CONSUMER FINANCIAL PROTECTION BUREAU
INFORMATION COLLECTION REQUEST – SUPPORTING STATEMENT
PART B
DEVELOPMENT OF METRICS TO MEASURE FINANCIAL WELL-BEING OF
WORKING-AGE AND OLDER AMERICAN CONSUMERS
(OMB CONTROL NUMBER: 3170-XXXX)

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Selection Methods

The Consumer Financial Protection Bureau has contracted with CFED to collect data to support the development of metrics to measure financial well-being and financial ability. CFED has, in turn, subcontracted with Vector Psychometric Group, LLC (VPG) to conduct the psychometric analysis in support of the development of these metrics. VPG uses methods based on item response theory (IRT) in its metric development process. IRT-based methods are not probabilistic—they do not assume that the data being analyzed are from a random or statistically representative sample. IRT-based analysis does however require a large, socio-demographically diverse sample. Large, diverse samples help ensure stable parameters around a wide-range of possible item responses and enables analysts to test the performance of the metrics in different subgroups of interest.

CFED and its subcontractors will work with Survey Sampling International (SSI) to collect data from working-age and older Americans that are sufficiently large and socio-demographically diverse. [1] Specifically, CFED and its subcontractors will provide SSI with socio-demographic recruitment targets. The 2010 Census with respect to income, education, employment status, marital status, age, gender, race/ethnicity, presence/ages of children and geography will be used as a guideline for recruitment targets when possible to help ensure socio-demographic diversity. SSI will recruit until the specified socio-demographic targets are met and then they will close the project to those subgroups. SSI is a global company and one of the largest providers of sampling, data collection and data analytics services. SSI adheres to World Association for Social, Opinion and Market Research (ESOMAR) standards and is Grand Mean audited through Sample Source Auditors and annually audited by Ernst & Young. In 2012, SSI successfully completed 29 million surveys/questionnaires in 78 countries.

SSI will collect for this project **completed** questionnaires from 14,300 Americans (7,800 working-age and 6,500 older Americans), drawn from its opt-in panel of Americans who are age 18 and older and have access to the internet from some location. In the future and with independent OMB review and approval, the CFPB may field the questionnaire developed via this requested information collection to a randomly selected, representative sample of the US

---

[1] 'Working-age Americans' include Americans ages 18-61 and 'older Americans' includes Americans ages 62 and older.

population in order to examine the instruments' performance in such a population.

SSI's opt-in convenience sampling frame (the persons who have the potential to be part of the sample that SSI will collect for this study), includes "the unique audience across a set of SSI web "properties" [generally online advertisements] who notice the invitation to join and are motivated to act on it."[2] SSI asks persons who respond to their invitations to complete a brief online questionnaire that includes socio-demographic questions. After receiving the responses to this questionnaire, SSI utilizes a suite of quality control procedures including digital fingerprinting, to avoid duplication. Additionally, SSI samples incoming participants with its blending questionnaire, which asks behavior and psychographic questions in order to monitor the consistency of the sample coming into the system.

The sampling frames, for the purposes of this study, include working-age Americans (US residents ages 18 to 61) and older Americans (ages 62 and older) who are actively participating in an SSI panel. From these sample frames, SSI will develop "out go" or contact groups. Persons in the "out go" group will receive invitations to take the questionnaires being fielded as part of this project. The contact group will be sufficiently diverse to ensure that the respondents meet the target demographics. Because some individuals are more likely than others to respond to the participation invitation, the contact group will likely resemble the target demographics to a lesser degree than the group of participants completing the questionnaires. This is because SSI takes into account the fact that the take-up rate varies by target group. For example, the contact group is likely to contain a much larger number of males ages 18 to 24 than females ages 45 to 54 given the differences in each group's likelihood of responding (the older females being more likely to respond than the younger males). SSI monitors participation rates within the target demographics (age, education, income, gender, race/ethnicity, geography) and adjusts the "out go" as necessary to ensure the resulting data set meets the target demographics.

Following the psychometric analysis of the item-level data, VPG will compare online, paper-and-pencil as well as phone versions of the questionnaire using a third data set provided by SSI. The analyses performed using this data set will focus on the stability of the item parameters across these different modes.

Sample	Population universe	SSI sampling frame	Persons receiving an invitation from SSI to participate in the survey	Number of SSI-fielded completed surveys
Working-age Americans	~140 million individuals in the US ages 18 and older who access the internet from some location	1.02 million individuals ages 18 and older participating in SSI panels	Approximately 280,000 SSI panel participants	7,800 SSI panel participants ages 18-61

---

[2] See "Mixing the Right Sample Ingredients: A New Source Recipe." Pete Cape, Kristin Cavallaro and Jackie Lorch. Survey Sampling International Memo.

Older Americans	~23 million individuals in the US ages 62 and older who access the internet from some location	180,000 individuals ages 62 and older participating in SSI panels	Approximately 120,000 SSI panel participants	6,500 SSI panel participants ages 62 and older

2. Information Collection Procedures

VPG relies on IRT-based methods in its metrics development process. IRT-based methods are not probabilistic—they do not assume that the survey data being analyzed are from a random sample. IRT-based analysis does however require a large, socio-demographically diverse sample. Large, diverse samples help ensure stable parameters around a wide-range of possible survey responses and enable analysts to test the performance of the metrics in different subgroups (for example, low-, middle- and high-income groups).

The contractor will work with SSI to obtain data from 7,800 working-age Americans and 6,500 older Americans whose socio-demographic profiles meet the specified demographic targets.

No population inferences will be made using this data.

3. Methods to Maximize Response Rates and Address Issues of Non-Response

Survey respondents will come from existing panels managed by Survey Sampling International (SSI). Agreements regarding incentive payments were established between SSI and the panelist at the time s/he was recruited. For example, when signing up for *OpinionWorld*, a panel administered by SSI, respondents are able to choose from a variety of incentives including donations to a preferred charity, entries into a drawing for a monetary prize, etc. The incentives respondents will receive are not specific to this survey. SSI offers a wide variety of incentives in order to increase diversity of its sample frames because different types of rewards will motivate different respondents to participate in a survey.

For a questionnaire of this length and type, a typical rate of return in convenience samples, defined as opening the email and completing the questionnaire, is roughly 3 to 5 percent. This rate of return is typical of online panels. The participation rate or the percent of panelists entering the project (i.e., clicked on the link and made it to the welcome and introduction) who completed the questionnaire is roughly 80 percent for SSI panels.

High return and participation rates for this study will be encouraged through several means, including carefully worded and market research tested invitation emails, thoughtful planning of the participant experience that capitalizes on the latest survey research (for example, ensuring that questionnaires conducted as part of the panel are short enough to completed before respondent fatigue sets in), and follow-up/reminder emails to non-respondents. In addition, SSI has a strong set of incentives designed to appeal to respondents across demographic and behavioral groups to ensure participation without biasing results. Ensuring a quality participant experience is central to SSI's survey recruitment strategy.

SSI will also try to maximize rate of return by emphasizing the importance of this study in its participation invitation emails. SSI will also send a reminder email to non-respondents offering them second and third opportunity to participate.

4. <u>Testing of Procedures or Methods</u>

The entire purpose and planned process for this information collection is to test and refine item wording and answer sets as well as evaluate modes of administration (online vs. telephone vs. paper and pencil). Using responses from three distinct rounds of data collection, the contractor will develop metrics to assess two constructs of interest (i.e., Financial Well- and Financial Ability) using modern psychometric methods (e.g., Netemeyer et al. 2003; Thissen and Wainer 2001; Wirth and Edwards 2007). A large pool of candidate items (questions and response sets) based on the finding from qualitative research, the project literature review, and a review of other financially related questionnaires was developed (see attached *Item Bank* document for an initial pool of questions on all anticipated topics). A panel of psychometric and content experts reviewed and refined the pool of candidate items so that the remaining items are targeted and well-worded. Cognitive interviewing of non-experts will further serve to refine the item pool and items. No more than 120 items (defined as a statement, typically one sentence, to which the participants will express their level of agreement) will be included in the refined item pool. This number of items is specified to ensure that respondents are not overly burdened when responding to the item pool.

In round I, SSI will field the candidate items online to approximately 2,500 working-age and 2,000 older Americans. A preliminary round of data will be collected to ensure there are no substantial problems with the items, as described in part A of the Supporting Statement. When the main data from the first round is received, it will be thoroughly investigated for each age-group individually and any necessary recoding or flagging of questionable response patterns (typically called "data cleaning"). Data cleaning efforts include, but are not limited to, identifying suspect response patterns (e.g., participants provides the same response for every item), detailing any missing data occurrences which may be problematic at later stages, checking for non-logical values (e.g., logically inconsistent responses across items), and reverse coding items as needed to ensure responses to all items are in the same direction (i.e., more agreement indicates a higher level of the construct of interest). With data cleaning complete, preliminary data analysis will begin.

Preliminary data analysis, by age group (working-age vs. older), will be conducted. For each item from the pool of candidate items, a frequency table of responses (and possibly a histogram which depicts the same information graphically) will be obtained and examined. Additionally, item-by-item 2-way contingency tables and a polychoric correlation matrix will be obtained. These will be examined to identify any item pairs in which responses are not related as expected. Classical test theory analyses (Cronbach's alpha, item total correlations, etc.) will then be conducted. Although these analyses will not be used in making final determinations of item inclusion on the final version of the instruments, such analyses provide recognizable results for those less familiar with modern scale development techniques and will provide a bridge to understanding the results of the modern scale development techniques. To that end, for each

construct, we will produce tables with the Cronbach's alpha (internal consistency) value for each scale, as well as the item-total correlation and "alpha if removed" value for each item.

Structural assessments will be conducted to ensure that an appropriate number of dimensions are employed to model the relationships among the items for each construct. Exploratory and confirmatory factor analysis (EFA and CFA, respectively) model parameters will be obtained from methods appropriate to the categorical nature of the data (e.g., Wirth & Edwards, 2007). The fit of each model will be evaluated using several different fit indices with their customarily accepted cut-off values indicating adequate fit (e.g., Bentler, 1990; Hu & Bentler, 1999; Browne & Cudeck, 1993Steiger & Lind, 1980; Tucker & Lewis, 1973). To ensure that preferred models are not over-fit to the data, each group within each data set will randomly be divided into two sub-samples: one for exploratory modeling and the other for confirmatory modeling.

Item response theory (IRT) calibration analyses will be conducted within each age group for each construct individually. The factor structure of the model for each construct will be the model that was preferred based on EFA and CFA results. Item parameter results from the IRT calibration of each construct will be presented in tables and in the form of trace line plots/item characteristic curves (ICCs) for each item. An ICC for a given item represents the probability of endorsing an item as a function of an individual's level on the underlying construct.

Using the IRT calibration results and expert opinion, the candidate items used in round I of data collection will be re-evaluated. Items that are performing poorly will be eliminated from further use. An examination of the remaining items will be conducted to determine any content areas that are not well-represented or areas on the construct continuum (e.g., ranges along the spectrum from low financial well-being to high financial well-being) that are not well-measured. Item development will be conducted to create items intended to fill any identified gaps.

Using the items that performed well in round I of data collection and the newly developed items, new iterations of the financial well-being and financial ability metrics will be created. The second round of online questionnaires (round II) will be fielded to 3,800 working-age and 3,000 older Americans. Analyses will follow the same structure as outlined previously. The IRT calibration results of round II of data collection and expert opinion will inform the selection of items to the final versions of the metrics for each construct.

As the final step of analyzing the phase II data, it is currently planned that analyses will be conducted to assess if the performance of items differs across the two age groups, working-age and older Americans. However, for these differential item functioning (DIF) analyses to be conducted, it is necessary that there is a sufficient amount of overlap across age-groups, with respect to the items selected for the final metrics. A feasibility review will be conducted to determine, for each construct, if the overlap of the final scales for each age group is sufficient to allow for comprehensive DIF testing across the age groups.

Assuming favorable results from the feasibility review and using the factor model(s) supported by the structural assessments, DIF analyses will be conducted across the working and older Americans age groups. The end result of such DIF analyses, if feasible, is that working-age and older American item parameters will be on the same metric, which will allow scores across the

age groups to be directly comparable to each other.

The final set of analyses is planned for the third round of data collection, which will contain responses from 3000 individuals (1,500 working-age and 1,500 older Americans) and use three different modes of item presentation: Online, Phone, and Paper-and-Pencil. For data from the phone and paper-and-pencil collection modes, CFA models will be fit using the factor structure mirroring that of the preferred model for each construct found using round II data, which was collected via online item presentation only. Using the models supported by the CFA analyses, IRT calibrations for these new presentation modes will also be conducted, individually for each combination of construct and presentation mode. Next, DIF analyses will be conducted to identify any item performance changes across the three presentation modes. Items with poor or inconsistent performance across modes are candidates for removal. These analyses will be conducted using the same procedures outlined in the previous paragraph. Finally, targeted analyses will be conducted to provide preliminary indications of the construct validity (e.g., convergent and discriminant validity) for the newly finalized metrics. Appendix E provides a table of the validation measures included in the data collection, as well as the expected direction of the relationship for each variable/metric with the newly developed metrics. Analyses (e.g., correlational, ANOVA) will be conducted as appropriate for each combination of variables, in consideration of the measurement properties and response distributions of the variables in question. Validation is an on-going process, thus it is expected that future, independent research examining the validity of the developed scales will be required.

5. Contact Information for Statistical Aspects of the Design

Anita Drever, Corporation for Enterprise Development, 202-027-0142, adrever@cfed.org

Leslie Eaton, Survey Sampling International, 703-282-0323
Leslie.Eaton@SurveySampling.com

Dee Warmath, Center for Financial Security, University of Wisconsin Madison,
 262-312-0606, warmath@wisc.edu

R.J. Wirth, Vector Psychometric Group, LLC, 919-768-7142, RJWirth@VPGCentral.com

Bibliography

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin, 107,* 238–246.

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Steiger, J. H., & Lind, J. M. (1980, June). Statistically based tests for the number of common

factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

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

Tucker, L. R, & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Wirth, R. J., & Edwards, M.C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*, 58-79.