Elle Tibbitts

Office of Research, Consumer Financial Protection Bureau, Washington, DC, USA
Corresponding author: Patrick Heck; Email: patrick.heck@cfpb.gov

(Received 4 January 2024; revised 12 June 2024; accepted 14 September 2024)

Abstract

A crucial first step in helping consumers improve their financial lives is understanding their financial circumstances and well-being. The Financial Well-Being (FWB) scale measures a consumer's subjective well-being related to aspects of their financial circumstances. It is available in standard-length (10-item) and abbreviated (5-item) versions, but no research has compared how completing either version may alter consumers' responses. Notably, the 5-item scale includes a higher share of reverse-coded (i.e., negatively framed) items. We hypothesize that the difference in item framing between scale versions influences participants' feelings about their financial situation, predicting that completing the 5-item FWB scale will result in more negative responses compared to completing the 10-item FWB scale. To test this hypothesis, we implement a randomized survey experiment using the Understanding America Study. In our experiment with nearly 6,000 participants, we find that completing the 5-item versus the 10-item FWB scale reduces FWB scores (average decline in the 5-item FWB score of 0.9 points, 95% CI [−1.552, −0.249]), and increases the share with a "low" 5-item FWB score by 5.0 percentage points, 95% CI [0.028, 0.071]), responses to individual scale items, and self-rated FWB. This pattern is strongest among lower-income respondents (average decline in FWB score of 2.3 points, 95% CI [−3.385, −1.171] and increases the share with a "low" 5-item FWB score by 8.1 percentage points, 95% CI [0.041, 0.121]). These findings highlight that FWB scale choice can have unexpected consequences. We discuss the implications for research on FWB and on the measurement of well-being more broadly.

Keywords: financial well-being; financial health; measurement; consumer decision-making

JEL Codes: I31; G41

1. Introduction

A crucial first step in helping consumers improve their financial lives is understanding their financial circumstances and well-being. To this end, measures of financial well-being (FWB) are used by financial educators and practitioners as an on-the-ground tool to provide awareness, education, and useful feedback to individual consumers in the United States (Kim, 2004; Postmus and Hetling, 2015; Way and Wong, 2010). Furthermore, evaluating the effect of financial education or other interventions on consumers' well-being entails first reliably measuring their financial situations.

A U.S. federal agency, the Consumer Financial Protection Bureau (CFPB), describes FWB as "the ultimate measure of success for financial literacy efforts" (CFPB, 2015a) and "the end goal of financial education" (CFPB, 2015b). With these goals in mind, the CFPB created and validated a widely used scale of consumer FWB that is available in a standard-length (10-item) and abbreviated (5-item) version (CFPB, 2015a). In this study, we implement a randomized experiment using the Understanding America Study – a large, nationally representative survey of U.S. households – to test if and how measuring FWB may unexpectedly alter the measure itself. Specifically, we study whether the decision to use the abbreviated FWB scale (relative to the standard-length FWB scale) alters consumers' (1) FWB scale scores, (2) scores on the individual items (i.e., questions) contained in the scale, and (3) self-rated FWB captured using a single item measure.

Many questions and scales have been used to measure consumers' financial circumstances. These measures, which can be either subjective or objective in nature, are thought to capture distinct (though conceptually related) constructs including FWB, capacity to absorb financial shocks, financial stability, financial health, and financial resilience, among others.[1] Measures of FWB are widely employed in both research and policy arenas (Tescher and Silberman, 2021), and the subjective component of well-being has been identified as critical to the broader definition and study of consumer wealth (Tully and Sharma, 2022).

We focus our research on the CFPB's FWB scale. The FWB scale, available in a 10-item standard-length version and a 5-item abbreviated version, measures a consumer's subjective well-being related to financial aspects of their lives. It is designed to assess whether "a person can fully meet current and ongoing financial obligations, can feel secure in their financial future, and is able to make choices that allow them to enjoy life" (CFPB, 2015b). Accordingly, the FWB scale is strongly associated with broad measures of economic security including household income, liquid savings, and home ownership (CFPB, 2017a; Sun et al., 2018), and tracks changes across the life course (Michael and Urban, 2021).[2] The FWB scale has been incorporated in research studies as a correlate or predictor of financial outcomes (Cohen, Hoagland, and Wiedrich, 2017; Dedmon, Lakshmanan, and Cabusora, 2019; and Powers and Edwards, 2020), in large-scale financial interventions and educational programs as a measure of effectiveness (Burke, Collins, and Urban, 2023; Hamilton, Roll, Despard, et al., 2022; Jaroszewicz, Jachimowicz, Hauser, et al., 2022), and in established and ongoing national surveys as a nationwide measure of financial wellness (e.g., the CFPB's Making Ends Meet Survey, CFPB, 2022). Over 70 unique research groups, firms, nonprofits, government agencies, and financial educators and practitioners use or recommend using the CFPB's FWB scale to measure FWB.[3]

Despite widespread use of the CFPB's FWB scale, little is known about if or how completing the scale or the specific items it contains may influence consumers' financial attitudes or beliefs. Research in psychology has broadly demonstrated how merely completing scale measures can alter people's attitudes or beliefs about what is being measured (Knowles, 1988). More narrowly, experimental research has found that exposure to stresses related to two components of FWB (e.g., "current money management stress"

[1] There is considerable debate over the optimal measurement of financial circumstances. Tescher and Silberman (2021) review many of the measures—objective and subjective—that relate to a consumer's overall financial health and well-being. Some of these include reports of whether one's financial situation has recently improved; whether consumers report having difficulty paying bills; and whether consumers fall into defined categories like "financially vulnerable" or "living on the edge." Our research focuses on financial well-being specifically and we do not take a position on the merits of this measure relative to others.

[2] Other measures conceptually related to financial well-being have also been shown to correlate with broad measures of wealth, well-being, and life satisfaction (for example, "financial satisfaction," see Ng and Diener, 2014).

[3] Data obtained from unpublished CFPB internal record keeping.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

and "expected future financial security") reduces participants' overall perception of their FWB (Netemeyer, Warmath, Fernandes, and Lynch, 2018). These insights motivate the logic underlying our research question: if measurement choices like which scale to use influence consumers' beliefs, attitudes, or mindset, then this may in turn alter consumers' responses on measures of FWB in unexpected or undesirable ways. Specifically, we ask: what is the effect of completing the 5-item FWB scale, versus the 10-item FWB scale, on measures related to consumers' FWB?

To answer this question, we compare the effect of completing the 10-item standard-length FWB scale, relative to the 5-item abbreviated FWB scale, on a series of outcomes related to FWB. First, we examine participants' computed "5-item FWB scores." This score provides a directly comparable measure because it is computed the same way for all study participants. For study participants who completed the 10-item scale, the 5-item FWB score is calculated using only the five questions included in the 5-item scale. For study participants who completed the 5-item scale, the same five questions are used to calculate the 5-item FWB score. We also compare the effect of completing either scale version on participants' responses to the individual scale items and self-rated level of FWB (measured separately).

The two FWB scale versions correlate highly with each other (in one report, $r = .94$, CFPB, 2017b) and have been previously validated as effective measures of the same core construct of FWB (CFPB, 2015a). However, two measures that are highly correlated can still produce scale scores, or even consumer experiences when completing the measure, that differ on average. Our experiment is the first to ask whether the choice to use the 5-item versus the 10-item FWB scale influences consumers' self-assessment of their financial circumstances.

Why might two versions of the same scale be expected to differentially influence consumers' assessments of their financial situation? There are some notable differences between these two scales beyond their length alone. One potentially important difference between the 10-item and 5-item scales is how many reverse-coded (i.e., positively versus negatively framed) items each version has. Although both scale versions contain items that vary in orientation, the 5-item scale contains proportionally more negatively framed questions (e.g., "I am concerned that the money I have or will save won't last") than the 10-item scale.

The CFPB's scale development technical report (CFPB, 2017b) explains that both framings were included during the scale validation process to capture positive and negative experiences, which are important components of well-being that should be measured using items oriented in each direction (National Research Council, 2013).[4] However, differences in exposure to question framing may influence consumers' experience and, therefore, their responses. To illustrate with an example, the first item in the 5-item scale is negatively framed: "Because of my money situation, I feel like I will never have the things I want in life." On the 10-item scale, however, this item is preceded by two positively framed items. If the immediate exposure to this negative item (i.e., when completing the 5-item scale) discourages consumers, or if exposure to the positive items preceding it (i.e., when completing the 10-item scale) shields them from this negativity, then we would observe different responses to this item based on which scale was administered.

Although our research is the first to examine question framing in the specific topic of FWB, there is a rich literature exploring best practices and quality improvement in

[4] The scale development technical report (CFPB, 2017b) notes that early work examined item framing in the initial pool of over 40 candidate items for the FWB scale, but our review of the scale development process found no details on how question framing influenced responses in the final scale versions or why the two scale versions differ in their relative proportions of negatively-framed items.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

questionnaire design (see Krosnick and Presser, 2018, for an overview and Weijters and Baumgartner, 2012, for specific recommendations related to negated and reverse-coded items). Incorporating questions with positive and negative framings is a common practice in scale measurement in part because it reduces acquiescence bias in survey responses (Likert, 1932), although there are tradeoffs involved in including both negative and positive items (see Chyung, Barkin, and Shamsy, 2018; Herche and Engelland, 1996). Our research is scoped to the measurement of FWB, but it contributes to the topics of scale construction and subjective well-being measurement more broadly by directly comparing two established scales that include a comparatively negative (or positive) collection of items.

Another notable difference between the 10-item and 5-item FWB scale versions is their reliability. While the two scales produce scores that are highly correlated, they differ in their level of internal consistency, or reliability. The initial FWB scale development report states that "unless time and effort constraints are a significant concern, the standard 10-item version is preferable because it produces scores with higher reliability and increased precision" (CFPB 2017b, p. 20).[5] This tradeoff between greater resource commitment and reliability is an important consideration when choosing between two scales, in part because a more reliable scale is expected to result in less measurement error. Our research explores differences between these scale versions beyond reliability alone.

In this research, we treat the 10-item scale as the "default" FWB measure. This is because the 10-item FWB scale was released as the CFPB's standard-length scale, due in part to its greater reliability than the 5-item abbreviated version. While we use this framing throughout the paper, we do not know consumers' "true" level of FWB or which measure best captures it, nor is it possible for any subjective survey measure to ascertain this ground truth. Rather, treating the 10-item scale as the default reflects the pragmatism of the status quo: the 10-item scale was established by the CFPB as the standard approach for measuring FWB, while the 5-item scale was included as a shorter but limited (i.e., less reliable) alternative. Thus, in our discussion of the findings we frame completing the 5-item FWB scale as resulting in *lower* FWB than the 10-item scale (as opposed to, for example, the 10-item FWB scale resulting in *higher* reported FWB than the 5-item scale).

Previewing our results, our research finds that the seemingly simple choice of which FWB scale to use has consequences that researchers and practitioners may not expect or desire. Specifically, we find that completing the 5-item (versus the 10-item) FWB scale results in lower reported FWB, as measured by the directly comparable 5-item FWB score. Among those assigned to complete the 5-item versus 10-item FWB scale, we find that the average 5-item FWB score is lower by 0.90 points, 95% CI [–1.552, –0.249], and the share of participants with a "low" 5-item FWB score is higher by 5.0 percentage points, 95% CI [0.028, 0.071]. We also observe similar declines in responses to the individual scale items and a self-rated measure of FWB. Additionally, we find that the negative effect of completing the 5-item (versus the 10-item) FWB scale is larger among those in lower-income households (relative to higher-income households). Among lower-income participants assigned to the 5-item (versus 10-item) FWB scale, the average 5-item FWB score is lower by 2.3 points, 95% CI [–3.385, –1.171], and the share with a "low" 5-item FWB score is higher by 8.1 percentage points, 95% CI [0.041, 0.121].

We proceed by developing our specific hypotheses and experimental design (Section 2) and providing more information on our study participants, data, measures, and analytic approach (Section 3). We then present our results (Section 4). Finally, we conclude with a

[5] Marginal reliability, which measures the strength of internal consistency of a set of metrics derived using item-response theory, is approximately 0.90 and 0.83 for the 10-item and 5-item versions, respectively. These reliabilities are generally high and indicate that only a small proportion of variation in scale scores is due to measurement error.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

Table 1. Questions and question presentation order in the CFPB's 10-item and 5-item Financial Well-Being (FWB) scales

	10-Item FWB scale (40% positive framing)		5-Item FWB scale (20% positive framing)
1	I could handle a major unexpected expense (+)		
2	I am securing my financial future (+)		
3	Because of my money situation, I feel like I will never have the things I want in life (–)	1	Because of my money situation, I feel like I will never have the things I want in life (–)
4	I can enjoy life because of the way I'm managing my money (+)		
5	I am just getting by financially (–)	2	I am just getting by financially (–)
6	I am concerned that the money I have or will save won't last (–)	3	I am concerned that the money I have or will save won't last (–)
7	Giving a gift for a wedding, birthday, or other occasion would put a strain on my finances for the month (–)		
8	I have money left over at the end of the month (+)	4	I have money left over at the end of the month (+)
9	I am behind with my finances (–)		
10	My finances control my life (–)	5	My finances control my life (–)

Notes: "+" and "–" symbols indicate positively and negatively framed questions, respectively. In the 10-item scale, questions 1–6 are framed by asking people "How well does this statement describe you or your situation" and provide five response options – Completely, Very well, Somewhat, Very little, and Not at all – while questions 7–10 ask "How often does this statement apply to you?" with the five response options being Always, Often, Sometimes, Rarely, and Never. The questions and responses for the 5-item scale follow similarly. See "Appendix: Survey budget tools and questions" for more details. CFPB materials instruct users of the FWB scale to "keep the scale questions in the same order and with the same wording as developed by the Bureau" (CFPB, 2019 p. 23).

discussion of the results including implications for practitioners and researchers (Section 5).

2. Hypotheses and experimental design

Our experiment tests whether the FWB scale version consumers complete – the 10-item standard-length version or the 5-item abbreviated version – influences how they respond to FWB-related measures. If consumers (or specific consumer groups or subpopulations) are systematically influenced by the choice to use one scale version over the other, then this poses an obstacle for the reliable and valid measurement of FWB.

Furthermore, we make a directional prediction for this effect. Examination of both scales reveals that the 5-item scale contains proportionally more negatively framed (or reverse-coded) items (e.g., "I am concerned that the money I have or will save won't last"), while the 10-item contains proportionally more positively framed items (e.g., "I can enjoy life because of the way I'm managing my money"). Table 1 presents the questions and question presentation order in the 10-item and 5-item FWB scales. If respondents are influenced by question framing such that negatively framed items are perceived as harsher and prompt more negative impressions of respondents' financial circumstances (than positively framed items), then the more negatively framed scale should result in greater negativity in participants' responses. We therefore hypothesize that exposure to a comparatively negative scale makes participants feel worse about their financial situation and predict the following:

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

Table 2. Two-by-two between-subjects experimental design

FWB Scale Condition (primary treatment)	Budget Order Condition (secondary treatment)	
	Budget Before Scale	*Budget After Scale*
10-item standard-length FWB scale	Participant completes the monthly budget then the 10-item FWB scale	Participant completes the 10-item FWB scale then the monthly budget
5-item abbreviated FWB scale	Participant completes the monthly budget then the 5-item FWB scale	Participant completes the 5-item FWB scale then the monthly budget

Notes: The FWB scale condition is our primary treatment and the budget order condition is our secondary treatment. We find no effects of the budget order condition or any interaction with the primary treatment.

Hypothesis 1: *completing the 5-item FWB scale will reduce consumers' financial well-being compared to completing the 10-item FWB scale. This may include their financial well-being score, responses to items in the FWB scale they complete, and a self-rating measure of FWB.*

Additionally, consumers who are struggling financially, or have access to fewer financial resources, may be particularly susceptible to the effect of completing the more negative 5-item FWB scale. We therefore further hypothesize the following:

Hypothesis 2: *any negative effect of completing the 5-item FWB scale will be greater among those with lower versus higher household income.*

We test these hypotheses using an online survey experiment. Study participants were randomly assigned to complete either the 10-item standard-length or the 5-item abbreviated CFPB FWB scale (i.e., the *FWB scale* treatment). Participants were also randomly assigned to a second manipulation: whether they completed the 5- or 10-item FWB scale before or after using a tool to estimate their monthly budget (i.e., the *budget-before-scale* or *budget-after-scale* treatment).[6] Our analyses incorporate participants' assignment to the budget-before-scale versus budget-after-scale treatment, and we find no significant effects of this secondary manipulation on any of our outcomes or interactions with our primary treatment. Table 2 presents the full two-by-two between-subjects experimental design.

After completing the 5- or 10-item FWB scale and the budget task, all participants completed the same series of questions related to their financial experiences, beliefs, and behaviors (see the appendix for all items). Key for this study and as one component of FWB, participants were shown a definition of FWB and asked to rate their own level of FWB on a seven-point scale from "very low" to "very high." This measure was separated from the FWB scale items by a complete survey page and several survey items. Including this measure as a primary outcome bolsters our conclusions by demonstrating that any effects of the experimental manipulations continue beyond completion of the FWB scale itself.

3. Study participants and data

Our experiment was implemented using the Understanding America Study (UAS), a nationally representative internet panel of U.S. households that includes nearly 13,000

[6] The budget task instructed participants to estimate dollar values for several specific categories of income (e.g., jobs, government benefits) and spending (e.g., housing, groceries, transportation) in a typical month (see appendix). This budget-timing manipulation was included to answer an alternate research question focused on whether completing a financial task—specifically, building a monthly budget—affects consumers' financial well-being.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

respondents and is maintained by the Center for Economic and Social Research at the University of Southern California. The UAS recruits panel members using address-based sampling, and unlike many internet-based surveys, does not require that panel members have preexisting online access. If a panel member does not have online access when they join the UAS panel, they are provided with a tablet and/or broadband access. This improves the representativeness of the UAS sample. The UAS data include a host of demographic and economic characteristics that are collected for all panel members when participants join the panel and are updated approximately every three months. These characteristics include age, gender, educational attainment, race and ethnicity, household income, and work status, among others. UAS typically contacts panel members once a month to complete a survey.

A random sample of 9,348 UAS panel members was invited to participate in our survey experiment between mid-October and early November 2022. Overall, 6,318 panel members took the survey, resulting in a response rate of 68 percent. We restrict our analysis sample to 5,768 survey respondents who (1) spent more than 3.4 minutes completing the survey, (2) completed both the FWB scale and budget sections of the survey, and (3) have demographic and economic (e.g., household income) information available. Five percent of survey respondents (315 people) were removed because they completed the survey in 3.4 minutes or less. The median amount of time spent on the survey was 7.3 minutes. We exclude an additional 214 people who did not complete the FWB scale or the monthly budget exercise and 21 people because their demographic characteristics were missing from UAS records.

Comparison of the initial random sample to our final analysis sample of 5,768 participants shows that our analysis sample is older on average (53 versus 51 years old) and less racially and ethnically diverse (share Black or Hispanic 21% versus 24%) (see Appendix Tables A1 and A2). While there appears to be some minor selective survey non-response, it does not vary by our primary subgroup of interest (household income) so it does not threaten the generalizability of our results.

Participants were randomly assigned to one of four treatment conditions. We test for statistically significant differences in mean characteristics across participants assigned to the different conditions. Table 3 shows that those assigned to the *5-item scale* versus *10-item scale* condition are more likely to be retired (27.5% versus 25.5%, respectively), but do not significantly differ on any other characteristics.[7] Our regression analyses, discussed in the next section, control for this difference and other participant demographic and economic characteristics.

4. Outcome measures and analytic approach

4.1. Measures

We analyze four interrelated primary outcome measures related to FWB (see Table 4). First, we use the 5-item FWB score, which we calculate using the item-response theory (IRT) weights provided in the FWB scale development technical report (CFPB, 2017b, p. 27). We use the 5-item FWB score as a primary outcome measure because it can be computed regardless of which scale participants completed; this measure ensures that we are examining the same outcome for all participants (i.e., it is an apples-to-apples comparison that avoids introducing any differences in scoring between, for example, the 5-item and 10-item scales). We find similar patterns when instead using participants' "actual" FWB

[7] We also test for and find no meaningful differences in demographic characteristics between the two budget task order conditions.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

Table 3. Descriptive statistics for study participants' characteristics

	Participants assigned to 5-item scale	Participants assigned to 10-item scale	All study participants
Age (%)			
18–29	7.0	7.7	7.4
30–39	14.8	16.8	15.8
40–49	19.5	18.6	19.1
50–61	24.1	23.1	23.6
62–69	16.5	16.1	16.3
70 or older	18.1	17.7	17.9
Male (%)	41.4	40.8	41.1
Married (%)	56.4	55.7	56.1
Number of children in the household (mean)	0.4	0.4	0.4
	(0.9)	(0.9)	(0.9)
Retired (%)	27.5	25.5*	26.5
Currently working (%)	57.2	57.8	57.5
Race and Ethnicity (%)			
White (non-Hispanic)	68.4	69.0	68.7
Black (includes Hispanic)	8.0	8.4	8.2
Hispanic	12.5	12.3	12.4
Other non-White	11.1	10.3	10.6
Educational attainment (%)			
High school degree or less	41.5	40.9	41.2
Associate's degree	13.3	14.1	13.7
Bachelor's degree or more	45.2	45.0	45.1
First generation immigrant (%)	10.2	10.5	10.3
Household income below $50,000 (%)	37.0	36.6	36.8
Observations	2,812	2,956	5,768

Notes: This table presents the characteristics of study participants overall and by treatment condition. Differences between the experimental conditions were tested using *t*-tests for continuous measures and chi-squared tests for categorical measures. Standard deviations in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

score (the 10-item score when completing the 10-item scale and the 5-item score when completing the 5-item scale).[8]

Although researchers may prefer continuous FWB scores, practitioners may prefer FWB information in discrete form (e.g., "low" or "not low") so they can provide concise feedback and guidance to an individual consumer. We therefore also report a binary version of the 5-item FWB score that captures whether a participant's 5-item FWB score is in the low range (i.e., "medium-low," "low," or "very low"), defined as a score below 50 (see

[8] We further discuss this finding in the results section.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

Table 4. Financial well-being outcome measures and descriptions

Outcome measure	Description
Primary outcomes	
5-item FWB score	Continuous 5-item FWB score computed regardless of FWB scale completed; range from 19 to 90 (low scores indicate lower FWB)
Low 5-item FWB score	Indicator for whether *5-item FWB Score* is in the low range (less than 50; "medium-low," "low," or "very low")
Self-rated FWB	Self-rated FWB after reading the definition of FWB; range from 1 to 7 ("very low" to "very high")
Low self-rated FWB	Indicator for whether *self-rated FWB* was low (choosing "medium-low," "low," or "very low")
Secondary Outcomes	
Never have what want in life	"Because of my money situation, I feel like I will never have the things I want in life;" range 1–5 ("completely" to "not at all")
Just getting by financially	"I am just getting by financially;" range 1–5 ("completely" to "not at all")
Concerned money won't last	"I am concerned that the money I have or will save won't last;" range 1–5 ("completely" to "not at all")
Money left over	"I have money left over at the end of the month;" range 1–5 ("never" to "always")
Finances control my life	"My finances control my life;" range 1–5 ("always" to "never")

Notes: This table presents the definitions for the primary and secondary outcome variables used in the estimation. All outcome variables are consistently measured for study participants who completed the 5-item and 10-item FWB scales. For example, we compute the 5-item FWB score for participants who completed the 10-item FWB scale, as well as for those who completed the 5-item FWB scale. For the secondary outcomes, questions 1–3 are framed by asking people "How well does this statement describe you or your situation" and provide five response options – Completely, Very well, Somewhat, Very little, and Not at all – while questions 4–5 ask "How often does this statement apply to you?" with the five response options being Always, Often, Sometimes, Rarely, and Never.

CFPB, 2018 for information developed for practitioners on FWB score ranges and how to interpret them). This binary approach adds insight because consumers who fall into these lower categories have been shown to report greater difficulty making ends meet and likelihood of having experienced a material hardship (CFPB, 2017a).

Third, we analyze participants' self-rated level of FWB, which was collected on a separate page after the FWB scale and several other survey questions had been completed. Participants were shown the definition of FWB and asked to rate their own level of FWB on a scale ranging from 1 (very low) to 7 (very high), with a midpoint of 4 (medium). This measure contributes to our analysis beyond participants' computed 5-item FWB scores because (1) it provides a response to a one-item measure that is distinct from the FWB scale items; and (2) participants answer this question several survey pages after they complete the FWB scale, allowing us to examine whether the effect of completing the 5- versus 10-item scale persisted beyond the measurement of FWB scores. Fourth, we analyze a binary version of this measure that asks whether participants reported having "medium-low," "low," or "very low" FWB according to the provided definition.

We additionally examine five secondary outcomes: the individual scale items in the 5-item FWB scale (see Table 4). Analysis of these secondary outcomes allows us to explore whether any overall effect we might observe is driven by all five items, concentrated within a single item, or something in between. It also provides some insight into how early during scale completion any differences begin to emerge. For example, as shown in Table 1,

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

Table 5. Summary statistics for primary and secondary outcomes

	All Participants		Lower-income participants		Higher-income participants	
	Mean	SD	Mean	SD	Mean	SD
Primary outcomes						
5-item FWB score	56.2	14.6	50.1	14.0	59.7	13.7
Low 5-item FWB score	29.6%		45.5%		20.4%	
Self-rated FWB	3.9	1.7	3.0	1.5	4.4	1.5
Low self-rated FWB	38.3%		61.6%		24.8%	
Secondary outcomes						
Never have what want in life	2.5	1.2	2.1	1.2	2.8	1.0
Just getting by financially	2.4	1.3	1.9	1.2	2.6	1.2
Concerned money won't last	2.1	1.3	1.7	1.3	2.3	1.2
Money left over	2.4	1.3	1.8	1.2	2.8	1.1
Finances control my life	2.3	1.2	2.0	1.3	2.5	1.1

Notes: This table presents the means of the four primary and five secondary outcomes for (1) all study participants (n= 5,768), lower-income participants (n = 2,124), and higher-income participants (n = 3,644). Lower-income and higher-income participants have household incomes of less than $50,000 and at least $50,000, respectively.

the 5-item scale starts with a negatively framed question (i.e., "Because of my money situation, I feel like I will never have the things I want in life"), while the 10-item scale asks two positively framed questions before this negatively framed question appears. Thus, differences between the 5- and 10-item scales may emerge starting as early as the first 5-item scale question.

Descriptive statistics for our primary and secondary outcome measures are shown for all participants and by participants' household income levels in Table 5. The average 5-item FWB score for all participants in our study is 56.2, which is in the upper end of the "medium-high" range of FWB. There is nearly a 10-point difference in the scores of lower-income (50.1) and higher-income (59.7) participants. Additionally, 29.6% of all participants have a 5-item FWB score in the "low" range, with more than twice as many lower-income participants in this range (45.5%) than higher-income participants (20.4%). We find a similar pattern for the self-rated measure of FWB. For example, the average self-rated FWB of 3.9 is near the scale midpoint (3.5), with lower-income participants dropping below the midpoint (3.0) and higher-income participants rising above it (4.4), on average. Descriptive statistics on the individual 5-item FWB scale items show overall averages at or just below the scale midpoint of 2.5, with sizable differences between lower- and higher-income participants.

4.2. Analytic approach

We estimate the effect of the FWB scale condition (5-item versus 10-item scale) on each outcome using Ordinary Least Squares (OLS) regressions. We estimate the following model:

$$Y_i = \alpha + \beta FWB5_i + \gamma_1 BudgetBefore_i + \gamma_2 X_i + \varepsilon_i \quad (1)$$

In this model, Y_i is the outcome for individual i (e.g., self-reported FWB), $FWB5_i$ is an indicator for random assignment to the 5-item (versus 10-item) FWB scale condition,

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

$BudgetBefore_i$ is an indicator for random assignment to completing the budget task before (versus after) the FWB scale, and X_i is a vector that denotes study participants' demographic and economic characteristics including age, gender, marital status, race and ethnicity, work status, and income (see Table 3 for a full list of characteristics). The parameter β is the estimated effect of taking the 5-item (versus the 10-item) FWB scale on the outcome of interest (e.g., self-reported FWB).[9]

We estimate a separate regression model for each of the four primary and five secondary outcomes. We adjust the p-values for multiple outcomes using the Romano and Wolf step-down procedure (Clarke, Romano, and Wolf, 2020; Romano and Wolf 2005). We make these adjustments treating our four primary outcomes as a set of outcomes and apply adjustments separately for our five secondary outcomes.

We hypothesize that the effect of scale assignment will differ based on participants' household income, so we estimate additional models to obtain estimated effects for lower-income and higher-income households. In our primary specifications, we consider a participant's lower income if their annual household income is less than $50,000 and higher income if their annual household income is at least $50,000. This dividing line of $50,000 is roughly twice the poverty threshold for the average participant in our study. We also conduct robustness checks on other income cut-offs – $35,000, $60,000, and $75,000 – and the findings are qualitatively similar in terms of magnitude and level of statistical significance.[10]

The model we estimate is:

$$Y_i = \alpha + \beta_1(FWB5_i * LowIncome) + \beta_2\big(FWB5_i * HighIncome\big) + \gamma_1 BudgetBefore_i + \gamma_2 X_i + \varepsilon_i \quad (2)$$

where β_1 is the estimated effect of taking the 5-item (versus 10-item) FWB scale on the outcome for lower-income people, while β_2 is the estimated effect for higher-income people.[11]

5. Results

We find that completing the 5-item FWB scale causes reductions in our primary and secondary outcomes relative to completing the 10-item FWB scale, supporting our first hypothesis. These effects are more pronounced among lower-income consumers, supporting our second hypothesis.

5.1. Primary outcomes

On average, participants randomly assigned to take the 5-item versus 10-item FWB scale have lower reported FWB as measured by our four primary outcomes.

[9] This regression approach improves the precision of our estimated effects by controlling for participants' demographic and economic characteristics.

[10] We examine these various income cut-offs because there is no one recognized "right" value. However, various researchers and practitioners define low-income as below 200 percent of the federal poverty level (e.g., Kaiser Family Foundation (https://www.kff.org/other/state-indicator/health-insurance-coverage-of-low-income-children-0-18-under-200-fpl), Annie E. Casey Foundation (https://datacenter.aecf.org/data/tables/8438-low-income-children-under-age-19-without-health-insurance), Zedlewski, Chaudry, and Simms, 2008.

[11] We also estimated models to test if the estimated effect of taking the 5-item (versus 10-item) FWB scale differs for younger (ages 18–62) and older (ages 62 and older) participants; we find no reliable evidence of different effects by age.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

Table 6. Effect of taking the abbreviated 5-item versus standard-length 10-item FWB scale on *primary FWB outcomes*, all participants and by income

	5-item FWB score	Low 5-item FWB score	Self-reported FWB	Low self-reported FWB
Estimated effects, full sample				
5-item FWB scale	–0.900**	0.050***	–0.068*	0.024*
	(0.332)	(0.011)	(0.038)	(0.011)
	$p = 0.013$	$p = 0.001$	$p = 0.079$	$p = 0.074$
Estimated effects, by income				
5-item FWB scale*low-income	–2.278***	0.081***	–0.122*	0.054**
	(0.565)	(0.021)	(0.064)	(0.020)
	$p = 0.001$	$p = 0.001$	$p = 0.052$	$p = 0.015$
5-item FWB scale*high-income	–0.096	0.031*	–0.037	0.006
	(0.410)	(0.013)	(0.046)	(0.014)
	$p = 0.848$	$p = 0.065$	$p = 0.710$	$p = 0.848$
Mean of dependent variable for those assigned to 10-item scale				
All	56.53	0.274	3.93	0.373
Low-income	51.08	0.417	3.06	0.594
High-income	59.68	0.191	4.44	0.246

Notes: This table presents OLS regression results for the impact of taking the 5-item (versus 10-item) FWB scale on our four primary FWB outcomes (5-item FWB score, low 5-item FWB score, self-reported FWB, low self-reported FWB). The impact estimates for all study participants (n = 5,768) are presented in the top panel. The impact estimates for lower-income (n = 2,124) and higher-income (n = 3,644) study participants are presented in the middle panel (see equation 2 in the "Analytic Approach" section for details). Lower income and higher income are defined as having household income of less than $50,000 and at least $50,000, respectively. All models also control for budget order, age, gender, marital status, number of children, work status (currently working, retired), race and ethnicity, educational attainment, income, and immigrant status (see Table 3 for details). Robust standard errors are in parentheses. The *p*-values are adjusted for multiple outcomes using the Romano and Wolf step-down procedure.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Focusing first on the 5-item FWB score, the average 5-item FWB score is lower by 0.90 points among those assigned to complete the 5-item versus 10-item FWB scale (see Table 6, top panel). This corresponds to a 1.6% decline relative to the 10-item condition (mean is 56.53; Table 6, bottom panel).[12] This 0.90 point difference is similar in magnitude to the average increase in FWB scores early in the pandemic. The average FWB score was found to increase by 1.04 points between June 2019 and June 2020 (Fulford, Rush, and Wilson, 2021) – from roughly nine months before to three months after the onset of the pandemic in March 2020 (and two months after the distribution of pandemic-induced Economic Impact Payments began in April 2020). Additionally, calculations based on our data show that our 0.90 point difference is not far from the roughly 1.3-point difference in the average FWB scores of (1) people with annual household incomes of $40,000 to $50,000 versus those with incomes of $50,000 to $60,000 and (2) people in their 20s versus those in their 30s. These comparisons provide a reference for how to interpret an average FWB difference of 0.90

[12] See the bottom panel of Table 6 for the mean of the dependent variable for those assigned to the 10-item scale condition, overall, and by participants' level of income. Also, see Appendix Table A3 for the full regression results for the primary outcomes.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

points between conditions in our study, which is approaching the difference in FWB associated with a $10,000 increase in income or an additional decade of age.

The share of participants with a low 5-item FWB score is higher by 5.0 percentage points among those assigned to the 5-item versus 10-item FWB scale condition, representing a substantial 18.1% increase relative to the 10-item condition (share is 0.274, or 27.4%).[13] These two FWB measures are based on the same underlying information (5-item FWB score) and yield converging results, but each provides a distinctly meaningful and interpretable outcome.[14]

Additionally, these results are similar or even larger when instead using participants' "actual" FWB scores as outcomes, that is, their 5-item score when completing the 5-item scale and 10-item score when completing the 10-item scale (and the shares of participants with low FWB based on these scores). This analysis (reported in Appendix Table A5) shows that the 5-item scale produces systematically lower FWB scores than the 10-item FWB scale by 1.97 points. We also find that the share of participants classified as having a "low" FWB score (calculated using the "actual" FWB score) is 5.4 percentage points higher when the 5-item vs. 10-item FWB scale is administered.

We find a similar pattern for participants' self-rated FWB, which is measured using a single question that was completed several survey pages after the FWB scale (Table 6). On average, self-reported FWB is lower by 0.068 points among participants who completed the 5-item versus 10-item FWB scale, which corresponds to a 1.7% decline relative to the 10-item condition (3.93). Self-rated FWB is measured in a narrower range (from 1 to 7, or "very low" to "very high") than the 5-item FWB score (from 19 to 90), so while the estimated effect size for self-rated FWB (0.068 scale points) is smaller than the estimated effect size for the 5-item FWB score (0.900 scale points), both effects correspond to a similar decline (1.6%–1.7%) relative to the 10-item condition. Aligning with these results, the share of participants with low self-reported FWB is higher by 2.4 percentage points for those who completed the 5-item versus the 10-item FWB scale condition, corresponding to a 6.4% increase relative to the 10-item condition (37.3%). This increase is smaller than, yet consistent, with our finding for the share with a low 5-item FWB score.

5.1.1. Primary outcomes by income

Consistent with our second hypothesis, we find evidence that the negative effect of completing the 5-item versus the 10-item FWB scale is larger among participants in lower-income (annual income below $50,000) versus higher-income (annual income of at least $50,000) households. In general, the treatment effects for lower-income participants are both larger in magnitude and have a higher level of statistical significance than the treatment effects for higher-income participants (Table 6, middle panel).[15]

Among lower-income participants, the 5-item FWB score is lower by 2.278 points for those assigned to the 5-item scale condition than in the 10-item scale condition, corresponding to a decline of 4.5% relative to the 10-item condition (51.08). This effect is more than twice as large as the full sample effect. Among higher-income participants, we find no statistically significant difference between FWB scale conditions on the 5-item FWB

[13] The shift toward lower FWB score categories for 5-item condition participants was generally consistent across the six possible categories (see Appendix Table A4). For example, 5.65% of participants in the 5-item condition received a "very low" score, while just 3.59% of participants in the 10-item condition were in this score category (a difference of 2.07 percentage points). Focusing on the "low" and "medium-low" score categories, this difference is 0.56 percentage points and 2.01 percentage points, respectively. Similarly, participants in the 5-item condition were generally *less* likely to fall into higher-scoring categories.

[14] As discussed above, researchers may prefer continuous FWB scores while practitioners may prefer FWB information in discrete form (e.g., "low" or "not low") so they can provide concise feedback to clients.

[15] See Appendix Table A6 for the full regression results for the primary outcomes by household income.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

score. We do find significant treatment effects for both lower- and higher-income participants for the "low 5-item FWB score" outcome, although the magnitude of the effect is more than twice as large for lower-income than for higher-income participants (8.1 versus 3.1 percentage points, respectively).[16]

For self-rated FWB, we find significant differences for only lower-income participants. Among participants in lower-income households, taking the 5-item (versus the 10-item) FWB scale results in (1) a lower self-rated level of FWB by 0.122 points (4.0%) and (2) an increase in the share of participants with low self-rated FWB by 5.4 percentage points (9.1%). For higher-income participants, the estimated treatment effects are substantially smaller and do not approach significance (p-values in the 0.7–0.9 range). We note, however, that for the self-rated FWB outcome measure the treatment effects for lower-income (–0.122) and higher-income (–0.037) participants are not statistically significantly different from one another ($p>0.1$). The treatment effects for the other three primary outcomes differ significantly between lower- and higher-income participants.

As mentioned above, we test the robustness of our findings to three additional income cut-offs – $35,000, $60,000, and $75,000. While there are some differences in the estimated treatment effects as the group defined as lower-income and higher-income changes, the findings are qualitatively similar (see Appendix Table A7). For the 5-item FWB score outcome, for example, the estimated treatment effects range between –2.278 and –1.579 points ($p<0.05$ in all models); these estimated treatment effects are statistically significantly larger (in absolute magnitude) than the estimated treatment effects for higher-income participants for all income cut-offs except $35,000. These estimated treatment effects are in a reasonably narrow range given the different income groupings and support our hypothesis that the negative effect of completing the 5-item versus the 10-item FWB scale is larger among lower- versus higher-income participants.

To summarize the primary results of our experimental manipulations: we find that across two distinct measures completed at different times during the survey, completing the 5-item FWB scale (compared to the 10-item scale) caused reliable declines in consumers' responses to questions about their FWB, and this decline was most prominent among lower-income consumers. These results suggest that lower-income consumers may be particularly influenced by the more negative (or less positive) content of the 5-item scale (compared to the 10-item scale) – a finding with practical implications for how FWB is measured in specific populations of interest. We discuss this further in the conclusion.

5.2. Secondary outcomes

We examine each of the five individual FWB scale items to learn whether the overall effects are driven by changes across all five FWB items, concentrated within a single item, or something in between.

For the full sample, study participants randomly assigned to complete the 5-item versus 10-item FWB scale have lower average scores for only one of the five items – "my finances control my life" (Table 7, top panel). The average response to this item is lower by 0.263

[16] Once again, the shift toward lower FWB categories in the 5-item condition was generally consistent among lower-income participants (Appendix Table A4). For example, while 11.24% of lower-income participants in the 5-item condition received a "very low" score, just 7.11% of lower-income participants in the 10-item condition were in this score category (a difference of 4.13 percentage points). Focusing on the "low" and "medium-low" score categories, this difference is 2.03 percentage points and 1.48 percentage points, respectively. These category differences do not follow a consistent pattern for higher-income participants, although they do produce a negative average treatment effect for the "low 5-item FWB score" primary outcome (Appendix Table A4).

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

Table 7. Effect of taking the abbreviated 5-item versus standard-length 10-item FWB scale on *secondary FWB outcomes*, all participants and by income

	Never have what want in life	Just getting by financially	Concerned money won't last	Money left over	Finances control my life
Estimated effects, full sample					
5-item FWB scale	0.028	–0.067	0.026	–0.032	–0.263***
	(0.028)	(0.031)	(0.031)	(0.029)	(0.029)
	$p = 0.554$	$p = 0.105$	$p = 0.554$	$p = 0.554$	$p = 0.001$
Estimated effects, by income					
5-item FWB scale* low-income	–0.127**	–0.234***	–0.119*	–0.047	–0.280***
	(0.051)	(0.053)	(0.053)	(0.051)	(0.054)
	$p = 0.026$	$p = 0.001$	$p = 0.056$	$p = 0.358$	$p = 0.001$
5-item FWB scale* high-income	0.119***	0.030	0.110**	–0.023	–0.252***
	(0.033)	(0.038)	(0.038)	(0.036)	(0.034)
	$p = 0.003$	$p = 0.642$	$p = 0.013$	$p = 0.642$	$p = 0.001$
Mean of dependent variable for those assigned to 10-item scale					
All	2.49	2.40	2.07	2.45	2.47
Low-income	2.14	2.02	1.73	1.84	2.13
High-income	2.69	2.61	2.27	2.80	2.66

Notes: This table presents OLS regression results for the impact of taking the 5-item (versus 10-item) FWB scale on our five secondary FWB outcomes (never have what want in life, just getting by financially, concerned money won't last, money left over, finances control my life). The impact estimates for all study participants (n = 5,768) are presented in the top panel. The impact estimates for lower-income (n = 2,124) and higher-income (n = 3,644) study participants are presented in the middle panel (see equation 2 in the "Analytic Approach" section for details). Lower income and higher income are defined as having household income of less than $50,000 and at least $50,000, respectively. All models also control for budget order, age, gender, marital status, number of children, work status (currently working, retired), race and ethnicity, educational attainment, income, and immigrant status (see Table 3 for details). Robust standard errors are in parentheses. The *p*-values are adjusted for multiple outcomes using the Romano and Wolf step-down procedure.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

points among participants who completed the 5-item versus 10-item FWB scale, corresponding to a 10.6% decline relative to the 10-item condition. The "my finances control my life" item is the only one that is negative and statistically significant for both lower- and higher-income participants (discussed directly below); this appears to drive the larger effect size for this item (compared to the other four items) in the full sample.

5.2.1. Secondary outcomes by income

Consistent with our hypotheses and primary analyses above, we find that the estimated treatment effects on the secondary outcome measures differ for lower-income and higher-income participants (Table 7, middle panel).

Among lower-income participants, those randomly assigned to complete the 5-item versus 10-item FWB scale have lower average scores for four of the five FWB items (all measures other than "I have money left over at the end of the month" are statistically significant ($p<.10$)). The largest treatment effects are for "my finances control my life"

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

(effect of –0.280, or 13.2%) and "I am just getting by financially" (effect of –0.234, or 11.6%), followed by "never have what want in life" (effect of –0.127, or 5.9%) and "concerned money won't last" (effect of –0.119, or 6.9%).[17]

Interestingly, the only item that does not differ between the two FWB scale conditions for lower-income participants is also the only item with a positive framing: "I have money left over at the end of the month." Furthermore, this item is the most objective question about one's finances. While the mechanism underlying the negative effects of the 5-item versus 10-item FWB scale is not clear, these observations suggest that lower-income consumers may be particularly affected by differences in negative or positive question framings between scale versions about their (comparatively) distressed financial situation.

Among higher-income participants, only one scale item – "my finances control my life" – is significantly lower for participants who completed the 5-item versus 10-item FWB scale (effect of –0.252, or 9.5%). This effect size is similar to the effect size among lower-income participants (effect size of –0.280). However, unlike the pattern for lower-income participants, we find that the treatment effects for two of the scale items – "never have what want in life" and "concerned money won't last" – are positive, indicating that higher-income participants who complete the 5-item scale have higher scores (higher well-being) than those who complete the 10-item FWB scale. For example, for the scale item "concerned money won't last," the average response to this item is *higher* by 0.110 points among higher-income participants who completed the 5-item versus 10-item FWB scale, corresponding to a 4.8% increase relative to the 10-item condition.

The pattern of results is not consistent among higher-income participants, with the 5-item versus the 10-item FWB scale resulting in a *lower* average rating (lower well-being) for one item and *higher* average ratings (higher well-being) for two items. The off-setting effects likely contribute to our more limited findings for the primary outcomes among higher-income participants and to our absence of an effect for some items in the full sample. It is unclear why higher-income participants show this mixed pattern. One possibility is that these participants enter a more future-oriented mindset when completing the 10-item scale (Greenberg and Hershfield, 2019): the second question on the 10-item scale, omitted from the 5-item scale, asks about future financial security ("I am securing my financial future"). Higher-income participants assigned to the 10-item scale could therefore focus more on their financial futures (e.g., retirement security), which may trigger stress about their financial future and lower their overall financial outlook – this is an avenue for future research.

5.3. Robustness check

To test the robustness of our results, we conduct falsification tests. Specifically, we estimate the effect of the FWB scale condition (5-item versus 10-item scale) on four objective survey measures that should not be affected by our manipulation: (1) monthly income, (2) monthly expenses, (3) monthly income minus expenses, and (4) an indicator for whether the participant handles all or most of the household finances.[18] The first three

[17] The 5-item FWB scale begins with a negatively framed question ("Because of my money situation, I feel like I will never have the things I want in life"), while the 10-item scale begins with two positively framed questions ("I could handle a major unexpected expense," and "I am securing my financial future"). Participants who complete the 10-item scale are not exposed to a "negative" question about their financial situation until the third item. As such, it is not surprising that we find a significant difference between the 5-item and 10-item FWB groups when examining even the first item in the 5-item FWB scale.

[18] The specific question is "which of the following statements best describes how involved you are in handling the finances (such as paying bills and budgeting) for yourself or your household?" with response options "I handle all or most of the finances," "someone else and I handle the finances about the same," and "someone else handles all or most of the finances."

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

Table 8. Falsification tests. Effect of taking the abbreviated 5-item versus standard-length 10-item FWB scale on household budget and involvement in household finances

	Monthly income	Monthly spending	Monthly income minus spending	Handle all or most of finances
5-item FWB scale	20.97	19.68	1.29	–0.001
	(405.05)	(97.83)	(394.85)	(0.012)
	$p = 0.959$	$p = 0.841$	$p = 0.997$	$p = 0.947$
Mean of dependent variable for those assigned to the 10-item scale	$5,589.19	$3,094.27	$2,494.92	0.670

Notes: This table presents OLS regression results for the effect of taking the 5-item (versus 10-item) FWB scale on four survey measures that should not be affected by our manipulation: (1) monthly income, (2) monthly spending, (3) monthly income minus spending, and (4) an indicator for whether the participant handles all or most of the household finances. Models for the three budget-related outcomes (income, spending, and income minus spending) are restricted to participants who completed the FWB scale before completing the monthly budget (n=2,914), as it allows us to test whether the scale condition affected these objective measures. The model for the involvement in household finance outcomes includes the full sample, as this question comes later in the survey and is always asked after the FWB scale. All models also control for budget order, age, gender, marital status, number of children, work status (currently working, retired), race and ethnicity, educational attainment, income, and immigrant status (see Table 3 for details). Robust standard errors are in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

measures were collected as part of the monthly budgeting exercise, while involvement in household finances was collected later in the survey. Models for the three budget-related outcomes are restricted to participants who completed the monthly budget after the FWB scale.[19] Imposing this restriction provides a cleaner test of whether the scale condition affects these objective measures.[20] Participants' involvement in household finances was collected later in the survey (after the FWB scale for all participants), so this model includes all study participants.

Consistent with expectations, we find no evidence that the FWB scale condition (5-item versus 10-item scale) is related to any of the four falsification test measures (Table 8). In all cases, the estimated coefficient is not statistically significant, with standard errors that are much larger than the point estimates. These analyses further support our main finding that completing the 5-item versus 10-item FWB scale causes declines in FWB scores and participants' self-rated FWB by demonstrating that the treatment has no effect on other unrelated measures.

6. Discussion and conclusion

Using a large-scale survey experiment conducted with nearly 6,000 participants, this study provides new evidence for how the FWB scale – a scale widely used to assess subjective financial circumstances – can alter consumers' FWB scale scores, responses to individual questions related to FWB, and self-rated FWB. Specifically, we find reliable negative effects of completing the abbreviated 5-item FWB scale compared to completing the standard-length 10-item scale (an average decline of 0.90 points, 95% CI [–1.552, –0.249]; an average increase in the share of participants with a "low" 5-item FWB score of 5.0 percentage points, 95% CI [0.028, 0.071]). Across multiple analyses, we find that completing the 5-item FWB scale leads to lower FWB scores (whether computed using their 5-item score or their 10-item score if the 10-item scale was completed; see Appendix Table A5) and, more

[19] Among our full sample of 5,768 study participants, roughly half (50.5 percent; 2,914 people) completed the monthly budget after the FWB scale.

[20] Our results are not sensitive to this restriction, with qualitatively similar results for the full sample and the sample of participants who completed the monthly budget before the FWB scale.

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

broadly, a worse assessment of consumers' financial circumstances than completing the 10-item scale.

These findings are consistent with our hypothesis that the more negative (and less positive) framing of the 5-item FWB scale (comprising 80% negatively framed items) causes lower levels of subjective well-being than the 10-item scale (comprising only 60% negatively framed items). Additionally, we find consistent evidence that the negative effect of completing the 5-item scale (relative to completing the 10-item scale) is more pronounced among lower-income consumers, both in overall FWB scores (an average decline of 2.3 points, 95% CI [−3.385, −1.171]; an average increase in the share of participants with a "low" 5-item FWB score of 8.1 percentage points, 95% CI [0.041, 0.121]) and in most individual 5-item FWB scale items.

This finding suggests that consumers with fewer financial resources are particularly susceptible to this effect of scale choice, possibly due to exposure to more negative or less positive framing of questions about one's financial situation.

This paper identifies the effect on reported FWB of using the 5-item versus 10-item FWB scale, but does not explore the mechanism(s) underlying our findings. Future research could investigate this using an experiment that systematically varies (1) the order in which each individual FWB item is presented and (2) whether that item is framed in positive or negative terms. In a fully factorial design, these manipulations would enable researchers to estimate the contributing effect of each individual item, as well as item framing, on the outcomes of interest (e.g., response to each item; overall FWB score). Such an experiment would also provide insight into the unexpected pattern of results for our secondary analyses (i.e., the individual FWB items) for higher-income participants.[21] A simpler yet still enlightening future direction might entail recoding all items to be positive (or negative) and comparing outcomes between participants who complete the (all-positive) 5-item versus the (all-positive) 10-item scale. If our hypothesis about differential exposures to negative (versus positive) items is the sole underlying mechanism, then such an experiment should produce no differences in FWB-related outcomes between these groups.

Although an exact mechanism remains to be identified, our findings raise an important implication for how FWB is measured: that the choice of which FWB scale to use is consequential in ways that researchers and practitioners may not expect or desire. The 5-item scale was developed as a secondary option to be used in cases where "time and effort constraints are a significant concern" (CFPB, 2017b, p. 20). Given that the 5-item (versus 10-item) FWB scale is found to systematically worsen lower-income consumers' subjective assessments of their financial situations, practitioners working with lower-income clients in the field, such as those engaged in financial coaching and empowerment programs, should consider using the 10-item FWB scale.

Our results also suggest that researchers should take care when considering the tradeoff between conserving resources by using the abbreviated 5-item scale and improving reliability by using the standard 10-item version. Our results suggest that scale length and reliability are not the only differentiating features in the choice to use one version over the other. However, our findings do *not* suggest that the 5-item FWB scale is no longer a valid measure at a cross-section or over time. We believe the 5-item FWB scale continues to be a good substitute when used for research and evaluation purposes. The strong correlation ($r > .9$) between the 5-item and 10-item FWB scale suggests that these measures capture similar information and would similarly covary with other measures of interest. Rather, our findings show that scores obtained from the 5-item and 10-item scales

[21] Recall that among higher-income participants, those assigned to the 5-item versus the 10-item FWB scale had a *lower* average rating for one item ("my finances control my life") and *higher* average ratings for two items ("never have what want in life" and "concerned money won't last").

https://doi.org/10.1017/flw.2024.11 Published online by Cambridge University Press

should not be directly compared with each other as though they are equivalent, as has been suggested in earlier FWB research.[22] Researchers should take note of this finding and not compare 10-item and 5-item FWB scores across surveys.

The implications of our findings reach beyond FWB into any domain where individuals' preferences, behaviors, attitudes, or aspects of well-being are measured using positively and negatively framed scale items. Subjective scale measures are ubiquitous across consumer finance, economics and psychology, and health and medicine, among other domains, and including questions with positive and negative framings is a well-accepted and common practice in scale development. Our research suggests that the balance of negatively versus positively framed questions alters consumers' reported FWB, which could very well generalize to other measures of well-being or to the practice of scale measurement more broadly. Future research should continue to explore the implications of question framing on individuals' outlook and well-being – and the mechanisms underlying these effects – in consumer finance and other domains.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/flw.2024.11.

Acknowledgements. We acknowledge and thank Abt Associates and the Understanding America Study for their assistance in implementing this research study. We also thank Jeremy Burke, Hector Ortiz, and Sangeetha Malaiyandi for providing feedback on earlier drafts of this work.

Disclaimer. The views expressed are those of the authors and do not necessarily reflect those of the Consumer Financial Protection Bureau or the United States.

References

Burke, Jeremy, J. Michael Collins, and Carly Urban. *Does State-Mandated Financial Education Affect Financial Well-being?* Madison, WI: Center for Financial Security, 2023.

Chyung, Seung Youn, Julie R. Barkin, and Jennifer A. Shamsy. 2018. "Evidence-based survey design: The use of negatively worded items in surveys." *Performance Improvement* 57: 16–25.

Clarke, Damian, Joseph P. Romano, and Michael Wolf. 2020. "The Romano–Wolf multiple-hypothesis correction in Stata." *The Stata Journal* 20: 812–843.

Cohen, Parker, Emily Hoagland, and Kasey Wiedrich. *Measuring Financial Capability and Well-Being in Financial Coaching Programs*. Washington, D.C.: Prosperity Now, 2017.

Collins, J. Michael, and Carly Urban. "Measuring financial well-being over the lifecourse," in *Financial Literacy and Responsible Finance in the FinTech Era, 45–63*. New York: Routledge, 2021.

Consumer Financial Protection Bureau (CFPB). *Measuring Financial Well-Being: A Guide to Using the CFPB Financial Well-Being Scale*. Washington, D.C.: Consumer Financial Protection Bureau, 2015a.

Consumer Financial Protection Bureau (CFPB). *Financial Well-Being: The Goal of Financial Education*. Washington, D.C.: Consumer Financial Protection Bureau, 2015b.

Consumer Financial Protection Bureau (CFPB). *Financial Well-Being in America*. Washington, D.C.: Consumer Financial Protection Bureau, 2017a.

Consumer Financial Protection Bureau (CFPB). *CFPB Financial Well-Being Scale: Scale Development Technical Report*. Washington, D.C.: Consumer Financial Protection Bureau, 2017b.

Consumer Financial Protection Bureau (CFPB). *Financial Well-Being Score Ranges*. Washington, D.C.: Consumer Financial Protection Bureau, 2018.

Consumer Financial Protection Bureau (CFPB). *Getting Started with Measuring Financial Well-Being: A Toolkit for Financial Educators*. Washington, D.C.: Consumer Financial Protection Bureau, 2019.

Consumer Financial Protection Bureau (CFPB). *Making Ends Meet in 2022: Insights from the CFPB Making Ends Meet survey*. Washington, D.C.: Consumer Financial Protection Bureau, 2022.

Dedmon, Michael, Valliappan Lakshmanan, and Haidee Cabusora. *Financial Security in Low- to Moderate-Income Communities: The Consumer Financial Protection Bureau's Well-Being Scale in a Direct Services Setting*. Washington, D.C.: Consumer Financial Protection Bureau, 2019.

[22] The CFPB's 2015 guide to using the CFPB financial well-being scale states that "the scores from the abbreviated scale can be compared directly to the scores from the standard version" (CFPB, 2015a, p. 13).

Fulford, Scott, Marie Rush, and Eric Wilson. *Changes in Consumer Financial Status During the Early Months of the Pandemic*. Washington, D.C.: Consumer Financial Protection Bureau, 2021.
Greenberg, Adam Eric, and Hal E. Hershfield. 2019. "Financial decision making." *Consumer Psychology Review* 2: 17–29.
Hamilton, Leah, Stephen Roll, Mathieu Despard, Elaine Maag, Yung Chun, Laura Brugger, and Michal Grinstein-Weiss. *The Impacts of the 2021 Expanded Child Tax Credit on Family Employment, Nutrition, and Financial Well-Being, Findings from the Social Policy Institute's Child Tax Credit Panel (Wave 2)*. Washington, D.C.: Brookings Institution, 2022.
Herche, Joel, and Brian Engelland. 1996. "Reversed-polarity items and scale unidimensionality." *Journal of the Academy of Marketing Science* 24: 366–374.
Jaroszewicz, Ania, Jon Jachimowicz, Oliver Hauser, and Julian Jamison. 2022. "How Effective is (more) money? Randomizing unconditional cash transfer amounts in the US." *Social Science Research Network Working Paper*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4154000.
Kim, Jinhee. 2004. "Impact of a workplace financial education program on financial attitude, financial behavior, financial well-being, and financial knowledge." *Proceedings of the Association for Financial Counseling and Planning Education* 22: 82–89.
Knowles, Eric S. 1988. "Item context effects on personality scales: Measuring changes the measure." *Journal of Personality and Social Psychology* 55: 312.
Krosnick, Jon. A. and Stanley Presser. "Question and questionnaire design," in *Handbook of Survey Research*, edited by J. D. Wright & P. V. Marsden. San Diego, CA: Elsevier, 2018.
Likert, Rensis. 1932. "A technique for the measurement of attitudes." *Archives of Psychology*, 22: 55.
National Research Council. *Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Experience*. Washington, D.C.: The National Academies Press, 2013.
Netemeyer, Richard G., Dee Warmath, Daniel Fernandes, and John G. Lynch Jr. 2018. "How am I doing? Perceived financial well-being, its potential antecedents, and its relation to overall well-being." *Journal of Consumer Research* 45: 68–89.
Ng, Weiting and Ed Diener. 2014. "What matters to the rich and the poor? Subjective well-being, financial satisfaction, and postmaterialist needs across the world." *Journal of Personality and Social Psychology* 107: 326–338.
Postmus, Judy L., Andrea Hetling, and Gretchen Hoge. 2015. "Evaluating a financial education curriculum as an intervention to improve financial behaviors and financial well-being of survivors of domestic violence: Results from a longitudinal randomized controlled study." *Journal of Consumer Affairs* 49: 250–266.
Powers, Sara, and Sonya Edwards. 2020. "Empowering through education: Understanding financial well-being among low-income older adults." *Innovation in Aging* 4: 732.
Romano, Joseph P., and Michael Wolf. 2005. "Exact and approximate stepdown methods for multiple hypothesis testing." *Journal of the American Statistical Association* 100: 94–108.
Sun, Sicong, Olga Kondratjeva, Stephen P. Roll, Mathieu Despard, and Michal Grinstein-Weiss. *Financial Well-Being in Low-and Moderate-Income Households: How Does it Compare to the General Population?* St. Louis, MO: Washington University, Social Policy Institute, 2018.
Tescher, Jennifer, and David Silberman. *Measuring the Financial Health of Americans*. Washington, D.C.: Brookings Institution, 2021.
Tully, Stephanie M., and Eesha Sharma. 2022. "Consumer wealth." *Consumer Psychology Review* 5: 125–143.
Way, Wendy L., and Nancy Wong. 2010. "Harnessing the power of technology to enhance financial literacy education and personal financial well-being: A review of the literature, proposed model, and action agenda." *Center for Financial Security Working Papers* 10: 1–20.
Weijters, Bert, and Hans Baumgartner. 2012. "Misresponse to reversed and negated items in surveys: A review." *Journal of Marketing Research* 49: 737–747.
Zedlewski, Sheila, Ajay Chaudry, and Margaret Simms. *A New Safety Net for Low-Income Families*. Washington, D.C.: Urban Institute NJ1, 2008.

Cite this article: Heck, Patrick, Caroline Ratcliffe, & Elle Tibbitts (2025). Consumer financial well-being: Does scale choice alter the measure? *Journal of Financial Literacy and Wellbeing*. https://doi.org/10.1017/flw.2024.11