May 2017

CFPB Financial Well-Being Scale

Scale development technical report

Consumer Financial Protection Bureau

Table of contents

1. Introduction

There is a growing consensus that the ultimate goal of financial education is to support individual financial well-being. The vision for the U.S. National Strategy for Financial Literacy is "[s]ustained financial well-being for U.S. individuals and families."[1] This vision is consistent with that of the OECD's International Network on Financial Education (INFE),[2] which describes the ultimate goal of financial literacy as "individual financial well-being."[3]

Researchers and practitioners have relied on a variety of measures to assess financial education program outcomes and to infer financial well-being. While existing measures provide information on factual knowledge, applied financial numeracy, or behaviors like savings or credit use, these measures generally do not take into account consumers' own perspectives on their financial lives.

The 2015 Consumer Financial Protection Bureau (CFPB or Bureau) report, *Financial well-being: The goal of financial education*, provided the first consumer-driven definition of

---

[1] See Financial Literacy and Education Commission, "Promoting Financial Success in the United States: National Strategy for Financial Literacy." Available at http://www.treasury.gov/resource-center/financial-education/Documents/NationalStrategyBook_12310%20(2).pdf (2011) (accessed May 19, 2017): p. 7.

[2] The Organization for Economic Cooperation and Development (OECD) is a unique forum where the governments of 34 democracies with market economies work with each other, as well as with more than 70 non-member economies to promote economic growth, prosperity, and sustainable development. To learn more, see http://usoecd.usmission.gov/mission/overview.html.

[3] See OECD INFE, "Measuring Financial Literacy: Questionnaire and Guidance Notes for Conducting an Internationally Comparable Survey of Financial Literacy." Paris: OECD. Available at http://www.oecd.org/finance/financial-education/49319977.pdf (2011) (accessed May 19, 2017): p. 3.

financial well-being for the financial education field.[4] This technical report describes the next milestone in our financial well-being research program: the development of a scale – a set of questions along with a scoring procedure – to measure individual financial well-being.

This report is designed to serve as a resource for researchers and others who seek the technical details related to the development of the scale and the scoring procedures. It describes the process used to develop the new financial well-being scale and scoring procedures, including the scoring rubric and the reliability of scores derived from its use.

Readers seeking to learn about the methodology and detailed findings related to the consumer-driven financial well-being definition, established in the initial phase of our research, should refer to the CFPB's 2015 report, *Financial well-being: The goal of financial education*. Because it lays the groundwork for this scale, that report is an important reference source for this technical report.

The consumer-driven financial well-being definition developed by the CFPB is multi-faceted, not directly observable, and describes a continuum ranging from feeling severe financial stress to being highly satisfied with one's financial situation. Through our prior work, we learned that consumers' perception of financial well-being corresponds to their sense of financial security and financial freedom of choice, in the present and when considering the future.

As part of the scale development process, we therefore sought to learn whether financial well-being could be represented by a single measure.[5] Our findings confirmed that one multidimensional measure could reflect a consumer's perception of financial well-being. This overall measure is the focus of this report.

The CFPB Financial Well-Being Scale, described in greater detail in the rest of this report, is:

- A consumer-driven measure based on a definition of financial well-being that draws on insights from both consumers and financial practitioners. Evaluation of scale items

---

[4] Available at consumerfinance.gov/reports/financial-well-being/.

[5] Measures of the individual elements (i.e., financial security and freedom of choice now and in the future) were not explored in this work.

included cognitive interviewing and testing to ensure accurate comprehension of questions, as well as review by academic experts.

- A measurement tool developed over multiple waves of quantitative testing using both classical scale development methods and state-of-the-art Item Response Theory (IRT) techniques to identify appropriate items and how best to combine those items to produce reliable scores of the financial well-being construct.

- A common metric adjusted for age group[6] and mode of administration that allows an "apples-to-apples" comparison of scores across consumers. It can also be used to assess a consumer's current state of financial well-being, to track their progress over time, and to understand how other factors, including program interventions, affect financial well-being.

- A free and publicly available survey instrument and measurement scale.

Use of the scale does not require a thorough understanding of the technical details in this report. Those seeking to use the scale "off the shelf" may refer to a companion publication, *Measuring financial well-being: A guide to using the CFPB Financial Well-Being Scale*.[7]

---

[6] Prior qualitative findings suggested that while descriptions of financial well-being by working age and older adults were largely consistent, minor systematic differences could exist. Therefore the scale development process investigated and ultimately adjusted for such differences.

[7] This user guide is available at consumerfinance.gov/financial-well-being.

2. Defining financial well-being

Our research program on financial well-being began with a carefully-designed set of research activities to understand and formally define financial well-being. Our goal was to develop a definition of financial well-being for all adults that could be applicable across different economic, geographic, and age-related contexts.[8] Here, we briefly summarize the approach and findings from the first phase of this project.[9]

In order to arrive at a definition of financial well-being that reflected the consumer perspective, CFPB's research team listened to people around the country talk about what financial well-being means to them.[10] As a result of this work we developed the following definition.

> Financial well-being is a state of being wherein 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.

---

[8] "Adult" refers to those ages 18 and older. Further "working age" is ages 18 to 61, while "older" is over age 62, a proxy for separating (or close to separating) from the full-time workforce. Certain age-based distinctions are specified in the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, 12 U.S.C. § 5493(g)(1).

[9] For more information on the methodology and detailed findings from this first phase of research, see *Financial well-being: The goal of financial education*, available at consumerfinance.gov/reports/financial-well-being/.

[10] The research team responsible for conducting the research and analysis to develop this definition of financial well-being, as well as to develop the scale to measure financial well-being, included Bureau staff as well as a team of research contractors led by the Corporation for Enterprise Development (CFED), including the University of Wisconsin-Madison Center for Financial Security, the Urban Institute, ICF International, and Vector Psychometric Group. The research was funded under a competitive award; contract number CFP-12-Z-00019.

Overall, perspectives on the meaning of financial well-being were remarkably similar for consumers of different ages, incomes, and locations around the country.

Combined with a review of research and consultation with leading experts, the CFPB found that financial well-being includes the following four elements:

- Having control over day-to-day, month-to-month finances.

- Having the capacity to absorb a financial shock.

- Being on track to meet your financial goals.

- Having the financial freedom to make the choices that allow you to enjoy life.

Another way to think about this is that financial well-being is the feeling of having financial security and financial freedom of choice, in the present and when considering the future.

FIGURE 1:   THE FOUR ELEMENTS OF FINANCIAL WELL-BEING

	Present	Future
Security	Control over your day-to-day, month-to-month finances	Capacity to absorb a financial shock
Freedom of choice	Financial freedom to make choices to enjoy life	On track to meet your financial goals

The definition reflects a concept of financial well-being that is inherently subjective, and, therefore, not directly observable.

Grounded in the consumer insights gathered from this first stage, the research team next undertook a state-of-the-art process to develop a scale to measure financial well-being, based on the construct definition described above.

3. Overview of a typical approach to scale development

In this section, we describe a typical approach to scale development. This discussion provides a backdrop for Section 4 in which the specific method used to develop the CFPB Financial Well-Being Scale is presented.

3.1 What is a scale?

A scale is an instrument, made up of multiple questions or "items," that yields a composite measure of something that is not directly observable, such as an attitude or ability.[11] Such scales are often described as "reflective scales," in which one's level of the attribute being measured is reflected in the response given to individual items.[12] Therefore, a scale was the natural choice for a tool to measure this inherently subjective concept of financial well-being. Response options for each item in a scale can range in complexity from a simple "Yes" or "No" response to a Likert scale or continuum, such as the level of agreement or disagreement along a range of response

---

[11] A scale can also refer to a response scale, which is the way a survey question is rated or measured. For example, one of the response scales used in the CFPB Financial Well-being Scale asks respondents "How well does this statement describe you?" and allows them to respond with one of 5 options, from "Describes me completely" to "Does not describe me at all."

[12] Bagozzi, R. P. (2011). Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations. *MIS Quarterly, 35*(2), 261-292.

options.

The quality of a scale is judged by its reliability and validity.

- Reliability describes how consistent and precise the scores produced by the scale are. The more reliable the scale, the less measurement error it is producing.

- Validity identifies whether the scale is measuring the intended construct. There are multiple, interrelated types of validity but they can be divided into two broad groupings: content-related validity, which assesses whether the scale relates to its underlying theoretical concepts, and criterion-related validity, which is used to determine whether the scale can predict performance on a related measure.

3.2 Typical scale development process

Scale development is an iterative process that begins with many more items than will go into the final scale. The multiple rounds of data collection and analysis serve to winnow down the initial pool of items and narrow in on the ones that best reflect the underlying construct. While there is variability depending on the particular project, the process and standards that are commonly followed when developing "reflective" scales include:

1. Establishing the concepts to be measured based on relevant theory, prior research, and qualitative evidence.
2. Creating a relatively large initial or "candidate" set of questions, with the goal of refining the set of questions into a shorter final scale based on iterative testing.
3. Gathering expert and user feedback on instructions, each individual question, and response set(s) to ensure that they relate to the intended construct, make sense to respondents and are interpreted as intended.

4. Conducting psychometric analyses[13] on the data gathered for candidate items through iterative surveys to create the scale model, narrow down the items in the final scale, and assess the reliability of the measure.[14] Examinations regarding the invariance of item properties across targeted grouping variables (e.g. age, sex, mode of administration,[15] etc.) may also be done.

5. Using validation questions in the surveys to check for content and criterion-related validity.

Establishing the validity of inferences derived from the use of a new scale is an ongoing process, as more data are collected over time and through different studies.[16]

3.3 Scoring a scale

The value of a scale is realized in its application. Applications may involve use of the scale to better understand a given respondent or use in a broader analysis involving the construct being measured. Applying a scale requires the combination of individual scale item responses into a

---

[13] Analyses may include classical test theory techniques, such as Cronbach's alpha and item-total correlations, factor analysis (both exploratory and confirmatory), and/or IRT analyses.

[14] A survey is the typical way in which a scale is deployed or implemented. A survey encompasses all aspects of the research process; the scales developed for this project were part of a larger survey that respondents received, which also included demographic and validation questions.

[15] The relevance of data collection methods and modes to resulting data quality and interpretation is well-documented. See, e.g., Chapter 5 of Groves et al. (2009) *Survey methodology*.

[16] There are multiple criteria necessary to assess the psychometric soundness of a new behavioral measure. The American Psychological Association (1985) suggests that scales should demonstrate content validity, criterion-related validity, construct validity, and internal consistency. Content validity pertains to the adequacy with which a measure assesses the domain of interest. Criterion-related validity describes the relationship between the new measure of interest and another independent measure. Construct validity is concerned with the relationship of the measure to the underlying attributes it is attempting to assess. Internal consistency refers to the homogeneity of the items in the measure or the extent to which item responses correlate with the total test score. There are specific practices that can be utilized to establish evidence of validity and reliability of new measures.

single "score."

There are a number of ways individual item responses can be combined to produce a single score. Many instruments rely on "summed scores" (i.e., simply summing or taking the mean of all relevant individual item responses), although this method assumes that all items are equally good indicators of the construct of interest. This assumption is often hard to support. In other cases, a score is obtained by weighting each item by its factor loading or relatedness to the scale before summing or averaging.

An increasingly popular and psychometrically rigorous alternative is to use scores obtained via item response theory (IRT), which allows for each item's *relatedness* to the concept (e.g., financial well-being) and *degree of severity*,[17] as well as respondent group (e.g., age) properties, to be accounted for when scoring. The assumptions underlying these IRT-based scores – broadly speaking that items may differ in their relationship to the construct of interest - are often more likely to be met in real-world situations and can offer superior reliability compared to other scoring methods (e.g., summed scores).[18]

---

[17] Degree of severity means how "much" of the attribute being measured, such as financial well-being, is needed for a respondent to choose a certain item response.

[18] See Edwards (2009) for an introduction to IRT models and further explanation of the concepts of relatedness and severity as item properties.

4. Development of the CFPB Financial Well-Being Scale

The process for developing the CFPB Financial Well-Being Scale followed the five steps outlined in Section 3, with a few adjustments to increase precision of the scores. The primary enhancement to the process was the incorporation of IRT modeling.[19] IRT is the method of choice for standardized educational testing and other domains in which score precision is critical, such as health outcomes.

While both classic approaches and IRT seek a score that indicates a given respondent's level of financial well-being (or other latent construct), classic approaches take more of an additive approach – summing (or weighting and summing) responses to individual items to reflect the respondent's level of financial well-being. IRT asks "what is the probability that a given respondent will have a specific financial well-being score, given how they answered the questions we asked?" IRT and related methods (such as structural equation modeling) then go further to examine whether the probability varies as a function of respondent characteristics or the mode of survey administration (e.g., self- versus interviewer-administered).

This section describes our process in the following order: concept and initial item pool development (Section 4.1); survey data collection and analysis to create the scale model and select scale items (Section 4.2); testing for age group considerations and mode effects (Section

---

[19] For general information about IRT and its applications, see Embretson and Reise (2013) or Thissen and Wainer (2001). For a specific example of how IRT has been applied to the development of a scale measuring financial knowledge, including a description of the IRT process and its contribution to measurement precision, see Knoll and Houts (2012).

4.3); and validation (Section 4.4).

The project goal was to create a financial well-being scale that could be used for all adults in the United States. Due to indications of potential age-related differences in perceptions of financial well-being from the initial qualitative research,[20] the research team chose to examine whether working age consumers might answer questions systematically differently from their older counterparts. Thus, all data collected in this study consisted of two separate samples - older consumers (aged 62 years and older) and younger consumers (aged 18-61 years). Initial analyses were conducted separately with data from these two groups, as expert opinion suggested that there were likely meaningful differences in how respondents would think about financial well-being as well as answer questions depending on age. Although data collection and analysis were separated by age group, the general design of the study was the same across the two groups. Analysis of differential item functioning[21] between the two age groups is discussed in Section 4.3, below.

Table 1, below, provides a high-level overview of the number of participants at each stage of data collection (interviews and surveys) throughout the scale development process.[22] In Phase 1 of the project, qualitative interviews were conducted with 59 adult consumers and 30 financial practitioners.[23] Information gleaned through these interviews, along with expert opinion, was

---

[20] See Section 3.2 in *Financial well-being: The goal of financial education,* available at consumerfinance.gov/reports/financial-well-being/, for a discussion of this particular finding.

[21] Differential item functioning (DIF) occurs when groups (such as defined by age or mode of survey administration) have different probabilities of providing a particular response to an item in the multi-item scale after controlling for overall scale scores. An item is labeled as having DIF when people with the same level of a latent trait but from different groups have an unequal probability of giving a response.

[22] The qualitative interviews used to develop the financial well-being construct definition and initial item pool were approved by the Office of Management and Budget (OMB) under OMB No. 3170-0036. All further information collection activities used to refine and select items for the scale were approved under OMB No. 3170-0043.

[23] For detailed information on the qualitative interviews, see *Financial well-being: The goal of financial education*, available at consumerfinance.gov/reports/financial-well-being.

used to develop an initial pool of candidate items. The initial pool of items was assessed through cognitive interviews with 19 adult consumers to test and refine the item wording. Three rounds of surveys totaling over 14,000 respondents were used to develop and confirm the scale model, narrow down the questions to the best indicators of the construct, and determine how survey responses may differ when self-administered online versus when administered by telephone, for younger and older adults separately. More detail on each of these activities follows below.

TABLE 1:   NUMBER OF PARTICIPANTS FOR EACH PHASE OF DATA COLLECTION

Interviews	Adults 18-61	Adults 62+	All Consumers	Practitioners
Qualitative interviews	41	18	59	30
Cognitive interviews	12	7	19	

Surveys	Adults 18-61	Adults 62+	All Consumers
Round 1 survey (online)	3,542	958	4,500
Round 2 survey (online)	6,350	1,549	7,899
Round 3 survey			
Online	500	500	1,000
Telephone	519	481	1,000
Total Survey Respondents*	**10,911**	**3,488**	**14,399**

*Subset of survey respondents included in final scale development analysis sample = 10,804.

4.1 Concept and initial item pool development

In order to develop an initial pool of candidate items (i.e., statements likely to reflect underlying financial well-being), the study team reviewed the transcripts of the one-on-one qualitative interviews that were conducted during the first phase of this research project with a diverse set of consumers and financial practitioners around the United States. Drawing heavily on the phrasing and language used by consumers during the interviews, the team developed potential items for testing. In addition, the team conducted a review of prior research surveys to learn what questions had been previously asked on related topics and incorporated some of them into the initial item pool. Items in the initial pool were intentionally designed to be income and life-stage agnostic. For example, none of them refer to specific dollar amounts, so that all items can be viewed in the context of the respondent's own economic situation and standards.

These items were reviewed by academic experts and 47 were selected to be field-tested with consumers using cognitive interviewing to ensure that respondents adequately understood all of the questions and to make refinements to the question wording where needed. These consumers represented diverse educational levels, socio-economic status, and life stages. This process identified any potential comprehension problems that could impact survey respondents' answers to the items. During this process, alternative wordings for items and responses were also tested. Because survey questions are either negatively or positively worded, the testing process included an analysis of how the direction, also known as polarity, of the items influenced responses. Following refinements to statement wording and discarding one item that was deemed repetitive, the study team moved 46 candidate items forward to the next stage of scale development.

4.2 Survey data collection & analysis

Using the refined pool of items, the research team fielded three rounds of surveys via a large nationally-recognized provider of survey data. No respondents participated in more than one round. The first two rounds of surveys were fielded online to large, diverse groups of consumers.[24] As the study moved from round one to round two, exploratory and confirmatory factor analysis, as well as IRT modeling, were used to develop, refine and shorten the scale from the initial set of items. The third survey round was fielded both online and by telephone and was used to confirm and refine the scale developed following the second round of online data collection, and to test for mode effects. Each round of survey data collection is briefly described below, followed by a description of the data analysis conducted to develop the scale model and to select the items.

---

[24] IRT scale development methods require a large and diverse sample of respondents, but not one that is nationally representative. Accordingly, survey respondents were recruited by Survey Sampling International (SSI) such that the sample was roughly similar to the U.S. population based on Census demographics for working age and older consumers, but not nationally representative. Some respondents were removed from the analysis sample due to data quality concerns. See Appendix A for demographics of survey respondents included in the scale development analysis sample.

ROUND 1

The first survey round fielded the 46 candidate items to 4,500 respondents, divided, as described in Table 1, between adults ages 18 to 61 and adults ages 62 and older. The sample was split fairly evenly between men and women, and included respondents across multiple racial and ethnic categories and areas of the country. The purpose of this first survey round was to collect data needed to begin to develop the scale model (via empirical determination of the most appropriate structure underlying the candidate financial well-being items).

ROUND 2

The purpose of the second survey round was to collect data needed to (1) confirm the initial scale model developed from round one survey data with a new data set, (2) examine age-group-related differences in responding to the items (3) select items for a preliminary version of the scale based on item content and item performance, and (4) develop the scale scoring. This survey round fielded the 44 candidate items[25] resulting from analysis of Round 1 survey data analysis to 7,899 respondents with a mix of adult ages, racial and ethnic categories, income categories, and U.S. geographic regions.

ROUND 3

The third and final round of survey data collection served two purposes: (1) to provide one more test of the reliability of the scale, and (2) to compare multiple modes or methods of survey administration: self-administered and interviewer administered. The 12 selected items, based on the analysis of the data collected in round two, were fielded online in round three to 1,000 respondents and by telephone to another 1,000 respondents. This type of "mode-testing" assesses whether respondents answer questions differently depending on how questions are presented to respondents (e.g., read aloud by interviewer or read directly from text by the respondent).

---

[25] These 44 items included an A/B split which allowed for examination of positive and negative wording for particular items.

4.2.1 Model development

The CFPB Financial Well-Being Scale was developed using modern measurement theory analyses (including factor analysis and IRT modeling). [26] The goal was to develop a measure with strong psychometric properties that can be easily used across varied populations in both experimental and survey research.

The statistical analysis undertaken to develop the CFPB Financial Well-Being Scale included basic descriptive statistics as well as exploratory and confirmatory factor analysis using standard statistical programs to allow for the empirical determination of the most appropriate structure underlying the candidate financial well-being items. This process sought to answer whether 1) financial well-being could be represented by a single measure and 2) if only a single concept was being measured with the items in the scale.

The researchers tested several different models, including a unidimensional model, several multidimensional models defined *a priori* by substantive item content, and other models indicated by initial exploratory analyses that reflected methodological considerations (such as positive versus negative item wording). The model found to fit the data best took the form of a bi-factor model.[27] This model contains one factor related to the latent financial well-being construct, which applies to all 10 scale items, plus two additional factors/dimensions that are associated with the polarity of the item (i.e., whether the item was phrased negatively or positively).[28] The second and third factors apply only to items whose wording matches the direction of the appropriate secondary factor. Hence each individual item has only two factors. See Appendix B for further description of the model used.

---

[26] Psychometric analyses were conducted by Vector Psychometric Group.

[27] See Gibbons, R.D., & Hedeker, D. (1992). Full information item bi-factor analysis. *Psychometrika, 57,* 423-436.

[28] The observed dependencies among items in the CFPB Financial Well-Being Scale related to the positive/negative orientation of the item wording is consistent with the conclusion of the 2013 National Research Council publication *Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Experience,* that "Both positive and negative emotions must be accounted for in [experienced well-being] measurement, as research shows that they do not simply move in an inverse way…Therefore, assessments of [experienced well-being] should include both positive and negative dimensions in order for meaningful inferences to be drawn."

The final model provides us with an overall, general "financial well-being" score that is conceptually sound, while also removing "noise" in the item responses stemming from the positive and negative wording of the items. While the four substantive elements previously mentioned are not explicitly modelled (e.g., there is no "freedom of choice" score that is calculated), items were selected for inclusion in the final version of the CFPB Financial Well-Being Scale to ensure that the four substantive areas were all represented.

4.2.2   Item selection

Using the model structure empirically supported during the model development phase, IRT analyses were then conducted to obtain objective values regarding item performance, otherwise known as "item parameters."[29] Each item had one parameter/value that described how strongly related to the general concept of financial well-being the item is and several severity parameters/values that are used to describe how "much" financial well-being was needed to provide a certain response (for example, what level of financial well-being is needed to answer "Strongly Agree" to "I am securing my financial future").

In deciding which item to include in the scale, the research team considered the following criteria: (1) Items more strongly related to general financial well-being were preferred over those less strongly related to our concept of interest; and (2) a wide range of severity items across all items was desirable given the "general purpose" nature of the scale. [30] We wanted a wide range of severity values to ensure that we could accurately measure low financial well-being individuals, high financial well-being individuals, and everyone in between.

The analysis of item performance following the first round of survey data collection began with 46 financial well-being candidate items. By Round 3, only the 12 best-performing questions, based on Round 2 results, were presented to respondents. Of these, 10 questions were selected

---

[29] IRT analyses were conducted using the software package flexMIRT® 2.0. See Houts, C. R., & Cai, L. (2013). flexMIRT user's manual version 2: Flexible multilevel multidimensional item analysis and test scoring. Chapel Hill, NC: Vector Psychometric Group. Available at: https://flexmirt.vpgcentral.com/downloads/flexmirtusermanual.pdf

[30] See Edwards (2009) for further discussion of these considerations.

for the final version of the scale, based on item performance (IRT results), overall scale reliability, and item content (ensuring that the four substantive areas of interest were all represented). The number of items tested at each phase is shown in Table 2.

TABLE 2: NUMBER OF ITEMS FOR EACH PHASE OF DATA COLLECTION

Data Collection Phase	Tested Items
Cognitive Interviews	47*
Phase 1	46
Phase 2	44
Phase 3	12
Final Scale	10

*One item was deemed repetitive and discarded following cognitive testing

For the final 10-item scale, we employ two different 5-point categorical response sets, providing labels that indicate the content of the item:

Describes me: "Describes me" response set, where the participants are told to select the response that indicates how well each statement describes them, where "5" indicates that the statement "Describes me completely" and "1" indicates that the statement "Does not describe me at all."

Frequency: "How often" response set, where the participants are told to select the response that best indicates how often they would be able to make this statement about themselves using a 5-point scale, where "5" indicates "Always" and "1" indicates "Never."

Abbreviated Scale

A further refinement was to select a subset of the 10-item scale that would still provide a useful indicator of financial well-being in settings where collecting 10 items is not feasible. The items were selected based on item performance (as before, how strongly related to financial well-being and the severity values for each item), and to ensure all four substantive elements of financial well-being (control, capacity to absorb a shock, being on track to meet goals, and freedom of choice) were represented. The correlation between scores from the 5-item version and the 10-item version of the scale estimated for the same groups of consumers was quite high, 0.94 (p <

0.001) where 1.0 means perfectly correlated. However, 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.

4.3 Age-group and mode differences

Initial models were tested for younger and older consumers separately but with the goal, if supported by the data, to find a common model across the two age groups. This would allow for direct comparisons between groups and the use of the same questions and response sets (but potentially different scoring) for the different groups. Further, the third round of the data used different modes of data collection in order to test whether responses remained consistent across self-administered and interviewer-administered modes. The item parameters (relatedness and severity, as described earlier) for self-administration were developed through online data collection, and the item parameters for interviewer administration were developed through telephone data collection.[31] Thus, four groups were produced and examined:

- Respondents ages 18-61 who self-administered the scale

- Respondents ages 18-61 who had the scale read to them by an interviewer

- Respondents ages 62 and older who self-administered the scale

- Respondents ages 62 and older who had the scale read to them by an interviewer

While the analyses support using the same underlying model structure (a bi-factor model with specific factors for negative and positive wording) for the CFPB Financial Well-Being Scale in all 4 groups, we found meaningful differences in item performance for some items among the

---

[31] Although the study team also collected about 500 pencil and paper surveys using respondents recruited through mall intercept, the research team did not have confidence in these data and they were dropped from this analysis. Prior studies support the idea that Web-based and paper surveys will have similar results under similar conditions. Groves et al. (2009) find that "for a large array of nonthreatening survey items, interviewer-administered surveys yield similar results whether administered by telephone or face to face, and self-administered methods yield similar results whether they are done on paper or via the Web."

groups. Therefore, different item parameters were estimated for the 4 groups. These item parameters are available in Appendix B.

Because these item parameters were estimated from one large multiple-group model, the resulting scores from all 4 groups are directly comparable to each other (that is, they are on the same metric – a score of 45 in the older-interview group means the same thing as a 45 in the younger-self-administered group). Further, analyses showed that all scoring rubrics (sets of item parameters) from this model displayed marginal reliability at 0.89 or better for the standard (10 item) version of the scale.[32] The marginal reliability of the abbreviated (5-item) scale is also above .80 for all groups. Marginal reliability can be interpreted similar to Cronbach's alpha. Table 6 in Appendix C has reliability estimates for each scoring rubric.

4.4 Validation

To assess the validity of the inferences made using scores from the financial well-being scale, each of the three rounds of survey data collection included the scale questions as well as various validation measures: other questions whose answers we expect to have positive or negative associations with financial well-being, as measured by the scale. Although there were no existing, precise measures of financial well-being available to refer to, there are related concepts such as financial satisfaction, credit scores, and experiences of economic shocks and material hardships. Thus, if the new scale is measuring what it is supposed to be measuring, there will be a statistically significant relationship in an expected direction between those measures and CFPB Financial Well-Being Scale scores. The validation measures used in this project came predominantly from existing scales for related concepts and other items frequently used in

---

[32] Unlike classical test theory (e.g., Cronbach's alpha) which assumes that all scores from a scale are equally well estimated/equally reliable, IRT acknowledges that the reliability of the scores can vary (e.g., average scores may be well estimated while extreme scores may be less well estimated). For this reason, in IRT reliability is represented by a function/curve that may change over the range of the latent construct, rather than a single value. For comparison to more traditional results, however, the marginal reliability statistic provides a single value summary of the marginal reliability (a weighted average of the reliability values over the whole continuum). See Thissen & Wainer (2001; pgs 117-119) for a more detailed discussion.

surveys.

Appendix D provides a full list of these validators and the corresponding data:

- Credit rating (self-reported)

- Confidence in one's ability to come up with $2,000 in 30 days

- Having three months of expenses in savings

- Self-rated current financial situation

- Experience of negative economic events (household income shocks, foreclosure, health emergency, debt in collections)

- Inability to afford food and/or medical care (material hardship)

An analysis of the relationship between scores on the new scale and these measures shows the expected relationships in both direction and magnitude. For instance, analysis found a positive association between respondents' financial well-being scores and their answers to the question, "How would you rate your current financial situation?"[33] as well as with commonly used measures of financial resilience, such as having three months of expenses in savings and confidence in one's ability to come up with $2,000 in 30 days.

Other validation items displayed similar statistically significant results. Respondents' financial well-being scores were positively correlated with self-assessed credit quality and negatively correlated with having a debt in collections, as well as likelihood of experiencing economic hardships in the last year. These results are directionally consistent with *a priori* hypotheses about the expected relationship between these other financial concepts and financial well-being.

The CFPB Financial Well-Being Scale scores are also correlated in the expected direction with commonly used indicators of socio-economic status such as income and education.

---

[33] The questions had 7 possible responses ranging from "Extremely bad" to "Extremely good."

5. CFPB Financial Well-Being Scale

5.1 Items

Through the analytical process described in Section 4, the research team developed a 10-item scale to measure financial well-being. Table 3 shows the 10 items that make up the final standard form of the CFPB Financial Well-Being Scale. The first column contains the text of each item and the second column provides the response options associated with those items. The items cover all four elements of the CFPB's definition of financial well-being: control over finances, capacity to absorb a financial shock, being on track to meet financial goals, and having the financial freedom to enjoy life, using both positive and negative phrasing.

TABLE 3:   CFPB FINANCIAL WELL-BEING SCALE

Questions	Response Options
How well does this statement describe you or your situation?	
1. I could handle a major unexpected expense	
2. I am securing my financial future	Describes me completely
3. Because of my money situation, I feel like I will never have the things I want in life* †	Describes me very well
4. I can enjoy life because of the way I'm managing my money	Describes me somewhat
	Describes me very little
5. I am just getting by financially* †	Does not describe me at all
6. I am concerned that the money I have or will save won't last * †	

Questions	Response Options
How often does this statement apply to you?	Always
7. Giving a gift for a wedding, birthday or other occasion would put a strain on my finances for the month*	Often
8. I have money left over at the end of the month†	Sometimes
9. I am behind with my finances*	Rarely
10. My finances control my life* †	Never

* Denotes items for which the response options are "reverse coded."

† Denotes items that are part of the abbreviated (5 item) scale.

The abbreviated 5-question version of the scale (items denoted with a † in the table above) still provides adequate coverage of the four elements of financial well-being as well as both positive and negative phrasing. The correlation between the 10 and 5 item versions of the scale is 0.94 (*p<0.001*). Both versions of the scale can be found formatted as printable questionnaires at consumerfinance.gov/financial-well-being.

5.2 Scoring

5.2.1   CFPB Financial Well-Being Scale scores

In this section we detail two possible methods for obtaining scores from the CFPB Financial Well-Being Scale: table-based scoring and software-based scoring. Both methods rely on the results of our IRT model. The use of IRT-based scoring procedures, rather than simple summed scores, automatically places the obtained scores on a common metric, regardless of group (i.e., age group or administration mode) or use of 5- versus 10-item version. This ensures the scores are directly comparable to one another. Further, the recommended IRT-based scoring also incorporates group values (the varied means and variances from the four age- and mode-specific groups reported earlier), resulting in more precise scores for each individual than if item responses were just summed or averaged, with no consideration of administration mode or age group.

Typical IRT-based scores are on the standard normal distribution – that is, a "bell-shaped" curve that has a mean of zero and a standard deviation value of one. For ease of interpretation and communication, the CFPB Financial Well-Being Scale scores are whole numbers between 0 and 100. To get to the CFPB Financial Well-Being Scale metric from a raw IRT score, the value is multiplied by 15, added to 50, and then rounded to the nearest whole number:

This provides a score distribution that ranges from approximately 0 to 100 and is centered at 50.

CFPB Financial Well-Being Scale Score = round ((IRT FWB score*15) + 50).

5.2.2   Scoring Methods

We recommend two methods for scoring the CFPB Financial Well-Being Scale: table-based IRT scoring or software-based IRT scoring. The former is a simple and expedient way to approximate a directly estimated IRT score when a respondent answers all the questions in the scale. The table-based method should not be used if a respondent does not answer all questions in the scale. Software-based scoring allows for direct estimation of individual IRT scores using the individual's full response pattern, provides the most precise individual scores, and can be used to produce common-metric (i.e. comparable) scores even if some questions in the scale are not answered.

Table-based IRT scoring

As a convenience to researchers and practitioners who do not have access to IRT software, or prefer not to use the scoring program described in Appendix F, the research team provides a table-based scoring procedure—a look-up table—that allows researchers to find an approximate IRT-based score for any given sum score. The table-based scoring procedure has two steps: calculating the total response value (i.e., the raw summed score) for each respondent and then looking that summed score up in the table to find the associated CFPB Financial Well-Being Scale score (on the 0-100 metric). This procedure is demonstrated in the scoring worksheets provided in Appendix E.

The table-based method should not be used if a respondent does not answer all questions in the scale (meaning 9 or fewer of the items answered on the 10 item scale), because skipping items reduces the overall number of items contributing values to the final score. For instance, an individual who answered all of the questions and has a total response value of 30 out of a total

possible 40 likely does not have the same financial well-being as an individual who skipped two of the items and therefore has a total response value of 30 out of a total possible 32.

The CFPB Financial Well-Being Scale score reported in the look-up table is the score for the most likely answer pattern for any given total response value or sum of answers.[34] To simplify communication of financial well-being scores, the tables shown in Appendix E translate the IRT scores onto a 0-100 point scale.[35] Because the table-based method uses summed scores rather than an individual's actual response pattern, the method is an approximation of directly estimated IRT scores when there are no missing item responses. Table-based scores show correlations of over .90 with response pattern IRT scores.

This table-based method of scoring holds two advantages over simply taking a sum or average of scores across the items. First, because it relies on IRT scoring for its foundation, the lookup table offers a more precise estimate than methods that assume each item contributes equally to underlying financial well-being. Further, the table-based scores account for variation introduced by item polarity, age group of respondent, and mode of data collection. Using the simple average or sum as the final score ignores these important sources of variation. Second, scores produced using table-based IRT scoring can be directly compared to data collected by survey researchers and financial education practitioners that are using either of the two IRT scoring methods provided in this report for either the 5-item or 10-item versions of the CFPB Financial Well-Being Scale. To the extent researchers want a score comparable across studies, scale versions, etc. and a simple method of score calculation, these look-up tables may be the most expedient option.

An important question is what to do with respondents who fail to complete all questions in the scale, as simply discarding such observations could be inefficient and costly. The most direct solution is for researchers to use IRT software for scoring instead of the table-based method. Software-based IRT scoring (described below), which uses the estimated item parameters

---

[34] This method has been described as a 'psuedo-EAP' score and is described in Bonifay (2014).

[35] Not shown in these tables is the conversion from the summed totals into IRT-based EAP (Expected *a posteriori*) scores, which is the intermediate step before the EAP score is transformed to a 0 -100 scale.

provided in Appendix B and the full response pattern of answers given by each individual, can calculate scores even for respondents with missing item responses.

Software-based IRT scoring

The software-based IRT scoring method allows researchers to estimate the most precise financial well-being scores for each respondent. This approach uses the individual's full response pattern, rather than the sum of the responses, to determine a financial well-being score for each individual in the data, providing more precise scores than summed scores and even scores found via the provided look-up tables. Additionally, this method can accurately estimate scores even with missing responses. This method requires the use of IRT-specific software (e.g., flexMIRT, IRTPRO), using IRT-specific packages or toolboxes in more general purpose statistical software (e.g., R, Stata, MPLUS), or programming the model and item parameters in a statistical package yourself. Appendix F describes procedures for scoring responses according to this method using a custom Stata package.[36]

Because differential item functioning was found across both age groups and survey modes, item parameters are provided in Appendix B for scoring four different types of scale respondents:

- Respondents ages 18-61 who self-administered the scale

- Respondents ages 18-61 who had the scale read to them by an interviewer

- Respondents ages 62 and older who self-administered the scale

- Respondents ages 62 and older who had the scale read to them by an interviewer

Use of the appropriate parameters produces a financial well-being score that is directly comparable across age groups, modes, and scale versions (10 versus 5 item).

---

[36] This Stata package was developed for the CFPB by Abt Associates, under a competitive award; contract number TPDCFPBPA130014.

6. Conclusion

The CFPB Financial Well-Being Scale was built to measure what consumers say defines financial well-being to them: control over day-to-day, month-to-month finances; the capacity to absorb a financial shock; being on track to meet financial goals; and the financial freedom to make the choices that allow for the enjoyment of life. The data analysis conducted during the development of the scale model reinforced the idea that the items can be used to measure financial well-being as a cohesive general construct.

With this new free and reliable scale, it is now possible to measure and compare financial well-being across individuals and over time. The scale provides practitioners and researchers a way to understand financial well-being using a tool that was developed using large samples and rigorous statistical analyses. The end result of this work is a scale that can provide reliable financial well-being scores that have been supported by our initial validation exercises. These scores offer a complement to objective measures such as income, credit scores, and other common measures of financial situation, in that they were purposefully designed to reflect a respondent's holistic and subjective sense of their financial situation.

At the CFPB, we are using this measurement tool to continue our work to uncover what specific types of knowledge, skills, behavior, and attitudes tend to help people to improve their financial well-being, with the goal of helping move consumer financial capability-building practices forward in ways that can benefit American consumers.

We hope this tool will prove useful to others in their programmatic or policy work to support improvements in consumer financial well-being, or in kindred research efforts. We look forward to engaging with the broader field to collectively advance our shared understanding of how to support consumers in achieving their life goals and greater financial well-being.

7. References

Bagozzi, R. P. (2011). Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations. *MIS Quarterly, 35*(2), 261-292.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46(4)*, 443–459.

Bonifay, W. E. (2014). An illustration of the two-tier item factor analysis model. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling* (pp. 207-225). New York, NY: Routledge.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural models* (pp.136-162). Newbury Park, CA: Sage Publications.

Cai, L. (2013). flexMIRT® version 2: Flexible multilevel multidimensional item analysis and test scoring [computer software]. Chapel Hill, NC: Vector Psychometric Group, LLC.

Cai, L., & Hansen, M. (2013). Limited-information goodness-of-fit testing of hierarchical item factor models. *British Journal of Mathematical and Statistical Psychology*, *66*(2), 245–276.

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods, 16(3)*, 221-248.

Consumer Financial Protection Bureau. (2015). *Financial well-being: The goal of financial education.* Retrieved from consumerfinance.gov/reports/financial-well-being

Consumer Financial Protection Bureau. (2015). *Measuring financial well-being: A guide to using the CFPB Financial Well-Being Scale.* Retrieved from consumerfinance.gov/financial-well-being

Edwards, M. C. (2009). An introduction to item response theory using the Need for Cognition Scale*. Social and Personality Psychology Compass*, *3* (4), 507-529.

Embretson, S. E., & Reise, S. P. (2013.*) Item response theory (2nd ed.)*. Hove, UK: Psychology
   Press.

Gibbons, R. D., & Hedeker, D. (1992). Full information item bi-factor analysis*. Psychometrika,
   57*(3), 423-436.

Gibbons, R. D., Darrell, R. B., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D.
   J., Frank, E., Grochocinski, V. J., & Stover, A. (2007). Full-information item bifactor
   analysis of graded response data. *Applied Psychological Measurement, 31(1),* 4-19.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau,
   R. (2009) *Survey methodology* (2nd ed.)*.* Hoboken, NJ: John Wiley & Sons.

Knoll, M. A. Z., & Houts, C. R. (2012). The financial knowledge scale: An application of item
   response theory to the assessment of financial literacy. *Journal of Consumer Affairs*,
   46(3), 381–410.

National Research Council. (2013). *Subjective well-being: Measuring happiness, suffering, and
   other dimensions of experience. In A. A. Stone & C. Mackie (Eds.), Panel on measuring
   subjective well-being in a policy-relevant framework.* Committee on National Statistics,
   Division of Behavioral and Social Sciences and Education. Washington, DC: The
   National Academies Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.).* New York, NY:
   McGraw-Hill, Inc.

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of
   Educational Measurement*, *28*(3), 237-247.

Thissen, D., & Wainer, H. (Eds.). (2001). *Test scoring*. Mahwah, NJ: Erlbaum.

APPENDIX A:

Survey respondent demographics

TABLE 4:   DEMOGRAPHIC CHARACTERISTICS OF COMBINED ANALYSIS SAMPLE OF RESPONDENTS TO ALL THREE ROUNDS OF SCALE DEVELOPMENT SURVEYS

	Number	Percent
	10,804	100.00%
Sex		
Female	5,770	53.41%
Male	5,034	46.59%
Age Group		
18-24	1,336	12.37%
25-34	1,622	15.01%
35-44	1,725	15.97%
45-54	1,914	17.72%
55-61	1,329	12.30%
62-69	1,415	13.10%
70+	1,463	13.54%
Race		
Not reported	91	0.84%
White	7,780	72.01%
Black	1,072	9.92%
Asian	434	4.02%
Hispanic	929	8.60%
Other race	498	4.61%
Income		
Not reported	665	6.16%
Less than $20,000	1,511	13.99%
$20,000 - $29,000	1,213	11.23%
$30,000 - $39,999	1,303	12.06%
$40,000 - $49,999	1,039	9.62%

	Number	Percent
$50,000 - $59,999	1,000	9.26%
$60,000 - $74,999	1,099	10.17%
$75,000 - $99,999	1,302	12.05%
$100,000 - $149,999	975	9.02%
$150,000 or higher	574	5.31%
Don't know / Refused	123	1.14%
Education		
Not reported	261	2.42%
Some high school	327	3.03%
Completed high school	1,876	17.36%
Some college	3,048	28.21%
College degree	2,990	27.67%
Some post-graduate classes	731	6.77%
Master's degree	1,054	9.76%
Ph.D. or professional degree	458	4.24%

APPENDIX B:

Model and item parameters

Appendix B provides a description of the CFPB Financial Well-Being Scale model. First, we describe the scale model and how it was estimated. Then we discuss how to apply the estimated model to responses to produce a score, followed by a presentation of the item parameters from the estimated model. Instructions for how to prepare and score response data using the estimated item parameters via a custom Stata package are provided in Appendix F.

CFPB Financial Well-Being Scale model

The model that best fit the financial well-being item responses, based on the type of methods widely used in psychometrics and test scoring called Item Response Theory (IRT), is an "item bifactor" graded response model. The model is discussed by Gibbons et al. (2007) and Cai, Yang, and Hansen (2011). In this model, two conditions hold. First, each item "loads" on a general factor and also loads on one (and only one) additional factor (so each item is informed by exactly two factors – hence the "bi" of bifactor). Second, responses to each item are monotonic; meaning a higher level of the general factor raises the probability of higher responses (responses are ordered on a Likert scale).

The CFPB Financial Well-Being Scale has ten items with 5 potential ordered responses (0, 1, 2, 3, 4), with numerical values associated with the verbal labels described in Section 5.1 of the full report. For our specific bifactor model, there are three factors total. The first factor captures the respondent's general level of financial well-being and applies to all 10 items while the second and third factors are wording-specific (positive and negative, respectively) and apply only to items whose wording matches the direction of the appropriate secondary factor.

IRT Model Specification

The graded response model is a probabilistic model in which latent factors (i.e., unobserved person- and item-specific values) determine the probability of the answers respondents provide. An individual respondent is assumed to have a value for each latent factor in a vector of latent factors θ which are multiplied by item-specific coefficients α to form an index value θα that can be compared to cutoffs specific to each question to compute the probability of each given answer being observed, conditional on the factors.

That is, there are three types of parameters estimated in the model – thetas (i.e., individual-specific latent factor values or ability scores) and two types of item-specific parameters (i.e., cutoffs for each level of answer and multipliers/slopes related to the thetas). These item-specific parameters correspond, respectively, to the item severity parameters and degree of relatedness parameters described conceptually in the main document, and measure how much variation is produced in individual responses to items by the same amount of variation in latent factors. In the model, the primary factor theta zero, written θ_0, reflects the trait of interest, in this case, financial well-being. There are two additional thetas reflecting additional factors that are related to positively and negatively worded-items, respectively.

In the model, each item has parameter values, represented by alpha (α) which can be interpreted as the degree to which the item relates to the factor, also called slopes. For our bifactor model α_0 is the slope on the primary factor, θ_0, measuring financial well-being, and α_s is the slope on the additional factor, θ_s, for s= 1, or s= 2. For any item that has a non-zero slope on the first additional factor, that item's slope parameter relating it to the second additional factor must be zero, and vice versa.

The second type of item-specific parameter has to do with item location or intercept parameters. In the graded response model, each location parameter or cutoff represents the point on the multivariate theta distribution at which the probability of responding in category k or higher surpasses 0.50. The probability of a given response on the Likert scale is the difference between the two adjacent cumulative response probabilities. For an item with K response categories (in which K-1 is the highest possible response), this can be expressed as:

$$P(y = k|\,\theta_0, \theta_s\,) = \ P(y \geq k|\,\theta_0, \theta_s\,) - \ P(y \geq \ k + 1|\,\theta_0, \theta_s\,) \hspace{3cm} Eq.\ 1$$

$$where\ P(y \geq 0\ |\theta_0, \theta_s\,) = 1,$$

$$P(y \geq 1\ |\theta_0, \theta_s\,) = \frac{1}{1+\ e^{\{-d_1+\ \alpha_0\theta_0+\ \alpha_s\theta_s\}}},$$

$$\vdots$$

$$P(y \geq K - 1 | \theta_0, \theta_s) = \frac{1}{1 + e^{\{-d_{k-1} + \alpha_0 \theta_0 + \alpha_s \theta_s\}}},$$

$$P(y = K | \theta_0, \theta_s) = 0$$

and where, d_1 is the cutoff between the first and second response options indicating the level of the index value (sum of factors or thetas multiplied by coefficients alpha) below which the lowest response (coded 0, indicating the lowest level of financial well-being) has the highest probability of being chosen.

When the index value (the sum of latent factors or theta multiplied by coefficients alpha) exactly equals d_1 then the probability of answering 0 or more than 0 is one half. The cumulative logistic distribution function converts the sum of factors and the item-specific cutoffs into probabilities; these probabilities are used in a maximum likelihood estimator (described in the next section, and in Gibbons et al. 2007) to estimate cutoffs for each item, and individual-specific θ_0's.

Estimation was conducted by Vector Psychometric Group, LLC using flexMIRT® 2.0 (Cai, 2013), a commercially available IRT-software package, using the defaults of that program described in flexMIRT user's manual version 2.[37] Because differential item functioning (DIF)[38] was found across both age groups and survey modes, item- specific parameters were estimated for scoring four different types of scale respondents:

- Respondents ages 18-61 who self-administered the scale

- Respondents ages 18-61 who had the scale read to them by an interviewer

---

[37] "Flexible multilevel multidimensional item analysis and test scoring. Chapel Hill, NC: Vector Psychometric Group, available at: http://flexmirt.vpgcentral.com/downloads/flexmirtusermanual.pdf.

[38] Differential item functioning (DIF) occurs when groups (such as defined by age or mode of survey administration) have different probabilities of endorsing a given item on a multi-item scale after controlling for overall scale scores. An item is labeled as having DIF when people with the same level of a latent trait but from different groups have an unequal probability of giving a response.

- Respondents ages 62 and older who self-administered the scale

- Respondents ages 62 and older who had the scale read to them by an interviewer

The program assumes a normal distribution for each latent factor. Respondents ages 18-61 who self-administered the scale were used as the reference population and assumed to have standard normal latent factors in the population (that is, latent factor means of 0 and standard deviations of 1); the means and standard deviations of other groups' latent factors are estimated, relative to that group, by the software.

Applying the Model

The probabilities of observing a response conditional on a set of factor values, calculated using Equation 1 and using the estimated item parameters in Table 5, below, are used to score responses using the estimated model. The method of scoring proceeds by computing the probability of an observed response pattern (the product of all item-specific probabilities, given estimated parameters) weighted by a factor value θ_0 and the prior probability of that factor value $f(\theta_0, \theta_1, \theta_2)$ where f() is the multivariate standard normal density. This value is summed over all possible factor values θ_0 and divided by the sum over all factor values θ_0 of the prior probability of that factor value $f(\theta_0, \theta_1, \theta_2)$ times the probability of the response pattern. That is, the integrals capturing the "true expected a posteriori" (TEAP) scores, assuming infinitely many calculations are possible,

$$\text{TEAP} = \frac{\iiint \theta_0 \left[\prod_{j=1}^{10} P(y_{ji}=k_{ji}|\theta_0,\theta_1,\theta_2)\right] f(\theta_0,\theta_1,\theta_2) d\theta_0 d\theta_1 d\theta_2}{\iiint \left[\prod_{j=1}^{10} P(y_{ji}=k_{ji}|\theta_0,\theta_1,\theta_2)\right] f(\theta_0,\theta_1,\theta_2) d\theta_0 d\theta_1 d\theta_2} \qquad \textit{Eq. 2}$$

(where the P functions are given by Equation 1 above using estimated parameters α and cutoffs d in Table 5) for each person *i* are approximated by sums

$$\text{EAP} = \frac{\sum_{\theta_0=\theta_L}^{\theta_0=\theta_H} \sum_{\theta_{Pos}=\theta_L}^{\theta_{Pos}=\theta_H} \sum_{\theta_{Neg}=\theta_L}^{\theta_{Neg}=\theta_L} \theta_0 \left[\prod_{j=1}^{10} P(y_{ji}=k_{ji}|\theta_0,\theta_1,\theta_2)\right] f(\theta_0,\theta_1,\theta_2) d\theta_0 d\theta_1 d\theta_2}{\sum_{\theta_0=\theta_L}^{\theta_0=\theta_H} \sum_{\theta_{Pos}=\theta_L}^{\theta_{Pos}=\theta_H} \sum_{\theta_{Neg}=\theta_L}^{\theta_{Neg}=\theta_L} \left[\prod_{j=1}^{10} P(y_{ji}=k_{ji}|\theta_0,\theta_1,\theta_2)\right] f(\theta_0,\theta_1,\theta_2) d\theta_0 d\theta_1 d\theta_2} \qquad \textit{Eq. 3}$$

taken over discrete steps of size dθ from θ_L to θ_H, to get the estimated EAP. This numerical integration (commonly known as fixed-point quadrature) is done to ensure compatibility with

scoring procedures used in flexMIRT® (other kinds of numerical integration procedures will give slightly different estimated EAP values). This process is depicted for a single dimension in figure 2 (setting the specific factor to zero for simplicity of exposition).

Each "step" in the sum is the area of the rectangle of height f times the width of the step $d\theta$ (divided by the sum of all rectangles) multiplied by the product of all probabilities P: the product of probabilities is the chance of seeing the observed response pattern conditional on that value of θ. This is a Riemann integral approximated by numerical integration that gets arbitrarily close to the true value of the integral as the number of steps increases (i.e. as $d\theta$ gets smaller) and θ_L and θ_H increase in absolute magnitude. In both the integrals (Eq. 2) and the sums (Eq. 3), the probability P for any refused, don't know, or otherwise missing response is set to one in the calculations, so those answers contribute no information (since multiplying by one does not change the overall product).[39]

As an illustrative example of how the software calculates an EAP "theta" score for the calculation illustrated in Figure 2, below, assume a situation in which a person aged 62 or older surveyed by phone (Group 1, in Table 5, below) answers one question only. The prior density is a simple normal curve, shown in the top left panel of figure 2. Suppose a person responds only to "I could handle a major unexpected expense" with "Describes me very well" or a 3 in numerical terms (and for simplicity fix the specific factor to be zero for the one calculation), meaning that 2.205 times θ is likely to lie between 1.088 and 2.119, per the first row in Table 5. That is, the probability of answering 3 or lower (rather than 4) is one half when the sum of factors is 2.119 divided by 2.205, and the probability of answering 2 or lower (rather than 3 or higher) is one half when the sum of factors is 1.088 divided by 2.205, meaning probabilities of observing a response of 3 are highest between about one half and one, as shown in the top right panel of Figure 2.

The convolution, or the distribution of probabilities of the observed response 3 multiplied by prior densities, is shown in the bottom left panel of Figure 2. The sum of the convolution of P, f,

---

[39] Note that different estimated scores can be obtained for different numerical approximations to the triple integrals shown, but for large numbers of steps from extreme values of the underlying normal distribution, small changes do not change computed scores within rounding error.

and θ in the final panel (bottom right panel of figure 2) is divided by the sum of the convolution of P and f (bottom left) to get the EAP theta of .675 which is multiplied by 15 and added to 50 to get the final CFPB Financial Well-Being Scale score of 60. The software does this calculation for each value of the specific factor, and each question answered, to construct the sum over all items, to get EAP scores and the final CFPB Financial Well-Being Scale scores.

FIGURE 2:   ILLUSTRATIVE CALCULATION OF EAP "THETA" SCORES
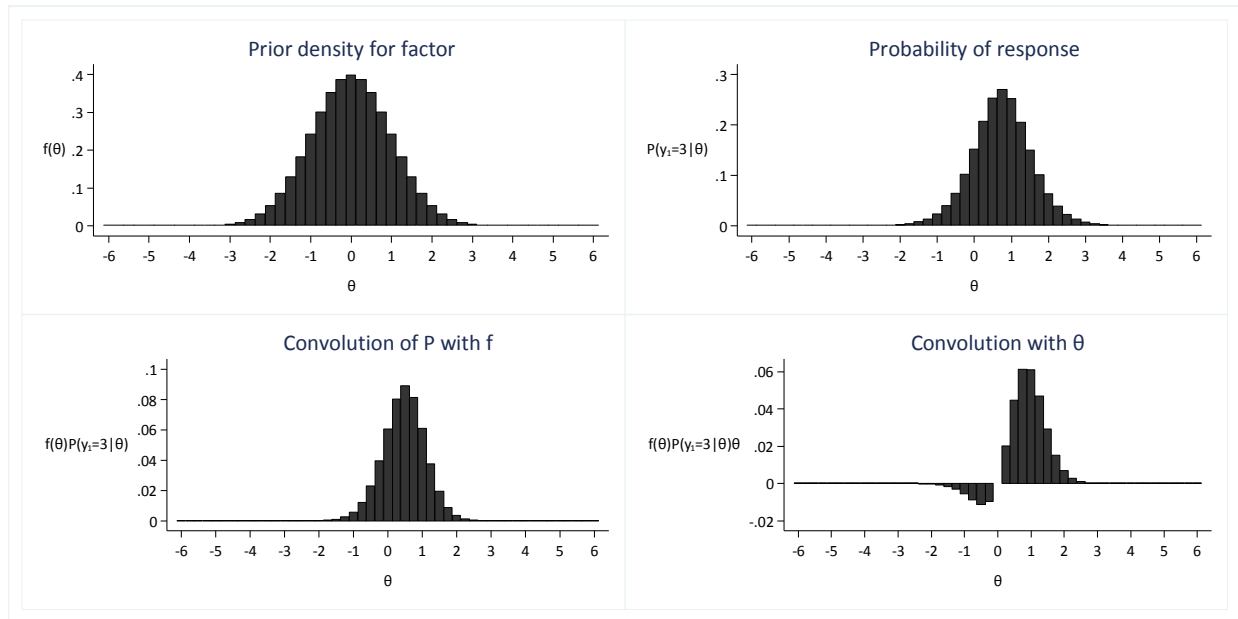
Note: This example assumes an older person surveyed by phone (group 1) answers one question only, to illustrate steps in calculating EAP scores.

Item parameters

The estimated item parameters used for scoring the model are shown in Table 5. The three alphas are α_0 (the estimated slope parameter for factor θ_0 or financial well-being as shown in Eq. 1), α_1 and α_2 (the estimated slope parameters for the specific factors θ_s, for s= 1, or s= 2). The other values in the table are the item intercept parameters where *d4* the cutoff for the highest Likert scale response (4) and the next 3 items are the cutoffs (*d3, d2, d1*) for each subsequent response.

The parameters estimated and shown in the table are used to estimate an *expected a posteriori* (EAP) score (Equation 2), which is a latent maximum likelihood mean factor estimate based on

probabilities computed using the estimated cut points for each categorical response (Equation 3 above). For example, d_4 is the point at which a factor sum ($\alpha_0\theta_0 + \alpha_1\theta_1 + \alpha_2\theta_2$, where we always take one of the item-specific alphas α_1, α_2 to be zero for the specific factor that does not apply to the item) yields a 50-50 chance of selecting 4 versus 3, 2, 1, or 0. Likewise, d_3 is the factor sum that yields a 50-50 chance of selecting 4 or 3 versus 2, 1, or 0, and so on.

The estimated parameters shown below in Table 5 include all combinations of respondent age groups and data collection modes, across all 10 items. These parameters are the basis of the scoring lookup tables shown in Appendix E, and the scoring program available for Stata, described in Appendix F, using the formulas shown in Eq. 3.

TABLE 5: IRT ESTIMATED PARAMETERS BY AGE GROUP AND MODE

AGE 62+, INTERVIEWER-ADMINISTERED (GROUP=1)

Item	α_0	d_4	d_3	d_2	d_1	α_1	α_2
I could handle a major unexpected expense	2.205	2.199	1.088	-0.359	-2.574	2.598	0
I am just getting by financially	1.321	2.026	0.274	-0.862	-1.937	0	-0.296
I am securing my financial future	2.214	1.659	1.059	-0.423	-2.144	1.956	0
I am concerned that the money I have or will save won't last	2.221	1.923	0.425	-1.445	-3.36	0	-0.998
Because of my money situation, I feel like I will never have the things I want in life	3.094	2.514	1.309	-0.445	-2.475	0	-0.573
I can enjoy life because of the way I'm managing my money	3.840	2.147	2.149	0.169	-1.735	2.391	0
I am behind with my finances	2.093	2.115	1.096	0.322	-0.357	0	0.522
My finances control my life	2.142	1.497	1.064	-0.33	-1.634	0	0.159
Giving a gift for a wedding, birthday or other occasion would put a strain on my finances for the month	2.062	2.872	1.111	-0.268	-1.535	0	0.347
I have money left over at the end of the month	1.745	1.526	0.756	-0.378	-1.484	1.463	0

AGE 62+, SELF-ADMINISTERED (GROUP=2)

Item	α_0	d_4	d_3	d_2	d_1	α_1	α_2
I could handle a major unexpected expense	2.418	2.578	0.412	-2.041	-4.774	2.598	0
I am just getting by financially	2.233	2.026	0.568	-0.997	-2.623	0	-0.296
I am securing my financial future	2.665	2.284	0.849	-1.694	-4.548	1.956	0
I am concerned that the money I have or will save won't last	2.221	1.923	0.425	-1.445	-3.36	0	-0.998
Because of my money situation, I feel like I will never have the things I want in life	3.094	2.514	1.309	-0.445	-2.475	0	-0.573
I can enjoy life because of the way I'm managing my money	3.508	2.147	1.228	-1.515	-4.323	2.391	0
I am behind with my finances	3.413	2.115	1.807	0.16	-1.778	0	0.522
My finances control my life	2.862	1.995	1.049	-1.026	-3.109	0	0.159
Giving a gift for a wedding, birthday or other occasion would put a strain on my finances for the month	3.691	2.872	1.594	-0.638	-3.194	0	0.347
I have money left over at the end of the month	3.203	1.972	1.069	-0.97	-2.902	1.463	0

AGE 18-61, INTERVIEWER-ADMINISTERED (GROUP=3)

Item	α_0	d_4	d_3	d_2	d_1	α_1	α_2
I could handle a major unexpected expense	1.205	2.199	-0.259	-1.995	-3.736	2.598	0
I am just getting by financially	2.233	2.026	0.568	-0.997	-2.623	0	-0.296
I am securing my financial future	3.155	1.659	1.253	-0.946	-3.222	1.956	0
I am concerned that the money I have or will save won't last	2.221	1.923	0.425	-1.445	-3.36	0	-0.998
Because of my money situation, I feel like I will never have the things I want in life	3.094	2.514	1.309	-0.445	-2.475	0	-0.573

Item	α_0	d_4	d_3	d_2	d_1	α_1	α_2
I can enjoy life because of the way I'm managing my money	2.983	2.147	1.29	-1.231	-3.455	2.391	0
I am behind with my finances	2.023	2.115	1.131	-0.126	-1.437	0	0.522
My finances control my life	1.238	1.542	0.4	-1.057	-2.296	0	0.159
Giving a gift for a wedding, birthday or other occasion would put a strain on my finances for the month	2.465	2.872	1.528	-0.081	-1.802	0	0.347
I have money left over at the end of the month	1.571	1.972	0.628	-0.818	-2.236	1.463	0

AGE 18-61, SELF-ADMINISTERED (GROUP=4)

Item	α_0	d_4	d_3	d_2	d_1	α_1	α_2
I could handle a major unexpected expense	2.418	2.199	0.412	-2.041	-4.774	2.598	0
I am just getting by financially	2.233	2.026	0.568	-0.997	-2.623	0	-0.296
I am securing my financial future	3.155	1.659	1.253	-0.946	-3.222	1.956	0
I am concerned that the money I have or will save won't last	2.221	1.923	0.425	-1.445	-3.36	0	-0.998
Because of my money situation, I feel like I will never have the things I want in life	3.094	2.514	1.309	-0.445	-2.475	0	-0.573
I can enjoy life because of the way I'm managing my money	3.508	2.147	1.228	-1.515	-4.323	2.391	0
I am behind with my finances	3.413	2.115	1.807	0.16	-1.778	0	0.522
My finances control my life	2.862	1.995	1.049	-1.026	-3.109	0	0.159
Giving a gift for a wedding, birthday or other occasion would put a strain on my finances for the month	3.691	2.872	1.594	-0.638	-3.194	0	0.347
I have money left over at the end of the month	3.203	1.972	1.069	-0.97	-2.902	1.463	0

Source: Vector Psychometric Group. MIRT parameters using bi-factor graded response model using flexMIRT® software.

APPENDIX C:

Reliability statistics

Marginal reliability statistics describe the percentage likelihood that the differences in scores produced by the scale are due to true or actual variance in the respondents' financial well-being, rather than to statistical error. In other words, marginal reliability in IRT is the ratio of true score variance to observed score variance.[40]

Marginal reliability statistics can range from 0 to 1, where higher values indicate more reliable scores, similar to Cronbach's alpha in classical test theory. The minimum standard for a scale to be considered generally credible is .70,[41] so the CFPB Financial Well-Being Scale, at or above .80, can be considered highly reliable for all combinations of respondent age, survey mode, and scale version.

TABLE 6:   MARGINAL RELIABILITY BETWEEN AGE GROUPS AND MODES FOR CFPB FINANCIAL WELL-BEING SCALE

	Standard 10 Item Scale		5 Item Scale	
Mode	Age 18-61	Age 62+	Age 18-61	Age 62+
Online	0.89	0.90	0.82	0.83
Phone	0.90	0.89	0.84	0.82

---

[40] Sireci, S.G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237-247.

[41] Nunnally, J. & Bernstein, I. (1994) Psychometric Theory. New York: McGraw Hill, 3rd ed.

APPENDIX D:

Validation statistics

The following table provides information on the relationship between the CFPB Financial Well-Being Scale score and select validation items collected during the scale development process.

TABLE 7: SPEARMAN'S CORRELATIONS BETWEEN 10 ITEM IRT SCORES AND VALIDATION ITEMS

Variable	Mean	(SD)	N	FWB Spearman Corr.
How would you rate your current credit record?	3.86	1.21	6,854	0.53***
(1) Very bad to (5) Very good				
How confident are you that you could come up with $2,000 if an unexpected need arose within the next month?	2.93	1.18	7,030	0.64***
(1) I am certain I could not come up with $2,000				
(2) I could probably not come up with $2,000				
(3) I could probably come up with $2,000				
(4) I am certain I could come up with the full $2,000				
Have you set aside funds that would cover your expenses for 3 months if you or someone in your family lost a job, got sick or had another emergency?	0.49	N/A	7,310	0.54***
(1) No (2) Yes				
On a scale of 1 to 7 where 1 means "Extremely bad" and 7 means "Extremely good" how would you rate your current financial situation?	4.72	1.61	7,222	0.63***
(1) Extremely bad to (7) Extremely good				
There has been a time in the last year when I needed food but couldn't afford to buy it.	0.18	N/A	10,166	-0.43***
(1) No (2) Yes				
In the past 12 months, there was at least one time when I or someone in my household needed to see a doctor or go to the hospital but did not go because we couldn't afford it.	0.14	N/A	10,166	-0.35***
(1) No (2) Yes				

Variable	Mean	(SD)	N	FWB Spearman Corr.
Had a debt in collections	0.09	N/A	1,218	-0.30***
(1) No (2) Yes				
Count of negative financial shocks experienced in past 12 months† (out of 10 possible options)	0.53	0.83	10,166	-0.31***

Notes: *** = p < .0001.

N varies per item because different validation items were fielded in different rounds of data collection.

† Calculated as the number of events selected by respondents in response to the following question: In the past 12 months, did you or your family living with you experience any of the following events? (1) I lost a job (2) I had my work hours and/or pay reduced (3) My spouse/partner lost a job (4) My spouse/partner had their work hours and/or pay reduced (5) Received a foreclosure notice (6) A business I owned had financial difficulty (7) Had a health emergency (8) Divorce (9) Had a baby (10) Death of primary breadwinner.

In addition to these validators, the following table provides information on the relationship between the CFPB Financial Well-Being Scale score and education and income. As shown below, the scale score is positively correlated with income and education.

TABLE 8:   SPEARMAN'S CORRELATIONS BETWEEN 10 ITEM EAP SCORES AND SOCIO-ECONOMIC DATA

Variable	Mean	(SD)	N	FWB Spearman Corr.
Annual Household Income [midpoint of bins]	62,948	48,013	9,378	0.38***
Less than $20,000 [$10,000]				
$20,000 - $29,999				
$30,000 - $39,999				
$40,000 - $49,999				
$50,000 - $59,999				
$60,000 - $74,999				
$75,000 - $99,999				
$100,000 - $149,999				
$150,000 or higher [$200,000]				
Education	3.71	1.42	9,846	0.19***
(1) Some high school to (7) PhD or professional degree				

APPENDIX E:

CFPB Financial Well-Being Scale scoring worksheets

As a convenience to researchers and practitioners who do not have access to IRT software, or prefer not to use the Stata scoring program described in Appendix F, the research team has developed a table-based scoring procedure which approximates IRT procedures by providing the IRT-based score for the most likely answer pattern for any given raw score. This appendix contains both the standard version and the abbreviated version of the CFPB Financial Well-Being Scale scoring worksheets needed to employ this scoring method. For more information on this scoring method and guidelines about its use, see Section 5.2.2.

Standard version scoring worksheet

CFPB FINANCIAL WELL-BEING SCALE
Scoring worksheet

1. Select the person's answers, record the response value in the right hand column and add up the total values for each part of the questionnaire.

This statement describes me	Completely	Very well	Somewhat	Very little	Not at all	Response value
1. I could handle a major unexpected expense	4	3	2	1	0	
2. I am securing my financial future	4	3	2	1	0	
3. Because of my money situation, I feel like I will never have the things I want in life	0	1	2	3	4	
4. I can enjoy life because of the way I'm managing my money	4	3	2	1	0	
5. I am just getting by financially	0	1	2	3	4	
6. I am concerned that the money I have or will save won't last	0	1	2	3	4	

Part 1 subtotal: _____

This statement applies to me	Always	Often	Sometimes	Rarely	Never	Response value
7. Giving a gift for a wedding, birthday or other occasion would put a strain on my finances for the month	0	1	2	3	4	
8. I have money left over at the end of the month	4	3	2	1	0	
9. I am behind with my finances	0	1	2	3	4	
10. My finances control my life	0	1	2	3	4	

Part 2 subtotal: _____

Total response value: _____

2. Find the financial well-being score

How old is the person?
☐ 18-61   ☐ 62+

How did the person take the questionnaire?
☐ Self-administered
☐ Administered by someone else

Because scores vary based on age and how the questionnaire was administered, you must convert the total response value to a financial well-being score.

a. Find the row that corresponds to the total response value.

b. Follow that row across to the column that corresponds to the person's age and how the questionnaire was administered.

c. Record the final score.

Financial well-being score: _____

Learn more at
consumerfinance.gov/financial-well-being

cfpb Consumer Financial Protection Bureau

Total response value	Questionnaire self-administered		Questionnaire administered by someone else	
	18-61	62+	18-61	62+
0	14	14	16	18
1	19	20	21	23
2	22	24	24	26
3	25	26	27	28
4	27	29	29	30
5	29	31	31	32
6	31	33	33	33
7	32	35	34	35
8	34	36	36	36
9	35	38	38	38
10	37	39	39	39
11	38	41	40	40
12	40	42	42	41
13	41	44	43	43
14	42	45	44	44
15	44	46	45	45
16	45	48	47	46
17	46	49	48	47
18	47	50	49	48
19	49	52	50	49
20	50	53	52	50
21	51	54	53	52
22	52	56	54	53
23	54	57	55	54
24	55	58	57	55
25	56	60	58	56
26	58	61	59	57
27	59	63	60	58
28	60	64	62	60
29	62	66	63	61
30	63	67	65	62
31	65	69	66	64
32	66	71	68	65
33	68	73	70	67
34	69	75	71	68
35	71	77	73	70
36	73	79	76	72
37	75	82	78	75
38	78	84	81	77
39	81	88	85	81
40	86	95	91	87

Abbreviated version scoring worksheet

CFPB FINANCIAL WELL-BEING SCALE
Scoring worksheet

1. Select the person's answers, record the response value in the right hand column and add up the total values for each part of the questionnaire.

This statement describes me	Completely	Very well	Somewhat	Very little	Not at all	Response value
1. Because of my money situation, I feel like I will never have the things I want in life	0	1	2	3	4	
2. I am just getting by financially	0	1	2	3	4	
3. I am concerned that the money I have or will save won't last	0	1	2	3	4	

Part 1 subtotal: _____

This statement applies to me	Always	Often	Sometimes	Rarely	Never	Response value
4. I have money left over at the end of the month	4	3	2	1	0	
5. My finances control my life	0	1	2	3	4	

Part 2 subtotal: _____

Total response value: _____

2. Find the financial well-being score

How old is the person?
☐ 18-61　　☐ 62+

How did the person take the questionnaire?
☐ Self-administered
☐ Administered by someone else

Because scores vary based on age and how the questionnaire was administered, you must convert the total response value to a financial well-being score.

a. Find the row that corresponds to the total response value.

b. Follow that row across to the column that corresponds to the person's age and how the questionnaire was administered.

c. Record the final score.

Financial well-being score: _____

Total response value	Questionnaire self-administered		Questionnaire administered by someone else	
	18-61	62+	18-61	62+
0	19	20	22	24
1	25	26	30	30
2	29	31	33	33
3	32	34	36	37
4	36	37	39	39
5	38	40	42	42
6	41	43	44	44
7	43	46	47	46
8	46	48	49	49
9	48	51	51	51
10	50	53	54	53
11	53	55	56	55
12	55	58	58	57
13	57	61	60	59
14	60	63	63	62
15	63	66	66	64
16	65	69	68	67
17	68	73	72	70
18	72	76	75	73
19	76	81	80	77
20	82	90	87	84

Learn more at
consumerfinance.gov/financial-well-being

cfpb Consumer Financial Protection Bureau

APPENDIX F:

Instructions for software-based IRT scoring using Stata

This appendix provides instructions for applying the software-based IRT scoring method for the CFPB Financial Well-Being Scale using a custom Stata package (`pfwb.ado`) developed for the CFPB by Abt Associates. It uses the parameter estimates reported in Appendix B Table 5. Researchers who would like to use this Stata program to directly estimate CFPB Financial Well-Being Scale scores should take the following steps to prepare and score their data.

Preparing your data

Step 1: Name your variables.

This is perhaps the most important of the steps. Please pay special attention to providing your variables with the proper names. Variables capturing responses to the 10 financial well-being items in your data will need to be renamed according to Table 9. Table 9 details each item, mapping variable name (Item name) to question wording (Item). Ensure that the names of each item are set to reflect the variables named as in the "Item name" column. Items that are included in the 5-item short scale are indicated. The same parameters are used in the 5-item scale as in the 10-item (standard) scale.

TABLE 9:   ITEM DESCRIPTIONS

Item name	Item	Reverse coded	In 5 item scale	Response type
fwb1_exp	I could handle a major unexpected expense	No		Describe
fwb2_getby	I am just getting by financially	Yes	Yes	Describe
fwb3_secure	I am securing my financial future	No		Describe
fwb4_concern	I am concerned that the money I have or	Yes	Yes	Describe

Item name	Item	Reverse coded	In 5 item scale	Response type
	will save won't last			
fwb5_never	Because of my money situation, I feel like I will never have the things I want in life	Yes	Yes	Describe
fwb6_enjoy	I can enjoy life because of the way I'm managing my money	No		Describe
fwb7_behind	I am behind with my finances	Yes		Frequency
fwb8_control	My finances control my life	Yes	Yes	Frequency
fwb9_strain	Giving a gift for a wedding, birthday or other occasion would put a strain on my finances for the month	Yes		Frequency
fwb10_left	I have money left over at the end of the month	No	Yes	Frequency

Note: Responses for "Describe" items: 4=Describes me completely, 3=Describes me very well, 2=Describes me somewhat, 1=Describes me very little, 0=Does not describe me at all;

Responses for "Frequency items": 4=Always, 3=Often, 2=Sometimes, 1=Rarely, 0=Never.

Step 2: Set missing values to missing.

While the financial well-being scales do not include refused or "Don't know" responses, you may still encounter missing or non-substantive data. If this is the case, you will need to recode the data to indicate that it is missing. In Stata, this would mean coding those responses to "." to indicate missing.

Step 3: Recode responses.

Before you can score the data, you must ensure that responses are coded as 0 to 4 where 4 is always the most positive financial well-being response. For example, with "My finances control my life," a response of "Never" should be coded as 4 and a response of "Always" should be coded as 0. With "I am securing my financial future," a response of "Completely" should be coded as 4 and a response of "Not at all" should be coded as 0. The column in Table 9 labeled "Reverse coded" will help you in recoding your items.

The Stata program should exit with an error code if the items are not recoded properly. To determine whether items are coded properly, the program checks average correlations across negatively and positively worded items to ensure that the correlation is positive on average. [42]

Step 4: Create the indicators for your group codes: mode of administration ("self") and age group ("age18_61).

You will need to create the two indicators used by the program to define the groups:

- Indicator 1: "age18_61" – set this indicator to 1 if the respondent is ages 18 to 61 and to 0 if the respondent is ages 62 or older
- Indicator 2: "self" – set this indicator to 1 if the survey was self-administered and to 0 if the survey was completed in an interview.

This is also a crucially important step. Correct group codes ensure that the proper parameters are applied in estimating your scores. To determine which parameters to use in scoring each case, the program creates an indicator of the group to which each individual response belongs. These groups are determined based on age [18 to 61 year versus 62 plus years] and mode of survey administration [self-administered or interview]. Depending on your particular sample and mode of administration, you will have between 1 and 4 groups: 1: 62+, Interview; 2: 62+, Self-Administered; 3: 18-61, Interview; 4: 18-61, Self-Administered.

Using the Stata program to generate scores

Step1: Find and install Stata scoring program.

Using Stata, one can simply type "`ssc install pfwb`" to install a program that will score the items listed in Table 9 as long as the data are prepared as described above. One can also type "findit pfwb" to search for alternate download locations.

---

[42] This does not guarantee that the scoring has been done correctly, since data can exhibit some pathological correlations, or one variable might be wrong while average correlations are correct, but should prevent the most egregious scoring errors where no survey response have been corrected for reverse coding.

Step 2: Generate the CFPB Financial Well-Being Scale score.

After preparing the data as described above, simply type "`pfwb fwb`" to generate a new variable `fwb` which is the respondent's CFPB Financial Well-Being Scale score. The Stata help file also explains these steps; simply type "`help pfwb`" after installing the program to read the help file.