Center for Financial Security

UNIVERSITY OF WISCONSIN-MADISON

Using the Right Yardstick:

Assessing Financial Literacy Measures by way of Financial Well-Being

Maximilian D. Schmeiser*, *Federal Reserve Board of Governors*
Jason S. Seligman, *The Ohio State University*

Center for Financial Security
Working Paper 2011-CFS.2

October 2011

*Contact Author
Max.Schmeiser@frb.gov

An index to the Center for Financial Security's
Working Paper Series is available at:
http://cfs.wisc.edu/publicationlibrary/articles-workingpapers.aspx

cfs.wisc.edu

Using the Right Yardstick: Assessing Financial Literacy Measures by way of Financial Well-Being

Maximilian D. Schmeiser
Federal Reserve Board of Governors

Jason S. Seligman[1]
John Glenn School of Public Affairs, The Ohio State University

With the shifting of responsibility for retirement planning and risk from employers to employees, the growth of 401(k) plans and the recent financial crisis, researchers, policy makers and practitioners have become increasingly concerned with the financial literacy of the U.S. population. Along with this interest has come a need to develop measures of financial literacy in order to evaluate interventions to improve financial literacy and examine the implications of financial literacy for financial well-being. However, the measurement of financial literacy is in its infancy. Questions to assess financial literacy have only recently been included in major secondary data sets, and the limited data on financial literacy and financial outcomes has largely led to simple cross-sectional analysis to validate these measures as predictors of financial well-being. Moreover, questions used to assess a respondent's level of financial literacy that ask about knowledge of compound interest, inflation, and portfolio allocation/risk, may be salient in different ways across different segments of the population. Of particular concern is the relevance of these questions for measuring financial literacy among low-income populations who rely more heavily on Social Security benefits, and encounter very different forms of financial decisions and financial products than upper/middle income households. Using data from nine waves of the Health and Retirement Study (HRS) we examine whether existing measures of financial literacy are descriptive and/or predictive of successful household financial management and resilience to the recent financial crisis. Specifically, we assess whether correct responses to various financial literacy questions are significant predictors of successful asset accumulation, resilience to asset loss and changes in retirement expectations. We find that once individual characteristics are carefully controlled for correct responses to many of the financial literacy questions in widespread use are not significant predictors of asset accumulation or resilience to financial shocks.

[1] The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Financial Literacy Research Consortium. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA, any agency of the Federal Government, the Federal Reserve, or the Center for Financial Security at the University of Wisconsin-Madison. The Authors wish to thank two graduate students, Sudipto Banjeree and Dee Warmath for Research Assistance and are grateful for feedback from discussants and participants at the summer CFS FLRC workshop in Madison, WI.

In recent years, the responsibility for managing employees' retirement funds has been shifting from firms to the employees themselves. Following the broad move from defined benefit to defined contribution plans, financial well-being in retirement is now increasingly dependent on an employee's effective management of savings and portfolio allocation decisions across both the work (accumulation) and retirement (draw down) phases of life. Moreover, the recent financial crisis, precipitated by an over-leveraging of consumers and the proliferation of dubious mortgage products, has highlighted the importance of consumer financial knowledge to the macro economy. Consequently, people's ability to make informed judgments and effective decisions about the use and management of their income, financial products and portfolios, or their level of financial literacy, is drawing growing attention from educators, businesses, community organizations and government agencies (Hilgert et al. 2003). Policymakers are concerned that a substantial proportion of consumers lack basic financial knowledge and money-managing capacities, which are indispensable for them to ensure, and enhance, the financial well-being of their families.

Despite the proliferation of research examining the relationship between "financial literacy" and financial well-being, no current measures of financial literacy have been fully validated. Current measures ability to predict long-term financial well-being, resilience to financial shocks, and their relevance for various socio-economic groups facing different financial decisions salient to their financial well-being is more-or-less taken for granted. However without methodically validated measures, the most effective interventions to improve financial capability may be displaced by those of lesser merit. The lack of validation is due to a combination of the newness of academic research into financial literacy and the preponderance of a narrow view of

the financial skills people need to ensure financial well-being. For example, consider three questions that have been widely adopted as measures of financial literacy:

1. *Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow: more than $102, exactly $102, less than $102?*

2. *Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account?*

3. *Do you think that the following statement is true or false? Buying a single company stock usually provides a safer return than a stock mutual fund.*

While these questions may be salient to the financial well-being of higher socio-economic status respondents who frequently deal with financial institutions, have savings accounts, and purchase stocks and mutual funds, it is unclear that these questions effectively assess financial capability for individuals who live paycheck to paycheck, do not posses checking and/or savings accounts, have never even contemplated the purchase of a stock or mutual fund, and whose sole source of retirement income is likely to be Social Security Old Age benefits. Moreover, given the emphasis these questions place on one's mathematical ability, it is unclear they distinguish between actual financial capability and simple numeracy skill, which by itself is a strong predictor of income and assets.

Our study seeks to assess the validity of existing questions intended to measure financial literacy by taking advantage of recently available longitudinal data on responses to financial literacy questions and detailed information on assets available in four recent waves (2002-2008) of the Health and Retirement Study (HRS) [2].

We employ standard regression analysis across several waves of the HRS focusing on the ability of several fielded financial literacy modules to identify relative success and vulnerability

[2] The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and was conducted by the University of Michigan.

to measures of assets and retirement satisfaction, We also engage in more specific panels of cross tabulations to identify better and worse versions of questions in terms of assessing and predicting financial capability and well-being. Finally we review the methods and recent innovations in Item Response Theory with promise for moving further in this direction in a multinomial setting. This review makes the most of existing measures of financial literacy, and informs future research as to the relative merits of employing several competing sets of financial literacy questions -- areas in need of further research.

Background & Context

The work we are doing is prescriptive in as much as it identifies not "financial literacy" but rather salient measures of financial literacy using outcomes as a basis for comparison. In a sense this cuts a "Gordian Knot" – the embedded and related query as to whether financial literacy is useful. That is, a related approach would be to ask, *conditional on being accurately measured – does financial literacy actually improve financial outcomes*. Following from evidence that the improving financial education may not lead to improved financial decisions, Estelami and Hooman (2009) provide specific prescriptions on how financial literacy programs might address five poor financial decision patterns (hyperbolic discounting, short-term memory overload, attribute anchoring, poor knowledge of risk levels, and mental accounting). The authors highlight cognitive biases that exist in consumers' financial decisions and consider them in the context of financial literacy programs. Context matters, the old saying "know your audience," is all about context. Related to this herein we have looked at asset evolution and respondent success in answering three proximate questions asked in the 2002 HRS numeracy module:

-A- "What is 15 percent of 1,000?"

-B- "A pill cures 15 percent of people who have a disease.

If 1000 people have the disease and they all take the pill, how many people will be cured?"

-C- "A store is offering a 15% off sale on all TVs. The most popular television is normally priced at $1000.

How much money would a customer save on the television during this sale?"

Of course all three variants [*A, B, C*] test whether the respondent can identify that 15 percent of 1,000 is 150. The reader may suspect that C is the most successful variant – the 'easiest' but this is not the case. As Table 2a reveals, 70 percent of those responding to -*C*- get to "150" versus 82 percent asked the direct version -*A*-*;* variant -*B*- yields the lowest empirical success--62 percent. That said among medical professionals -*B*- may be more useful, but unless investment advice is similarly contextualized, -A- may still do the best job of identifying those who would benefit from taking a seminar on managing their finances, as opposed to recommending medications.

As further detailed in Table 2b the same basic pattern does not always hold however. For this more difficult question (testing whether the respondent can identify that 10 is 1 percent of 1,000) the direct numeracy question yields a 45 percent correct response rate, while the commerce context, -C- yields 54 percent success. Once again the medical context is least successful – yielding 28 percent. The questions in Table 2 may strike some as pithy; however those answering incorrectly hold much lower wealth in each wave of the HRS on average. Tables 2c and 2d afford the reader greater perspective on how respondents react to each question. Of note while across Tables 2a-2c the medical frame is always the most difficult, with a relatively easy question like 2d – it performs no differently than the direct numeracy version. Remarkably in this case a lottery frame performs worst—with 60 percent correct versus 73-75 percent for the other two, perhaps because people are confusing the odds with the payout.

What becomes clear across the 2002 HRS numeracy module is that the context and frame of a question matter. What is more, whether or not the frame difference appears pithy, inability to answer correctly is consistently correlated with lower asset balances. In a nutshell if you can fool even just some of the people some of the time it would appear to be profitable.

A significant body of literature is emerging examining the relationship between various individual financial behaviors, knowledge, and characteristics and long-term financial well-being. Meier and Sprenger (2010) try to measure each person's present bias and long-run discount factor. They found that impatience is associated with higher debt levels and time preferences are important in determining credit behavior. Seligman (2010) did not find any significant change in success with financial literacy questions related to patience but did find some evidence of increased confidence in financial ability.

Muller and Weber (2010) examine the relationship between financial literacy and mutual fund investment behavior. They use an internet survey conducted in May 2007 in cooperation with a large German newspaper. They employ an eight item quiz, self-assessment of financial expertise and find that while financial literacy is a strong predictor of knowledge about passive investment alternatives, a very small percentage of people select passively managed funds. They suggest "overconfidence" as an explanation. The study also reports higher literacy among males, purchasing funds on-line, working in the financial sector, higher education, higher income and being middle-aged. Seligman 2010 finds overconfidence within the very wealthy – who ostensibly mistake success for evidence of skill versus luck on the margin, and among non-white respondents who score poorly on a 12 item quiz.

Meier and Sprenger (2007) consider self-selection into financial counseling. In a first round selection filter they query *"Do you know what a credit score is?"* coding responses *"-a-*

No, I don't know; -b- Yes, I know but it is not important for me; -c- Yes, I know and it is important for me. And then balancing offers for counseling across the groups. They observe the most vulnerable as least likely to seek help. Seligman (2010) finds that low education non-white populations display strong overconfidence, suggesting that this may be due to language and cultural differences,.

The Credit Card Accountability, Responsibility and Disclosure (CARD) Act of 2009 mandated GAO to examine relationship between fluency in English and financial literacy. Their study—GAO (2010)—found that a lack of proficiency in English creates significant barriers to financial inclusion, including a greater probability of being unbanked, use of alternative financial services (e.g., payday lenders), and a somewhat higher susceptibility to predatory or fraudulent practices.

Huston (2010) provided a summary of the broad range of financial literacy measures used in the last decade, including the definitions of financial literacy used in the various studies. She found that the terms financial literacy, financial knowledge and financial education are often used interchangeably. Currently there is no standardized instrument to measure financial literacy. Kozup and Hogarth (2008) examine key factors for an effective financial literacy effort and find financial literacy is, in part, numeracy (the ability and comfort with manipulating numbers) mated to a set of critical thinking skills that allow one to weigh the pros and cons of a particular decision relative to one's own personal needs, values and goals.

Lusardi and Mitchell (2008) considered why female-headed households, are less likely to plan for retirement than other demographic groups. Using the 2004 HRS module on planning and financial literacy they found women who have higher levels of financial literacy are more likely to plan and plan effectively. However, this begs the question of why women are less financially

literate than men using their metric. Seligman (2010) finds that women are slightly less successful that male counterparts in answering a 12 item quiz, and that women display less self confident than score differentials alone would suggest. He concludes that women's financial vulnerability stems in part from a lack of self-confidence in capacity to make decisions relative to comparable males.

Agnew et al. (2008) used a controlled experiment and found that women are more likely than men to choose annuities for managing resources over the course of their retirement. Notably, women's financial literacy as measured via a 10-item quiz score was significantly lower than men. Women were also more likely to choose the annuity option, even when controlling for financial literacy and risk aversion.

Braunstein and Welch (2002) analyzed the NEFE 1997-1998 survey of its High School Financial Planning Program which measured financial knowledge, behavior and confidence. Their findings suggest that more successful literacy measures emerge as a result of six survey design considerations that place the target audience in context.

Remund (2010) provided an analysis of the many ways in which financial literacy has been defined and measured in an attempt to provide a clear and consistent definition. The study examines existing studies, programs and evaluations.

Mandell and Klein (2007) examined whether an incomplete understanding of the weakening or elimination of important safety nets has contributed to a lack of intrinsic motivation to learn and retain concepts of financial literacy. The study used a national JumpStart survey of high school seniors. It used three questions designed to measure motivation to be financially literate:

1. *Which of the following do you feel is the greatest cause of serious financial difficulty, …*

2. *How bad do you think it is for families who don't have enough money to pay their bills?*
3. *What do you think happens to older people when they retire if they haven't saved much money and don't have a good pension from their former jobs?*

The "push" style survey questions were designed by a committee of financial educators and found that increasing motivation can increase the "stickiness" of financial education efforts.

Worthington (2006) noted that changes in the financial services market in Australia required that consumer become more knowledgeable if they are going to manage their finances effectively. His study uses the ANZ Survey of Adult Financial Literacy in Australia. He found that financial literacy is highest amongst: those 50 to 60 years of age, professionals, business and farm owners, and university/college graduates. Financial literacy was found to be lowest amongst the unemployed, females and non-English speaking individuals with a low level of education. This finding is generally consistent with Agarwal et al. (2007) who found financial facility to be greatest around age 53, and decreasing thereafter.

Gerardi, Goette and Meier (2010) examined whether numerical ability (an element of financial literacy) played a role in the defaults and foreclosures in the U.S. subprime mortgage market and found a large and significant negative relationship between numerical ability and default/delinquency, even when controlling for other elements of financial literacy. Also, they found that higher cognitive ability was associated with lower rates of foreclosure. Of course these correlates may be more or less direct – lower numeracy may tend to cause people to choose mortgages with worse terms, but it may also reduce employment opportunities which then increase overall economic vulnerability. Regardless of the channel – and there is likely more than one at work-- if the goal of the financial literacy literature is to identify and address vulnerabilities, understanding how numeracy, framing and cognitive effort interact is important.

It requires a methodical consideration of the data on past questions to better understand what works, what does not, where myopia exist and, then following that effort how to best to go about addressing them over the lifecycle.

Item Response Theory

The most commonly used technique to identify a set of questions that best assess one's underlying ability in a given area, such as mathematics, on the basis of correct responses is Item Response Theory (IRT). (IRT is employed in the production of popular tests such as the Scholastic Achievement Test (SAT), for example.) IRT cannot be used to indentify whether different survey/test items measure the same underlying construct – only how people at different points on the ability scale respond or how those item characteristic curves (ICC) differ by subgroups. IRT models are generally based on the logistic function and come in three degrees of complication. The first being a one parameter model that measures a relative difficult parameter, b_j, across questions J based on observations across responders N. The parameter is estimated as the point along the ICC with maximum slope, or in the logit form, as the value for which a latent ability θ_n is equal to 0.5 (fifty percent). A two factor model adds a relative discrimination parameter, a_j, which identifies the sheerness of cut off between correct and incorrect answers along the ICC to question j based on the global test score across the full set of questions J. The discrimination and difficulty of each individual item are examined to determine where on the ability curve the item operates, with difficulty identifying the point on the ability scale and discrimination indicating how well the item distinguishes individuals in this area on the ability curve. Finally the fullest IRT model includes a final term c_j which allows the intercept of the ICC to vary – estimating the ease with which an item may be correctly guessed. Generally a one

parameter model assumes that all items are of equal discrimination and that concerns regarding guessing are negligible. Equation 1 describes a three parameter logistic model (3PLM):

$$P_j(\theta_n) = c_j + \frac{(1-c_j)\exp\{a_j(\theta_n - b_j)\}}{1+\exp\{a_j(\theta_n - b_j)\}} \quad (1.1)$$

Fayers and Machin (2000) offer a very straightforward treatment of Item Response Theory and Differential Item Function based on the above logistic framework for assessing the relative value of survey questions for estimating respondents' quality of life. As Baker (2001) describes, the objective of IRT is to find the ICC that best fits the observed proportions of correct responses. Thus IRT would take the correct responses to a variety of questions and be able to locate individuals on a financial literacy (ability) scale. Below we review some articles describing the use of IRT to develop scales for measuring ability in areas similar to financial literacy, such as reading literacy. We then highlight the first article, to our knowledge, to apply IRT to financial literacy questions.

Kaplan and Venezky (1993) use IRT to examine the value of the GED Tests as "a certification mechanism" (p. 2). GED Tests have been developed and refined over a long period, starting in 1942, with the objective of certifying a level of competency in five areas. The question is whether employers should use the GED to assume certain abilities of the individual job applicant relative to high school graduates and drop outs without a GED. They used a subsample from the Young Adult Literacy Survey (YALS), which administered three scales of functional literacy: Prose, Document and Quantitative. Scoring for each of these scales was accomplished with IRT.

The first step in their analysis was to examine mean differences in these literacy scores for demographic subgroups, which showed potentially significant differences. Using IRT scaling, items and subgroups were located on the scale at the point where subgroups or

individuals had an 80 percent probability of answering the item correctly (θ is set to 0.8). Each item/subgroup has a distribution of correct responses above and below this point, indicating the steepness of the item's characteristic curve. To compare items at the same location on the curve, you have to examine how rapidly the probabilities decline. Using this approach, the authors conducted regression with three blocks of variables being entered: demographics, literacy areas and educational route (GED, High School Dropout, High School Graduate). IRT allowed them to take into account individual-level differences and allowed them to explain previous findings of significant differences among educational routes. "The differences found in literacy abilities, therefore, must result from a selection rather than a training or practice factor; that is, those who studied for and completed the GED Tests must have had higher literacy skills on leaving high school than those who did not complete the GED Tests" (p. 26).

Bjorner et al. (2003) analyzed response to four well-established tests of headache impact, applying confirmatory factor analysis (CFA) and IRT. Their objective was to solve the dilemma between respondent burden and test precision, which they accomplished with a combination of computerized adaptive testing (CAT) to randomly select the items presented and IRT to score the results. The starting point for potential items was the four well-established scales that had been demonstrated over numerous years to be valid and reliable measures of headache impact. In this particular effort, IRT produced gains in test precision and responsiveness to changes over time. The underlying approach was 1) use factor analysis to develop or confirm the underlying measurement scale, 2) use IRT to examine item properties, and 3) test for differential item functioning at a subgroup level.

Thissen et al. (1983) demonstrate the application of IRT to the development of a scale to measure the propensity toward jealousy in romantic relationships. In their description, "the

essence of IRT is the specification of a statistical model for each item response in which the response is described as a function of the trait being measured and some item parameters" (p. 212). The approach they describe begins with preliminary iterative analysis of items in the available pool using factor analysis to eliminate items measuring other dimensions or latent traits. The ideal objective is a set of items measuring a single dimension, trait or ability with equal weighting that allows a simple sum to create the score.

Thissen et al. (1986) originally set out to shorten the 72-item Masher Forced Choice Sex Guilt Inventory (MFCSGI) using IRT, but ran into the problem of item bias. IRT allows the authors to distinguish between real group differences and test bias. Test bias refers to "a formalization of the intuitive idea that a test is less valid for one group of examinees than for another group in its attempt to assess examinee differences in a prescribed latent trait, such as mathematical ability" (Shealy and Stout 1991). "Test bias occurs if the test under consideration is measuring a quantity in addition to the one the test was designed to measure, a quantity that both groups do not possess equally" (p. 4). For example, with the MFCSFI they found that the scale contains more items that elicit guilty responses from women than from men, producing the observed gender differences in guilt on the scale. The analysis involved the estimation of item characteristic curves for each item for men and for women. These item characteristic curves are built from the probability of a correct response at each level of the underlying trait. The dependent variable, then, was sex guilt as indicated by the established MFCSGI.

Reise et al. (1993) use CFA and IRT to investigate whether mood ratings collected in Minnesota and China are comparable. The same scale (NA5) was used in both countries with the primary question being whether the trait scores for each group are on a common measurement scale. The focus of the analysis, then, is at the level of the latent variable

underlying the scales. CFA models account for the covariance between test items to accomplish this objective. IRT models use the item response function to "account for the relation between examinee level on a latent variable and the probability of a particular item response" (p. 557). The desired finding is that of measurement invariance, which would mean that the NA5 scale produces results on a common measurement scale across subgroups. One major difference in the two approaches is that IRT allows for analysis at an examinee-item level.

Osgood et al. (2002) use the graded response model of IRT to address issues related to current self-reported scales of crime and deviance. They apply this approach to evaluation of a 14-item measure of delinquency. The benefits of a multiple-item measure and summative scale apply "only when the research question at hand concerns a unifying construct that is reflected in the content of each of the items" (p. 269). In crime and deviance measures, there is the challenge of biased reporting on individual items that produces bias in the summed scale. In their words, "IRT is built on mathematical models that relate responses for a set of items to positions on a latent dimension" (p. 276). "An IRT model is an equation that translates positions on the latent dimension to probabilities for the possible responses to each item" (p. 278). IRT scoring "applies maximum likelihood methods to infer the position on the theta dimension that is most likely to produce the pattern of item responses observed for an individual" (p. 276). IRT has very high mathematical complexity that is not required for most social research problems (summative scores "prove quite adequate for most problems"). Its primary unique benefit is that it addresses a problem in standardized tests such as comparability of scores across versions of test and selection of items for computerized adaptive testing. They use another aspect of IRT, the graded response model, applying it to refinement of the MtF measure of delinquency. The article concludes with an assessment of whether and when IRT is worth the extra trouble it requires. In

their view, it depends on your objective. In developing a scale or evaluating a measure, IRT provides incremental information that can be valuable. Conceptually, however, IRT does not bypass the basic steps in measure construction, starting with an understanding of what can be measured. What these authors did not do was to take every item available regardless of what they measure substantively to see how a scale could be created. They started with the Mtf measure of delinquency.

MacDonald and Paunonen (2002) compare the results of IRT and CTT, finding many similarities between them. The primary difference observed was that IRT estimates held across most experimental conditions, while CTT estimates held only in some. IRT was less influenced by characteristics (or "vagaries") of the item pool. "This finding suggests that a test constructor's item selection decisions based on item difficulty and discrimination estimates are more likely to result in the best possible subset of test items with IRT methods" (p. 942). Item difficulty estimates were highly comparable between IRT and CTT. Item discrimination was comparable only for certain circumstances. When difficulty is within a narrow range, item discrimination estimates are comparable between IRT and CTT. When the range of difficulty widens, CTT estimates decrease.

In the above articles, the starting point for their analysis was a validated scale of an underlying trait or ability. Kaplan and Venezky (1993) examined an established high school equivalency test. Bjorner et al. (2003) analyzed responses to four well-known tests of headache impact. Funk and Rogge (2007) used IRT to assess the precision of eight well-validated self-report measures of relationship satisfaction. In his *Basics of Item Response Theory*, Baker (2001) notes that a starting point for IRT is a scale that can be used to tell how much of an ability someone has.

In a recent working paper, Knoll and Houts (2011) applied the IRT analysis to the preponderance of financial literacy questions available across a range of data sets to identify a set of questions that are most effective at discerning an individual's financial capability level. Specifically, beginning from 60 questions drawn from several waves of the HRS, the American Life Panel, and the National Financial Capability Study they identify the 20 questions which are most psychometrically-sound to field. Their study takes an important first step in validating financial literacy questions, it is unique in many ways – including its use of a preliminary software for sorting and estimating a two parameter IRT model and a very elaborate multi-dataset latticing of subsets or questions and subpopulations to inform it's identification and reporting of key difficulty and discrimination parameters. An extension of this model to the three parameter setting is conceptually straightforward, but technically this may be a non-trivial innovation on their current model.[3] Because of the pattern of queries and skips across the HRS financial literacy populations employing the HRS alone to identify question vs. respondent IRT parameters is very difficult. We have been unsuccessful in employing either the commercially available Multilog software authored, or the recently written Stata module OpenIRT with our 2002 – 2008 HRS question data due primarily to the limits of the subsamples and their overlap.

Data

Our primary data come from the Health and Retirement Study (HRS). The original HRS cohort consists of persons born between 1931 and 1941 and their spouses. Members of the cohort were first interviewed in 1992 and have been re-interviewed once every two years. Since then several new cohorts have been added. The original 1992 survey included questions

[3] Assuming a naïve guessing strategy for test takers/survey respondents the guessing parameter is simply one dividend by the number of options, however if answer choices collapse based on poor question design, or proneness of response bias then this approach must be modified, on a question by question basis.

regarding whether persons participated in defined contribution pension plans, and whether they could choose how to allocate these funds as well as a question regarding participation in financial seminars. This was replicated in the 2000 survey albeit for a much smaller sample, and with almost no overlap as described in Seligman and Bose (2006).

In 2002 variants of numeracy questions, like the one discussed above in the introduction to this paper, were offered, along with a module devoted to risk assessment. Additionally a large subsample of the HRS population took a three item financial literacy quiz with numeracy questions framed in the varying contexts – a question on percentages framed in medical terms, a question regarding division in a lottery frame and a question on compound interest provided in a savings accumulation frame. The 2004 wave of the HRS includes modules on risk, annuities, social security, and financial literacy wherein a set of three questions are asked of a subset of approximately 3,000 respondents to assess their level of financial literacy. These three questions were repeated in 2006 across a broader sample of over 17,000 respondents including replication for almost the full 2002 sample – allowing us to consider the 2002 application as a potential treatment.

In 2008 an expanded set of financial literacy/ economics education questions were fielded to a subset of HRS respondents. In addition to these, the questions on financial literacy were replicated on a small subset of respondents, of these some had been previously offered the questions in one or more waves from 2002 - 2006. Table 1 describes the full intersection of modules across respondents who answered at least one question from any two of the modules just described, the HRS contains extremely detailed information on respondent demographics, assets, health, income and employment. It also contains an oversample of African Americans, Hispanics, and residents of Florida.

The RAND version 'k' data provide more consistent measures of assets across waves – it is these measures which are employed for the asset evolutions offered in Table 2 and it is from these that we obtain all of financial, retirement satisfaction and regression control variables. However while the HRS offers the opportunity to consider a very broad set of behaviors and responses over the latter lifecycle, even after implementing the RAND imputations to increase the robustness of the data sample sizes quickly dwindle – for example, while over the panels many cohorts over 48,000 persons participate in the Core HRS data collection, of those only 12,773 observations hold asset balances for the first wave, and less than half (a maximum of 21,492 for the 1998 wave, or 17,339 for the 2008 wave for example) provide asset balances in any single wave. Building a robust panel out of these data is therefore not easy – even given the fine preassembly and refinements provided via the RAND work.

Using multiple waves of the HRS data one might investigate rates of financial literacy across various populations, which of the questions are most salient across socio-economic groups, the variation in responses across groups, and the ability of the questions to predict financial well-being. However because of the relatively small subsample of each HRS cohort that engages financial literacy questions in each wave merging the data tends to yield very small cell counts from which to work. In presented regressions small sample sizes are the norm and many specifications fail to run or have so few degrees of freedom that the F-test, t-test and R-squared measures imply that so-called "suggestive evidence" is anything but. In a few regression panels however we are able to identify the relative merits of questions and to see how subpopulations vulnerabilities vary across specifications. Table 1 presents the patterns of HRS subsample exposure to financial literacy over the core 2002 – 2008 biennial surveys. Of particular note are the highlighted strong sample sizes for the 2002 and 2006 financial literacy modules including a

subsample of roughly 12,000 individuals that is (at least) double exposed to these questions. With multiple waves measuring financial literacy, one might assess the consistency of responses over time, and we are able to identify the value of improving scores, both generically across module and within each of the three questions employed.

With the recent recession beginning in December of 2007, and the economic collapse accelerating through 2008, these data suggest an opportunity to examine whether current measures of financial literacy actually predict resilience to economic shocks, such as unemployment, severe drops in portfolio value, disability and home foreclosures, and whether or not they do so equally well for individuals across the socio-economic spectrum. Unfortunately however the timing of the 2008 interviews is ahead of the bulk of impacts affecting the US economy and financial markets, for example we estimate that the average respondent was interviewed in the context of an unemployment rate of roughly 5.6 percent, and well ahead of the Lehman Brothers bankruptcy on September 15th 2008. Figure 1 documents the collection of HRS 2008 wave interviews in the context of these unfolding events.

Greater promise in this area is to be found in later related data collections. In 2009 the HRS fielded their off-year Internet Survey and collected responses from 4,433 respondents, about a third of which come from the core HRS sample (the others come from samples such as the American Life Panel). The 2009 Internet Survey was fielded from March 2009 through August 2009. The 2009 Internet Data contain a sample drawn from respondents who reported internet access in the HRS 2008 Core survey, plus those who did not respond to the 2008 Core survey but had been selected for the 2003, 2006, or 2007 Internet surveys. Unfortunately there are almost no data from the HRS subsample which are full enough in terms of item responses to

merge with our regression data. The upcoming release of 2010 wave data offer more promise in this area, and an opportunity for future work.

Exploratory Analysis – Findings from Panels of Cross Tabulations

We evaluate questions in table-frames like that already presented in Table 2, looking for value in questions based on how correct and incorrect respondents' assets evolve. Second we report on the impact of repetition for the three questions contained in the 2004, 2006, and 2008 as a first or second treatment to see whether persons are consistent in their responses or learn and improve their answers over time. We also present evidence of guessing/inattentive listening behaviors via this same method. While we would very much have liked to have presented evidence for even just a single treatments' impact on assets, except for the 2006 data, for whatever reason not enough of those queried in 2004 and 2008 provide asset data for their households. Across the 2004 sample the number answering the financial literacy and the asset questions range from 30, to 5, or less before segregation into correct and incorrect answers. Requiring that the asset values be aligned at the observation level brings these counts to essentially nil in all cases. The 2008 data provide an even smaller sample than in 2004. The 2006 queries are more robust – but not for repeat respondents – of those repeating, regardless of whether they are correct or incorrect, cell counts for those with asset data plummet to the level of fewer than 20 observations in every case. The reader is reminded that the hypothesis we are testing to find preliminary evidence for or against validity as a measure of financial literacy is simply whether exposure to these problems constitutes a treatment motivating greater financial literacy. Using our panels of cross tabulations we next describe what a lousy question looks like,

what a latent vulnerability looks like, how reversing frame can impact the population that responds in the affirmative, and how competence can be identified as successfully self reported.

What does a lousy question look like; what does a latent vulnerability looks like?

Educators often tell their classes that there is no such thing as a "dumb question" – this is not true. A dumb question is one for which answering "correctly" or "incorrectly" does not distinguish a person's insight, preference, or latent ability in any meaningful way. But every query has the potential to measure something and this is where the demure are distinguished from the dumb – consider the following question asked in 2008:

You should put all your money into the safest investment you can find and accept whatever return it pays. (True or False?)

You might think that asking this question in 2008 would be a great idea because all around the world there was a 'flight to safety' and for a moment in time while this survey was in the field (albeit late in that period) sophisticated investors were accepting negative returns, just to hold US Treasuries which were perceived of as a safe investment. Thus asking this question might pick up on whether households were in fact reacting in some way much as professional money managers might with their money.[4] You might also think however that this is a very dumb question. If everyone is in a panic, then those answering "True" (the incorrect answer under normal circumstances) might not be distinguishable, and in fact you might suspect that the vast majority answered "True." As reported in the Table 3 68 percent did in fact answer this way. By the middle panel of Table 3 roughly a third of those reporting assets answered "True".,

[4] Hrung and Seligman Warren B. Hrung and Jason Seligman, 'Responses to the Financial Crisis, Treasury Debt, and the Impact on Short-Term Money Markets', *Federal Reserve Bank of New York Staff Report*, 481 (2011). document the true value of the Federal Reserve's Term Securities Lending Facility which in fact swapped higher risk paper for Treasuries over brief periods for a premium – the flight to quality was not non-stop, and further, not direct.

Moreover, prior to 2006, those answering "True" in 2008 tended to have higher reported assets – over 1992-2004 they hold from 97-131 percent of the assets of those who answer "False". Beginning in 2006 those answering "True" suffer larger declines on average from 04-06, and are holding lower balances in 2008. Of course the possibility exists that those answering "True" are simply older workers and retirees (The HRS population) whose spot in the lifecycle suggests asset preservation is important. The bottom panel of Table 3 suggests that this might be the case, as we see observation linked declines as a percent of asset balances overall. Those reporting "True" outperform over most periods including the 2000-02 and 2006-08, periods in which equities markets take meaningful losses. In any case this question deserves more investigation via multivariate analysis – as statistical power allows. And more importantly it should be repeated on the same population in years to come to see how consistently people are responding – in order to tease out short term reaction from long term strategy. Controlling for age, should the percentage reporting true decline that would be a strong sign of reaction to the turmoil of the moment. However, given their performance over the two latter broad market declines in these data it is likely that what we are seeing is a long term strategy, and that our own thoughts as researchers using these sorts of questions about which answer choice is correct may need some lifecycle modification.

How does reversing frame impact the population that responds in affirmative?

Consider the following question presented to 1,227 respondents in 2008:

If you start out with $1,000 and earn an average return of 10% per year for 30 years, after compounding, the initial $1,000 will have grown to [more/less] than $6,000 (True or False?)

Roughly half were asked the *[more]* version and half the *[less]* version. It matters – generally 81 percent answered the *[more]* version correctly while only 63 percent answered the *[less]* version correctly – of those reporting assets success was insignificantly greater in each case (83, and 67 percent respectively). This same pattern is found across many of the questions, especially in cases where the respondent is made to agree or disagree as in "True/False" contexts the population appears prone to agreement – question variants requiring an answer of "False" consistently appear more difficult – fewer surveyed answer correctly . Which version should be asked? If we are looking for vulnerability the answer is the harder one. For in the latter frame agreeability and guessing are punished and more easily detected – even if we may get more false positives. To see this latter point compare the asset evolutions across answer choices in the top and bottom frames. Those that get the easy version of the question wrong show asset balance differentials that average 42 percent (1-0.58) versus 33 percent (1-0.67) for the more difficult version. Although the differentials are on average greater for those who miss the easy version of the question, they are fairly large in either case – so asking the tougher question may be seen as essentially protective.

Confidently self-reporting real competence – what does it look like?

And finally, consider confronting this 2008 question:

I understand the stock market reasonably well (True or False?)

The reader may suspect that this would beg overconfidence – especially given the agreeability bias that we observed in the last dual frame just above; however, this does not appear to be the case. In fact the bottom panel of Table 4 suggests strong prudence/forbearance in the way respondents answered – just 27 percent of both the full and asset reporting samples

answered this question "True" – and it appears they are right, for they consistently hold roughly twice as much wealth as those who answer "False." Asking this question in 2000 might have yielded a different set of results, but asking it going forward is likely going to have to suffice as a check on robustness. If the question does hold up it may well be in part to the use of the word "*reasonably*" which modifies the statement.

What is the value of a question as treatment – does it motivate learning?

Often one senses that a survey population may learn from being asked a question. Certainly there are a lot of dollars invested in "push" surveys during elections – surveys that embed ideas in the respondent – often these frame a candidate negatively and attempt to appear innocuous. The question:

"Often people comment on the weakness of candidate [X]'s character – calling [him/her] a 'flip flopper' on the issues, how do you feel about candidate [X]'s character?"

is an example of this strategy. The HRS has asked the three financial literacy questions first introduced in 2004, one, two and sometimes three times of the same respondent. This repetition allows us to consider whether the respondents are improving following the initial response. If this were true it would be wonderful news – for it would suggest that, as with the Mandell and Klein (2007) study, the simple act of asking a question regarding financial literacy may motivate a person to go out and learn. Table 5 presents evidence on the repetition of an easy, medium and hard financial literacy question – as evidenced by rates of success on first response. The easy question focuses on calculating a percent: *If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?*

Table 5a reports that in 2004 89 percent successfully answer this question. However of the 421 respondents who will be asked this question again in 2006 and still get it wrong – success declines to 58 percent – suggesting a latent vulnerability. Of those that get the question incorrect the first time, the percentage responding correctly in 2006 is remarkably similar 59 percent. For those asked the question a third time success increases from 84 percent in 2004 to 89 percent in 2008, or 90 percent if they got it correct in both 2004 and in 2006. Not very impressive as far as treatments go. Either that or perhaps people are just very prone to careless errors.

Table 5b reports on the question focusing on division:

If 5 people all have the winning numbers in the lottery and the prize is two million dollars, how much will each of them get?

Only 57 percent of respondents get this question right in 2004, and that percentage goes down to 45 percent if they answer incorrectly when asked in 2006. Of those who get it right in 2004, 72 percent repeat that performance. Conditional on getting it wrong in 2004 46 percent correctly respond at the next go round. In 2008 first time responders get it right about 37 percent of the time. If a respondent had answer the question correctly prior to 2008, they have about a 53 percent chance of doing that again in 2008. A respondent who answered correctly twice in the past increases their probability of a correct response in 2008 to 71 percent, which is far less than one would expect.

Table 5c report on the question focusing on compound interest:

Let's say you have $200 in a savings account. The account earns 10 percent interest per year. How much would you have in the account at the end of two years?

This is the hardest of the questions – 14 percent of those surveyed get it right in 2004, and of those never before asked in 2008 only 2 percent correctly answer. Of those who got it wrong in 2004 21 percent get it right in 2006. While at first glance this may appear to provide evidence of learning, notice that of the 14 percent that get it correct overall in 2004 – the probability of getting it right in 2006 remains less than half, and those who got it wrong in 2004 perform 12 percent of the time, again very close to the proportion who got it right in 2004 but are destined to fail answering it correctly in 2006 (11 percent). Finally of the 12 respondents who got it correct in both 2004 and in 2006, precisely none get it right again in 2008. While the precision of this question makes it unlikely respondents are guessing, it does appear that the respondents are just very prone to careless errors.

For the first two questions discussed above it is likely that failure to repeat past success is evidence of guessing, or that respondents will be inattentive when asked the question again in subsequent waves of the HRS. Regardless of the reason for the subsequent error it may suggest vulnerability – for even if the case should be the latter and it should be due to boredom, it suggests that delaying you with paperwork at a car dealership, or at a home closing might arouse boredom and careless impatience. Should it be the later thing might even be worse. And should you not know you are guessing – well that is likely the worse case of all. The learning hypothesis is further refuted by the symmetry between the percent who get it right in 2004, but who will not in 2006, and the number of previously unsuccessful persons who will in fact answer correctly in 2006. If leaning was a large component of the 2004-06 improvement, you'd expect the number of respondents improving to greatly outweigh those who fail conditional on previous success – but in percentage and raw count the paths up and down are really quite equal over sample sizes of 421-423 (a); 851-872 (b), and 1,710 -1,729 (c).

Multivariate Analysis – Findings from Regression Analyses

The multivariate setting affords more opportunity for controlling for several interacting factors, conditional on the data on which it is employed. In this section we describe work across eight considered dependent variables and seven financial literacy module variants. Our dependent variables are organized as follows, first between financial measures of wealth and holistic measures of satisfaction. Within the wealth measures we employ measures of total assets, and all but housing assets as of 2008, we also employ the change in both asset measures between 2006 and 2008—four dependent variables in all. For the holistic retirement satisfaction measures we employ measures of retirement satisfaction taken in 2008 and also the change in this measure between 2006 and 2008. The number of observations for either of these measures is smaller than for the asset measures. We also attempt to employ two additional measures of retirement satisfaction relative to the years just ahead of retirement, and the change in this measure of relative pre-post retirement satisfaction between 2006 and 2008. The number of observations with either of these measures is prohibitively small and none of our regressions are successfully run. Table 6a provides univariate summary statistics for the full sample of each dependent, financial literacy, and control variable.

Included financial literacy question modules constructed and regressed upon the dependents are the 2002 Numeracy – percents module, containing four questions, the 2004 Social Security knowledge (two questions) the 2008 economics and investing module (eight questions), two subsets of this module, one containing five diversification questions and the second containing the residual three questions, and the matched 2002-2006 financial literacy questions (three questions asked twice, for a total of six.) Finally to test the learning/treatment

vs. guessing aspect of each repeated question the matched module is decomposed into three variables focused on evolving responses. These are coded as follows: ("0" if never correct, "1" if correct only in 2002, "2" if correct only in 2006, "3" if correct in both instances). It is unfortunate that the sample selection across modules does not allow us to include many of them in the same regression. In this way our specifications can be said to suffer from a lack of comparable financial literacy measures as controls. This may explain some of the volatility in coefficients on control variables across the specifications for each dependent variable.

Included controls are comprised of a dummy for whether the respondent or their spouse has a defined contribution pension plan within their first reported employer plan in any interview wave, their coefficient of risk aversion, level of education, year of birth, and whether or not they are female, of African heritage, of other non-white heritage, Hispanic, and whether they have ever reported applying for disability insruance. All of the control variables are selected for their ability to identify particular group characteristic driven vulnerabilities.

Tables 6b-g report on six panels of regression employing the six dependent variables with sufficient observations for analysis. Tables 6b-e focus on the Asset variables. In both the total wealth (6b) and the non-housing wealth specifications correct responses to the 2002-2006 financial literacy modules are strongly correlated with improved asset holdings in 2008, significant at or above the one percent level. For total wealth the impact is estimated to be roughly $459,000, of which the non-housing portfolio comprises roughly $203,000. In Table 6b this finding is strongly associated with repeated success with the compound interest question, itself associated with a positive impact of roughly $419,000 at the 95 percent confidence level. Across these two dependent variables there is some repeated evidence of the importance of education attainment in generating wealth, and some evidence that those coming from other non-

white heritages are in better financial position than the baseline white male respondent. Those reporting having ever applied for disability show consistent negative asset impacts relative to others; although the statistical weakness of this evidence leads it to be more suggestive and supportive of other work on disabled and involuntarily retired populations.[5]

Moving to changes in total and non-housing wealth (Table 6c and Table 6e) the financial literacy modules are not as frequently predictive of statistically significant economic improvements. For total wealth, the 2002-2006 financial literacy modules show some evidence of being associated with positive relative impacts on the order of $216, but this is at the weakly significant 10 percent confidence level. There is spotty evidence suggesting that those of non-Caucasian, non-African heritage outperform in terms of asset management over this time period and that those who have applied for disability underperform by a substantial margin. Moving to our last two included dependents on retirement satisfaction none of the tested financial literacy measures are associated with meaningful impacts on retirement satisfaction. It must be noted that several of the specifications suffer from small samples – these are highlighted in the table. In particular, within Table 6g the 2006-08 change in satisfaction suffers from very small sample sizes and should be discarded. These are identified by having their columns reports struck though. In what remains there is some evidence that African heritage is associated with greater retirement satisfaction and strong evidence that having ever applied for disability insurance is associated with lesser retirement satisfaction. Remarkably across all specifications it appears that knowledge of Social Security impacts groups differently than other forms of financial literacy. In particular those of African heritage would appear to benefit from this knowledge.

[5]Lachance and Seligman Marie E. Lachance and Jason S. Seligman, 'Involuntary Retirement: Prevalence, Causes and Impacts', (The Ohio State University, 2011). find strong evidence that those who exit the labor market for health reasons hold fewer assets in retirement.

Findings on Financial Literacy for Vulnerable Populations

Our analysis of the response pattern to existing financial literacy questions, as well as the value of correctly responding to these questions for predicting financial well-being, suggest a clear difference for vulnerable populations. The primary finding is that the financial literacy measures currently in use suggest that women and minorities have lower levels of financial ability than white males. Whether this difference in correct responses represents an actual difference in ability or simply is a byproduct of the nature and type of question being used is unclear at this point. Therefore, this issue requires further exploration, as any potential racial/gender bias in the assessment value of financial literacy questions may fundamentally bias research on financial literacy.

Conclusions

Herein we have detailed some of the more promising questions we have compiled and tallied. We are heartened by some of these findings which show that the yardstick does matter. We have identified the IRT protocol as being of particular promise for identifying questions of salience, but have been challenged by the lack of sample from which to run even the most basic regressions. In a sense we close by noting that financial literacy appears to be at a cross roads of sorts – on the one hand it would be nice to validate questions before fielding anything further in large doses, on the other hand it appears difficult to validate any of these measures based on the previous pattern of efforts to field them. We look forward to the release of the 2010 HRS, in hopes that it will better identify the value of financial literacy protocols for asset preservation – especially for those among traditionally vulnerable groups. Beyond this in terms of future research and policy we suggest that fielded financial literacy survey efforts be designed with

Item Response Theory based identification in mind – much as "experimental sections of the SAT are fielded each year. The recent introduction of a very large sample by FINRA would appear promising as a base from which to develop the experimental module variants from which better financial literacy questions can arise. The yardstick does appear to matter, but to date we've moved merely a foot in the direction of sound measures.

References

Agarwal, Sumit, et al. (2007), 'The Age of Reason: Financial Decisions Over the Lifecycle', *National Bureau of Economic Research Working Paper Series*, No. 13191.

Agnew, Julie R., et al. (2008), 'Who Chooses Annuities? An Experimental Investigation of the Role of Gender, Framing, and Defaults', *The American Economic Review*, 98 (2), 418-22.

Baker, Frank B. (2001), *Basics of Item Response Theory* (2nd edn.; Washington, DC: Office of Educational Research and Improvement) 186.

Bjorner, Jakob B., Kosinski, Mark, and Ware Jr, John E. (2003), 'Calibration of an item pool for assessing the burden of headaches: An application of item response theory to the Headache Impact Test (HIT™)', *Quality of Life Research*, 12 (8), 913-33.

Braunstein, Sandra and Welch, Carolyn (2002), 'Financial Literacy: An Overview of Practice, Research, and Policy', *Federal Reserve Bulletin*, 88 (11), 445-57.

Estelami, Hooman (2009), 'Cognitive drivers of suboptimal financial decisions: Implications for financial literacy campaigns', *Journal of Financial Services Marketing*, 13 (4), 273-83.

Fayers, Peter M. and Machin, David (2000), *Quality of Life* (West Sussex, England: John Wiley & Sons).

Funk, Janette L. and Rogge, Ronald D. (2007), 'Testing the ruler with item response theory: Increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index', *Journal of Family Psychology*, 21 (4), 572-83.

GAO (2010), 'Factors Affecting the Financial Literacy of Individuals with Limited English Proficiency', *GAO Highlights*, GAO-10-518 (May).

Gerardi, Kristopher, Goette, Lorenz, and Meier, Stephan (2010), 'Financial Literacy and Subprime Mortgage Delinquency: Evidence from a Survey Matched to Administrative Data', *Federal Reserve Bank of Atlanta Working Paper Series*, 2010 (10).

Hilgert, Marianne, Hogarth, Jeanne, and Beverly, Sandra (2003), 'Houshold Financial Management: The Connection Between Knowledge and Behavior', *Federal Reserve Bulletin*, 89, 309-22.

Hrung, Warren B. and Seligman, Jason (2011), 'Responses to the Financial Crisis, Treasury Debt, and the Impact on Short-Term Money Markets', *Federal Reserve Bank of New York Staff Report*, 481.

Huston, Sandra J. (2010), 'Measuring Financial Literacy', *Journal of Consumer Affairs*, 44 (2), 296-316.

Kaplan, D and Venezky, R. L. (1993), 'What can employers assume about the literacy skills of GED graduates?', (Philadelphia, PA: National Center on Adult Literacy).

Knoll, Melissa and Houts, Carrie (2011), 'A Standardized Scale to Measure Financial Knowledge', (Social Security Administration Working Paper).

Kozup, John and Hogarth, Jeanne M. (2008), 'Financial Literacy, Public Policy, and Consumers' Self-Protection—More Questions, Fewer Answers', *Journal of Consumer Affairs*, 42 (2), 127-36.

Lachance, Marie E. and Seligman, Jason S. (2011), 'Involuntary Retirement: Prevalence, Causes and Impacts', (The Ohio State University).

Lusardi, Annamaria (2008), 'Planning and Financial Literacy: How Do Women Fare?', *The American Economic Review*, 98 (2), 413-17.

Macdonald, Paul and Paunonen, Sampo V. (2002), 'A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory versus Classical Test Theory', *Educational and Psychological Measurement*, 62 (6), 921-43.

Mandell, Lewis and Schmid Klein, Linda (2007), 'Motivation and financial literacy', *Financial services review*, 16 (2), 105.

Meier, Stephan and Sprenger, Charles (2007), 'Selection into Financial Literacy Programs: Evidence from a Field Study', *Federal Reserve Bank of Boston Discussion Paper Series*, 07 (5).

--- (2010), 'Present-Biased Preferences and Credit Card Borrowing', *American Economic Journal: Applied Economics*, 2 (1), 193-210.

Müller, Sebastian and Weber, Martin (2010), 'Financial Literacy and Mutual Fund Investments: Who Buys Actively Managed Funds?', *Schmalenbach Business Review, Forthcoming*.

Osgood, D. Wayne, McMorris, Barbara J., and Potenza, Maria T. (2002), 'Analyzing Multiple-Item Measures of Crime and Deviance I: Item Response Theory Scaling', *Journal of Quantitative Criminology*, 18 (3), 267-96.

Reise, Steven P., Widaman, Keith F., and Pugh, Robin H. (1993), 'Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance', *Psychological Bulletin*, 114 (3), 552-66.

Remund, David L. (2010), 'Financial Literacy Explicated: The Case for a Clearer Definition in an Increasingly Complex Economy', *Journal of Consumer Affairs*, 44 (2), 276-95.

Seligman, Jason (2010), 'Financial Literacy: Evidence from the Cusp of Retirement', *UW-Madison Center for Financial Security Working Paper*.

Seligman, Jason and Bose, Rana (2006), 'Pension Design, Retirement Wellbeing, and the Residual Public Burden', *Policy Notes* (Carl Vinson Institute of Government, University of Georgia).

Shealy, Robin and Stout, William (1991), 'A Procedure to Detect Test Bias Present Simultaneously in Several Items', *Defense Technical Information Center OAI-PMH Repository*.

Thissen, David, Steinberg, Lynne, and Gerrard, Meg (1986), 'Beyond group-mean differences: The concept of item bias', *Psychological Bulletin*, 99 (1), 118-28.

Thissen, David, et al. (1983), 'An Item Response Theory for Personality and Attitude Scales: Item Analysis Using Restricted Factor Analysis', *Applied Psychological Measurement*, 7 (2), 211-26.

Worthington, Andrew C. (2006), 'Predicting Financial Literacy in Australia", Financial Services Review', *Financial Services Review*, 15 (1), 59-79.

Table 1: Repeated questioning of subsamples in HRS modules: cell counts (N) 2002-2008

	numeracy (percents) 2002	risk 2002	stock 2004	risk 2004	annuity / social security 2004	retirement 2006	risk 2006	econ / invest 2008	financial literacy 2002	financial literacy 2004	financial literacy 2006	financial literacy 2008
numeracy (pcnts) 2002	1,061											
risk 2002	0	1,244										
stock 2004	0	0	1,007									
risk 2004	74	93	74	1,235								
annuity / soc sec 2004	54	61	56	0	1,058							
retirement 2006	87	11	77	126	105	1,562						
risk 2006	11	796	7	116	105	0	1,404					
econ/investments 2008	96	95	75	44	35	127	120	1,517				
financial literacy 2002	890	1,072	1,048	881	504	1,048	911	976	14,648			
financial literacy 2004	3	7	8	196	371	277	279	303	71	3,388		
financial literacy 2006	887	1,057	860	1,113	966	1,550	1,393	1,439	11,902	2,949	17,570	
financial literacy 2008	3	6	1	4	9	9	9	8	34	48	80	124

Table 2a: Three 'easy' question variants with different rates of success – all have ramifications for asset evolution

(HRS 2002 Numeracy Module)

What is 15% of 1,000?

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 82	incorrect:	$71,872	$68,619	$79,734	$68,992	$105,962	$134,359	$154,837	$128,886	$215,584
N: 245	correct:	$203,243	$261,998	$381,009	$500,025	$386,706	$450,985	$514,786	$1,678,268	$613,046
N $a: 56	***25%***	*35%*	*26%*	*21%*	*14%*	*27%*	*30%*	*30%*	*8%*	*35%*
	Total	$173,281	$215,320	$309,520	$401,719	$325,453	$375,853	$428,132	$1,283,880	$513,680

A pill cures 15% of people who have a disease. If 1000 people have the disease and they all take the pill, how many people will be cured?

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 62	incorrect:	$119,037	$120,667	$84,504	$94,040	$107,919	$96,861	$185,576	$165,204	$127,050
N: 249	correct:	$289,460	$336,879	$376,859	$395,403	$463,214	$441,952	$558,255	$1,343,113	$644,785
N $a: 74	***26%***	*41%*	*36%*	*22%*	*24%*	*23%*	*22%*	*33%*	*12%*	*20%*
Total		$228,108	$259,042	$267,226	$285,446	$335,308	$316,464	$419,122	$891,312	$472,206

A store is offering a 15% off sale on all TVs. The most popular television is normally priced at $1000. How much money would a customer save on the television during this sale?

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 70	incorrect:	$ 99,179	$ 112,806	$ 205,727	$ 137,204	$ 136,647	$ 151,432	$ 157,461	$ 228,994	$ 222,183
N: 367	correct:	$ 249,442	$ 361,436	$ 318,555	$ 404,461	$ 484,564	$ 453,435	$ 497,028	$ 533,968	$ 569,878
N $a: 101	***38%***	*40%*	*31%*	*65%*	*34%*	*28%*	*33%*	*32%*	*43%*	*39%*
	Total	$ 201,762	$ 282,544	$ 283,839	$ 322,228	$ 378,825	$ 359,415	$ 388,366	$ 435,589	$ 455,297

Table 2b: Three 'difficult' question variants with different rates of success

(HRS 2002 Numeracy Module)

**The number 10 is what percent of 1,000?*

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
						(percentages are for incorrect:correct)				
% right: 45	incorrect	$203,171	$329,887	$281,347	$367,731	$331,972	$303,577	$389,409	$386,995	$347,940
N: 240	correct	$298,850	$280,562	$338,998	$300,489	$453,941	$399,841	$519,907	$1,922,808	$657,838
N $a: 62	***76%***	*68%*	*118%*	*83%*	*122%*	*73%*	*76%*	*75%*	*20%*	*53%*
		$244,176	$309,202	$305,669	$339,799	$380,372	$343,564	$441,608	$1,001,320	$468,456

**If the chance of getting a disease is 10 in 1,000, what percent of people will get the disease?*

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
						(percentages are for incorrect:correct)				
% right: 28	incorrect	$170,411	$185,837	$188,500	$208,328	$280,054	$286,914	$407,566	$414,955	$408,170
N: 265	correct	$263,134	$298,708	$419,183	$413,676	$517,603	$510,405	$574,477	$980,404	$708,948
N $a: 54	***56%***	*65%*	*62%*	*45%*	*50%*	*54%*	*56%*	*71%*	*42%*	*58%*
Total		$195,699	$216,620	$251,414	$264,332	$346,040	$346,778	$448,507	$561,972	$492,388

**If a customer saved $10 off a $1000 chair, what percent would the customer have saved off the original price?*

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
						(percentages are for incorrect:correct)				
% right: 54	incorrect	$129,328	$181,814	$216,165	$210,316	$250,385	$251,652	$240,045	$303,501	$284,545
N: 363	correct	$210,295	$294,250	$367,656	$472,263	$428,793	$505,119	$486,073	$1,247,193	$642,958
N $a: 117	***50%***	*61%*	*62%*	*59%*	*45%*	*58%*	*50%*	*49%*	*24%*	*44%*
Total		$174,234	$243,654	$301,052	$361,613	$351,685	$390,851	$375,569	$817,476	$480,344

Table 2c: Evidence of consistency in difficulty with the medical/disease frame

(HRS 2002 Numeracy Module)

*Which of the following percentages is the biggest: One percent, ten percent, or five percent?

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 91	incorrect	$75,308	$221,000	$168,950	$222,425	$136,267	$114,298	$59,750	$285,261	$211,123
N: 319	correct	$191,055	$216,619	$215,517	$271,554	$303,419	$319,419	$371,332	$423,676	$442,625
N $a: 77	***57%***	*39%*	*102%*	*78%*	*82%*	*45%*	*36%*	*16%*	*67%*	*48%*
Total		$185,120	$216,840	$213,034	$268,968	$294,735	$309,163	$355,353	$416,194	$429,764

*Which of the following percentages represents the biggest risk of getting a disease: One percent, ten percent, or five percent?

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 69	incorrect	$152,979	$256,693	$176,914	$170,527	$198,445	$193,704	$151,689	$240,312	$199,841
N: 307	correct	$154,428	$227,159	$313,406	$467,799	$345,068	$399,270	$383,114	$1,236,372	$507,825
N $a: 78	***57%***	*99%*	*113%*	*56%*	*36%*	*58%*	*49%*	*40%*	*19%*	*39%*
Total		$153,957	$236,757	$269,594	$377,489	$299,953	$334,091	$313,987	$930,914	$419,163

*Which of the following percentages represents the biggest discount in a sale: One percent, ten percent, or five percent?

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 86	incorrect	$120,506	$129,062	$140,647	$107,517	$231,048	$207,799	$218,737	$242,604	$281,572
N: 389	correct	$248,457	$284,279	$332,743	$334,266	$393,461	$387,786	$466,690	$883,421	$503,545
N $a: 118	***46%***	*49%*	*45%*	*42%*	*32%*	*59%*	*54%*	*47%*	*27%*	*56%*
Total		$234,710	$267,603	$310,517	$308,457	$374,513	$367,300	$438,661	$812,219	$480,510

Table 2d: Evidence of difficulty with a different sort of frame – the lottery

(HRS 2002 Numeracy Module)

**Which of the following is the most likely to happen: something that happens 1 in 100 times, something that happens 1 in 1000 times, or something that happens 1 in 10 times?*

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 73	incorrect	$182,897	$192,274	$203,663	$207,658	$248,512	$198,726	$301,397	$330,628	$260,129
N: 295	correct	$243,268	$264,768	$348,123	$480,279	$419,968	$473,706	$506,370	$1,769,226	$599,245
N $a: 83	***52%***	*75%*	*73%*	*59%*	*43%*	*59%*	*42%*	*60%*	*19%*	*43%*
Total		$229,612	$248,371	$317,888	$420,048	$384,850	$413,653	$463,351	$1,496,074	$535,088

**Which of the following represents the biggest risk of getting a disease: a 1 in 100 risk, a 1 in 1000 risk, or a 1 in 10 risk*

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 75	incorrect	$173,598	$156,084	$163,719	$178,608	$246,978	$249,660	$266,963	$351,880	$324,964
N: 290	correct	$168,745	$212,397	$221,921	$243,126	$266,624	$273,242	$365,059	$401,962	$447,135
N $a: 66	***82%***	*103%*	*73%*	*74%*	*73%*	*93%*	*91%*	*73%*	*88%*	*73%*
Total		$169,904	$198,949	$208,022	$227,946	$261,861	$267,693	$340,535	$390,224	$418,145

**Which of the following represents the biggest chance of winning a lottery: a 1 in 100 chance, a 1 in 1000 chance, or a 1 in 10 chance?*

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 60	incorrect	$151,451	$239,313	$264,173	$229,130	$294,527	$262,516	$274,519	$284,146	$344,969
N: 411	correct	$210,793	$271,833	$273,563	$327,856	$376,264	$392,230	$407,175	$488,784	$464,323
N $a: 122	***75%***	*72%*	*88%*	*97%*	*70%*	*78%*	*67%*	*67%*	*58%*	*74%*
Total		$187,245	$259,030	$269,791	$289,013	$344,624	$341,561	$355,407	$412,724	$422,007

Table 3: A relatively 'dumb' questions as distinguished from one with potential for evidencing latent vulnerability

**You should put all your money into the safest investment you can find and accept whatever return it pays.*

asset evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
				(percentages are for incorrect:correct)						
% 'True': 68	("True")	$272,302	$261,752	$321,549	$349,999	$430,023	$454,754	$632,515	$526,243	$500,975
N: 1,457	("False")	$247,351	$268,176	$270,048	$359,698	$398,866	$347,267	$495,596	$713,974	$560,679
	106%	*110%*	*98%*	*119%*	*97%*	*108%*	*131%*	*128%*	*74%*	*89%*
Total	|	$255,293	$266,153	$286,238	$356,655	$408,467	$382,371	$539,143	$653,537	$542,011

Sample evolution:		1992	1994	1996	1998	2000	2002	2004	2006	2008
N: Assets |	("True")	113	108	105	107	102	113	111	113	111
N: Assets |	("False")	242	235	229	234	229	233	238	238	244
Total	|	355	343	334	341	331	346	349	351	355

biennial change in assets:		92-94	94-96	96-98	98-00	00-02	02-04	04-06	06-08
incorrect	("True")	-2%	25%	3%	24%	-5%	38%	-17%	-6%
correct	("False")	6%	0%	34%	10%	-12%	47%	44%	-20%
	46%	*-35%*	*.*	*9%*	*235%*	*39%*	*80%*	*-37%*	*32%*
Total	|	4%	8%	23%	15%	-10%	44%	21%	-17%

Table 4: Reverse frame effects – evidence of agreeability and guessing | agreeability in context (bottom frame)

HRS (2008 Econ / Investment Decisions Module)

**If you start out with $1,000 and earn an average return of 10% per year for 30 years, after compounding, the initial $1,000 will have grown to [more] than $6,000*

	year	1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 81	incorrect	$154,078	$180,610	$228,505	$241,127	$209,542	$265,481	$219,131	$299,418	$300,621
N: 610	correct	$322,384	$256,072	$304,813	$394,647	$397,139	$387,964	$587,234	$542,897	$551,842
N $a :147	***58%***	*48%*	*71%*	*75%*	*61%*	*53%*	*68%*	*37%*	*55%*	*54%*
Total	|	$292,487	$243,495	$290,939	$368,360	$366,320	$367,274	$524,202	$503,414	$507,217

**If you start out with $1,000 and earn an average return of 10% per year for 30 years, after compounding, the initial $1,000 will have grown to [less] than $6,000*

year	year	1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% right: 63	incorrect	$251,371	$220,516	$211,585	$242,086	$221,265	$302,638	$536,127	$414,468	$316,751
N: 617	correct	$206,397	$327,409	$298,467	$424,062	$574,559	$465,787	$611,725	$758,165	$784,227
N $a :144	***67%***	*122%*	*67%*	*71%*	*57%*	*39%*	*65%*	*88%*	*55%*	*40%*
Total	|	$221,082	$292,761	$270,541	$365,112	$461,915	$410,654	$586,696	$644,374	$628,402

**I understand the stock market reasonably well.*

	year	1992	1994	1996	1998	2000	2002	2004	2006	2008
					(percentages are for incorrect:correct)					
% 'True': 27	false	$177,305	$224,247	$234,518	$296,692	$326,862	$290,238	$421,077	$533,834	$430,765
N: 1,502	true	$468,295	$384,289	$421,060	$522,062	$613,517	$624,192	$847,585	$960,184	$830,470
N $a: 353	***52%***	*38%*	*58%*	*56%*	*57%*	*53%*	*46%*	*50%*	*56%*	*52%*
Total	|	$254,050	$266,652	$284,153	$356,102	$404,656	$380,546	$538,157	$648,712	$538,378

Table 5a: Impact of repeating a financial literacy question on odds of success for a relatively easy question: 2004, 2006, 2008

Chance of Getting a Disease (10 percent of 1,000 (100))

	N	Percent Correct	Std Dev	
(2004)	3,201	83%	38%	*overall*
	2,251	89%	31%	*if correct in future (2006)*
	421	58%	49%	*if incorrect in future (2006)*
(2006)	15,808	77%	42%	*overall*
	2,672	84%	36%	*if second time asked*
	2,249	89%	31%	*if previously correct (2004)*
	423	59%	49%	*if previously incorrect (2004)*
(2008)	117	84%	37%	*overall*
	44	80%	41%	*if never before asked*
	36	89%	32%	*if third time asked*
	2	50%	71%	*if previously always incorrect*
	65	88%	33%	*if ever correct in past*
	31	90%	30%	*if always correct in past*

> Weak evidence of learning; evidence of guessing!

Table 5b: Impact of repeating a financial literacy question on odds of success for a -medium difficulty- question: 2004, 2006, 2008

Lottery Winnings Split Five Ways (20 percent of $2 million ($400,000))

	N	Percent Correct	Std Dev	
(2004)	2,932	57%	49%	*overall*
	1,396	71%	45%	*if correct in future (2006)*
	851	45%	50%	*if incorrect in future (2006)*
(2006)	13,665	53%	50%	*overall*
	2,247	62%	49%	*if second time asked*
	1,375	72%	45%	*if previously correct (2004)*
	872	46%	50%	*if previously incorrect (2004)*
(2008)	100	47%	50%	*overall*
	41	37%	49%	*if never before asked*
	30	60%	50%	*if third time asked*
	2	50%	71%	*if previously always incorrect*
	45	53%	50%	*if ever correct in past*
	17	71%	47%	*if always correct in past*

> Weak evidence of learning, further evidence of guessing!

Table 5c: Impact of repeating a financial literacy question on odds of success for a - difficult- question: 2004, 2006, 2008

Interest on Savings ($200 compounded at 110 percent annually for two years ($242))

	N	Percent Correct	Std Dev	
(2004)	2,778	14%	35%	*overall*
	363	44%	50%	*if correct in future (2006)*
	1,710	11%	31%	*if incorrect in future (2006)*
(2006)	12,082	12%	33%	*overall*
	2,073	18%	38%	*if second time asked*
	344	46%	50%	*if previously correct (2004)*
	1,729	12%	32%	*if previously incorrect (2004)*
(2008)	98	8%	28%	*overall*
	42	2%	15%	*if never before asked*
	29	14%	35%	*if third time asked*
	19	21%	42%	*if previously always incorrect*
	12	0%	0%	*if ever correct in past*
	2	0%	0%	*if always correct in past*

> Some evidence of learning, further evidence of guessing!

Table 6a: Summary statistics for variables employed in regression analysis – reporting on fullest univariate sample

Variable	Obs	Mean	SD	Min	Max
Dependent Variables:					
2008 Total Wealth	17,339	$484,275.80	$1,308,997.00	-$1,064,000.00	$38,100,000.00
2006-08 Change in Total Wealth	16,603	-$45,899.30	$1,948,118.00	-$92,500,000.00	$26,400,000.00
2008 Total Non-Housing Wealth	17,339	$317,445.70	$1,026,511.00	-$1,144,000.00	$31,600,000.00
2006-08 Change in Total Non-Housing Wealth	16,603	-$37,510.43	$1,787,095.00	-$92,100,000.00	$25,400,000.00
2008 Retirement Satisfaction	8,283	1.54	0.64	1.00	3.00
2006-08 Change in Retirement Satisfaction	6,248	0.00	0.58	-2.00	2.00
2008 Retirement Satisfaction relative to years just ahead	1,833	2.41	1.49	1.00	5.00
2006-08 Change in Retirement Satisfaction relative to years just ahead	729	0.06	1.43	-4.00	4.00
Financial Literacy Measures					
Numeracy (pcnts) 2002	1,031	2.44	1.18	0.00	4.00
Social Security Knowledge (2004)	1,055	0.69	0.67	0.00	2.00
Economics & Investing (2008)	1,447	4.74	1.73	0.00	8.00
" Diversification subsection	1,490	3.10	1.25	0.00	5.00
" Other questions	1,451	1.60	0.94	0.00	3.00
Financial Literacy (2002, 2006)	6,154	3.51	1.07	2.00	6.00
" Percentage question (medical frame)	10,648	2.27	1.08	0.00	3.00
" Division question (lottery frame)	8,001	1.67	1.25	0.00	3.00
" Compound interest (savings frame)	7,126	0.38	0.84	0.00	3.00
Included Controls:					
Defined Contribution (plan1)	48,103	0.20	0.40	0.00	1.00
Risk Aversion	6,494	3.32	1.04	1.00	4.00
Level of Education	30,641	2.93	1.44	1.00	5.00
Birth Year	30,686	1933	13	1890	1983
Female	48,103	0.36	0.48	0.00	1.00
African Heritage	48,103	0.10	0.29	0.00	1.00
Other Non-White Heritage	48,103	0.03	0.17	0.00	1.00
Hispanic	30,659	0.09	0.28	0.00	1.00
Ever Applied for Disability	28,404	0.15	0.35	0.00	1.00

Table 6b: Regressions employing 2008 Total Wealth as Dependent

Dependent:	2008 Total Wealth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Financial Literacy Measures:							
Numeracy (pcnts) 2002	27,322						
4 questions: (0-4)	(44,065)						
Social Security Knowledge (2004)		514,408					
2 questions: (0-2)		(370,317)					
Economics & Investing (2008)			54,702				
8 questions: (0-8)			(55,364)				
" Diversification subsection				**117,427***			
5:8 questions: (0-5)				(70,221)			
" Other questions					-35,822		
3:8 questions: (0-5)					(104,147)		
Financial Literacy (2002, 2006)						**459,419*** **	
3 questions: repeated in 2002, 2006 (0-6)						(167,600)	
" Percentage question (medical frame)							179,028
learning (0=never, 1=correct02; 2=correct06; 3=both)							(275,458)
" Division question (lottery frame)							198,217
learning (0=never, 1=correct02; 2=correct06; 3=both)							(158,441)
" Compound interest (savings frame)							**419,109** **
learning (0=never, 1=correct02; 2=correct06; 3=both)							(198,441)
Included Controls:							
Defined Contribution (plan1)	142,442	448,292	-155,029	-194,645	-101,967	-318,148	-323,740
binary	(97,687)	(553,214)	(204,339)	(197,606)	(201,010)	(384,900)	(386,858)
Risk Aversion	7,460	-392,029*	104,764	106,180	88,647	19,149	4,257
(1=least risk adverse, ..., 4=most risk adverse)	(51,852)	(222,449)	(101,581)	(94,786)	(101,290)	(166,682)	(167,115)
Level of Education	101,434**	111,565	171,076**	184,044**	195,940**	191,503	182,284
(1<HS,2=GED, 3=HS, 4=some college 5=BA+)	(39,566)	(185,404)	(80,220)	(75,501)	(83,184)	(158,405)	(160,757)
Birth Year	-14,593*	-51,319	-1,842	-931.0	-2,062	-27,849	-28,089
continuous	(7,926)	(51,562)	(17,729)	(16,243)	(17,858)	(32,953)	(33,114)
Female	245,474	-1.259e+06*	306,507	274,205	249,684	-47,379	-21,083
binary, (1=yes)	(161,486)	(637,072)	(256,298)	(230,816)	(258,647)	(431,780)	(437,450)
African Heritage	-92,154	491,293	-363,308	-333,943	-352,208	-131,332	-164,777
binary (1=yes)	(133,183)	(734,523)	(228,109)	(215,895)	(230,187)	(615,432)	(616,281)
Other Non-White Heritage	6,137	).672e+06***	106,689	115,171	67,513	3.908e+06***	4.001e+06***
binary (1=yes)	(231,440)	(1.423e+06)	(375,767)	(359,483)	(376,413)	(818,410)	(825,050)
Hispanic	48,413	682,039	-128,610	-106,584	-180,638	-2.055e+06**	-2.143e+06**
binary (1=yes)	(139,262)	(841,216)	(349,408)	(310,895)	(353,033)	(982,697)	(986,953)
Ever Applied for Disability	-37,764	-229,775	-301,186	-285,741	-329,135	-353,538	-355,991
binary (1=yes)	(142,809)	(652,513)	(271,660)	(253,379)	(273,298)	(566,552)	(568,562)
Constant	2.796e+07*	1.014e+08	2.821e+06	937,763	3.581e+06	5.290e+07	5.396e+07
	(1.551e+07)	(1.003e+08)	(3.441e+07)	(3.157e+07)	(3.467e+07)	(6.412e+07)	(6.443e+07)
Observations	76	112	82	85	82	342	342
R-squared	0.265	0.399	0.174	0.197	0.164	0.109	0.108

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6c: Regressions employing 2006-08 Change in Total Wealth as Dependent

Dependent:	2006-08 Change in Total Wealth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Financial Literacy Measures:							
Numeracy (pcnts) 2002	-13,571						
4 questions: (0-4)	(34,519)						
Social Security Knowledge (2004)		459,178					
2 questions: (0-2)		(324,048)					
Economics & Investing (2008)			38,056				
8 questions: (0-8)			(82,296)				
" Diversification subsection				**30,656**			
5 :8 questions: (0-5)				(105,876)			
" Other questions					49,812		
3:8 questions: (0-5)					(154,005)		
Financial Literacy (2002, 2006)						**215,757***	
3 questions: repeated in 2002, 2006 (0-6)						(116,940)	
" Percentage question (medical frame)							111,370
learning (0=never, 1=correct02; 2=correct06; 3=both)							(192,341)
" Division question (lottery frame)							123,255
learning (0=never, 1=correct02; 2=correct06; 3=both)							(110,633)
" Compound interest (savings frame)							**134,158**
learning (0=never, 1=correct02; 2=correct06; 3=both)							(138,563)
Included Controls:							
Defined Contribution (plan1)	17,478	483,037	-367,158	-363,793	-344,178	-105,589	-97,572
binary	(76,526)	(484,093)	(303,744)	(297,941)	(297,237)	(268,556)	(270,128)
Risk Aversion	-10,694	-314,446	-35,214	-26,429	-42,615	-52,010	-57,473
(1=least risk adverse, ..., 4=most risk adverse)	(40,620)	(194,655)	(150,997)	(142,914)	(149,780)	(116,299)	(116,690)
Level of Education	-3,804	-30,901	-25,232	-8,709	-26,517	68,812	71,018
(1<HS,2=GED, 3=HS, 4=some college 5=BA+)	(30,995)	(162,239)	(119,245)	(113,837)	(123,006)	(110,524)	(112,250)
Birth Year	-1,854	-28,457	-10,076	-12,525	-9,591	-7,769	-8,019
continuous	(6,209)	(45,120)	(26,354)	(24,491)	(26,407)	(22,992)	(23,122)
Female	-137,899	-1.248e+06**	-45,759	-17,706	-52,022	-287,602	-291,723
binary, (1=yes)	(126,505)	(557,472)	(380,980)	(348,014)	(382,468)	(301,266)	(305,454)
African Heritage	-56,698	733,167	143,946	118,550	138,296	52,156	31,969
binary (1=yes)	(104,333)	(642,747)	(339,079)	(325,517)	(340,383)	(429,406)	(430,324)
Other Non-White Heritage	20,412	-1.422e+06**	148,792	113,514	128,809	3.593e+06***	3.609e+06***
binary (1=yes)	(181,306)	(1.245e+06)	(558,568)	(542,013)	(556,610)	(571,029)	(576,099)
Hispanic	16,068	605,513	108,059	196,256	106,762	-1.693e+06**	-1.716e+06**
binary (1=yes)	(109,095)	(736,110)	(519,386)	(468,754)	(522,037)	(685,658)	(689,150)
Ever Applied for Disability	111,654	-26,016	1.123e+06**	1.074e+06**	1.127e+06**	-117,301	-121,085
binary (1=yes)	(111,874)	(570,984)	(403,816)	(382,034)	(404,131)	(395,300)	(397,004)
Constant	3.752e+06	5.663e+07	1.994e+07	2.468e+07	1.912e+07	1.471e+07	1.534e+07
	(1.215e+07)	(8.778e+07)	(5.115e+07)	(4.760e+07)	(5.126e+07)	(4.474e+07)	(4.499e+07)
Observations	76	112	82	85	82	342	342
R-squared	0.038	0.382	0.126	0.120	0.125	0.129	0.128

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6d: Regressions employing 2008 Total Non-Housing Wealth as Dependent

Dependent:	2008 Total Non-Housing Wealth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Financial Literacy Measures:							
Numeracy (pcnts) 2002	3,043						
4 questions: (0-4)	(34,723)						
Social Security Knowledge (2004)		313,933					
2 questions: (0-2)		(240,437)					
Economics & Investing (2008)			39,633				
8 questions: (0-8)			(53,184)				
" Diversification subsection				**98,204**			
5 :8 questions: (0-5)				(67,576)			
" Other questions					-49,061		
3:8 questions: (0-5)					(99,668)		
Financial Literacy (2002, 2006)						**202,821****	
3 questions: repeated in 2002, 2006 (0-6)						(92,765)	
" Percentage question (medical frame)							112,404
learning (0=never, 1=correct02; 2=correct06; 3=both)							(152,509)
" Division question (lottery frame)							119,427
learning (0=never, 1=correct02; 2=correct06; 3=both)							(87,722)
" Compound interest (savings frame)							**139,094**
learning (0=never, 1=correct02; 2=correct06; 3=both)							(109,868)
Included Controls:							
Defined Contribution (plan1)	100,040	237,661	-170,830	-210,667	-128,077	-61,409	-56,007
binary	(76,977)	(359,186)	(196,296)	(190,163)	(192,365)	(213,038)	(214,187)
Risk Aversion	2,404	-243,773*	83,429	85,340	70,572	-2,092	-6,077
(1=least risk adverse, ..., 4=most risk adverse)	(40,860)	(144,430)	(97,583)	(91,216)	(96,934)	(92,257)	(92,524)
Level of Education	70,031**	77,693	125,457	135,248*	149,218*	158,332*	158,125*
(1<HS,2=GED, 3=HS, 4=some college 5=BA+)	(31,178)	(120,378)	(77,062)	(72,657)	(79,606)	(87,675)	(89,004)
Birth Year	-7,132	-31,364	2,268	3,242	1,912	-15,910	-16,348
continuous	(6,246)	(33,478)	(17,031)	(15,632)	(17,090)	(18,239)	(18,334)
Female	121,096	-833,763**	235,555	211,602	184,098	-31,645	-34,401
binary, (1=yes)	(127,252)	(413,633)	(246,210)	(222,123)	(247,523)	(238,986)	(242,197)
African Heritage	-75,672	340,917	-336,276	-311,530	-324,099	-199,199	-211,023
binary (1=yes)	(104,949)	(476,905)	(219,131)	(207,764)	(220,287)	(340,635)	(341,208)
Other Non-White Heritage	3,286	-1.168e+06***	20,825	30,781	-9,808	2.450e+06***	2.468e+06***
binary (1=yes)	(182,376)	(923,910)	(360,977)	(345,944)	(360,224)	(452,982)	(456,794)
Hispanic	69,561	383,235	-230,606	-202,584	-279,093	-1.408e+06**	-1.430e+06***
binary (1=yes)	(109,739)	(546,178)	(335,655)	(299,186)	(337,849)	(543,913)	(546,433)
Ever Applied for Disability	-61,962	-212,730	-249,121	-230,164	-274,053	-209,163	-208,663
binary (1=yes)	(112,534)	(423,658)	(260,968)	(243,837)	(261,544)	(313,581)	(314,788)
Constant	1.362e+07	6.205e+07	-4.973e+06	-6.990e+06	-4.006e+06	3.027e+07	3.123e+07
	(1.222e+07)	(6.513e+07)	(3.306e+07)	(3.038e+07)	(3.318e+07)	(3.549e+07)	(3.567e+07)
Observations	76	112	82	85	82	342	342
R-squared	0.189	0.391	0.126	0.145	0.122	0.130	0.130

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6e: Regressions employing 2006-08 Change in Total Non-Housing Wealth as Dependent

Dependent:	2006-08 Change in Total Non-Housing Wealth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Financial Literacy Measures:							
Numeracy (pcnts) 2002	8,961						
4 questions: (0-4)	(29,927)						
Social Security Knowledge (2004)		295,312					
2 questions: (0-2)		(196,009)					
Economics & Investing (2008)			43,927				
8 questions: (0-8)			(81,856)				
" Diversification subsection				**36,598**			
5 :8 questions: (0-5)				(105,454)			
" Other questions					52,856		
3:8 questions: (0-5)					(153,245)		
Financial Literacy (2002, 2006)						**-5,468**	
3 questions: repeated in 2002, 2006 (0-6)						(72,093)	
" Percentage question (medical frame)							66,338
learning (0=never, 1=correct02; 2=correct06; 3=both)							(118,099)
" Division question (lottery frame)							53,737
learning (0=never, 1=correct02; 2=correct06; 3=both)							(67,929)
" Compound interest (savings frame)							**-108,385**
learning (0=never, 1=correct02; 2=correct06; 3=both)							(85,079)
Included Controls:							
Defined Contribution (plan1)	-32,283	287,042	-377,330	-380,372	-349,938	158,297	175,990
binary	(66,345)	(292,815)	(302,118)	(296,756)	(295,770)	(165,564)	(165,860)
Risk Aversion	-9,402	-186,667	-6,766	-2,159	-15,545	-56,606	-52,402
(1=least risk adverse, ..., 4=most risk adverse)	(35,216)	(117,742)	(150,189)	(142,345)	(149,041)	(71,698)	(71,648)
Level of Education	-5,794	-38,629	-17,658	3,347	-17,988	62,274	73,166
(1<HS,2=GED, 3=HS, 4=some college 5=BA+)	(26,872)	(98,134)	(118,606)	(113,384)	(122,399)	(68,137)	(68,922)
Birth Year	2,964	-16,307	-9,712	-11,222	-9,191	1,348	911.8
continuous	(5,383)	(27,292)	(26,213)	(24,394)	(26,276)	(14,175)	(14,197)
Female	-98,185	-787,063**	5,522	22,456	-3,773	-211,315	-239,195
binary, (1=yes)	(109,675)	(337,201)	(378,941)	(346,629)	(380,580)	(185,729)	(187,551)
African Heritage	-54,349	454,745	124,086	102,998	118,394	-65,718	-67,166
binary (1=yes)	(90,453)	(388,782)	(337,264)	(324,221)	(338,703)	(264,727)	(264,222)
Other Non-White Heritage	-4,793	-1.943e+06**	99,342	48,152	75,824	2.139e+06***	2.089e+06***
binary (1=yes)	(157,185)	(753,189)	(555,578)	(539,856)	(553,863)	(352,037)	(353,729)
Hispanic	24,801	330,336	27,720	153,740	24,056	-1.083e+06**	-1.050e+06**
binary (1=yes)	(94,581)	(445,255)	(516,606)	(466,889)	(519,461)	(422,705)	(423,142)
Ever Applied for Disability	55,224	-26,072	1.105e+06**	1.036e+06**	1.111e+06**	4,230	3,285
binary (1=yes)	(96,990)	(345,374)	(401,655)	(380,514)	(402,137)	(243,701)	(243,763)
Constant	-5.667e+06	3.257e+07	1.905e+07	2.198e+07	1.819e+07	-2.558e+06	-2.022e+06
	(1.053e+07)	(5.309e+07)	(5.088e+07)	(4.741e+07)	(5.101e+07)	(2.758e+07)	(2.762e+07)
Observations	76	112	82	85	82	342	342
R-squared	0.039	0.372	0.120	0.111	0.118	0.119	0.125

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6f: Regressions employing 2008 Retirement Satisfaction as Dependent

Dependent:	2008 Retirement Satisfaction (1=very, 2=somewhat, 3=not at all)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Financial Literacy Measures:							
Numeracy (pcnts) 2002	-0.147						
4 questions: (0-4)	(0.143)						
Social Security Knowledge (2004)		0.0634					
2 questions: (0-2)		(0.126)					
Economics & Investing (2008)			0.0130				
8 questions: (0-8)			(0.0476)				
" Diversification subsection				-0.00821			
5 :8 questions: (0-5)				(0.0551)			
" Other questions					0.0596		
3:8 questions: (0-5)					(0.0833)		
Financial Literacy (2002, 2006)						**0.00476**	
3 questions: repeated in 2002, 2006 (0-6)						(0.0468)	
" Percentage question (medical frame)							-0.0131
learning (0=never, 1=correct02; 2=correct06; 3=both)							(0.0662)
" Division question (lottery frame)							-0.00941
learning (0=never, 1=correct02; 2=correct06; 3=both)							(0.0421)
" Compound interest (savings frame)							**0.0173**
learning (0=never, 1=correct02; 2=correct06; 3=both)							(0.0540)
Included Controls:							
Defined Contribution (plan1)	0.0180	-0.474**	0.237	0.250*	0.210	0.0765	0.0739
binary	(0.281)	(0.184)	(0.142)	(0.137)	(0.145)	(0.100)	(0.102)
Risk Aversion	0.000443	0.0447	0.128	0.120	0.128	0.0338	0.0311
(1=least risk adverse, ..., 4=most risk adverse)	(0.209)	(0.0960)	(0.0877)	(0.0875)	(0.0851)	(0.0551)	(0.0556)
Level of Education	-0.160	0.0213	-0.0893	-0.0838	-0.104	-0.0816*	-0.0823*
(1<HS,2=GED, 3=HS, 4=some college 5=BA+)	(0.101)	(0.0654)	(0.0628)	(0.0602)	(0.0656)	(0.0441)	(0.0449)
Birth Year	-0.00408	-0.0189	0.0263	0.0262	0.0257	0.000860	0.000459
continuous	(0.0330)	(0.0355)	(0.0227)	(0.0227)	(0.0225)	(0.0157)	(0.0159)
Female	-0.839	0.118	-0.391**	-0.389**	-0.368**	-0.175	-0.169
binary, (1=yes)	(0.541)	(0.236)	(0.161)	(0.162)	(0.163)	(0.120)	(0.122)
African Heritage	-1.930***	-0.0418	0.00764	-0.0130	-0.0135	-0.173	-0.176
binary (1=yes)	(0.633)	(0.253)	(0.168)	(0.172)	(0.161)	(0.177)	(0.181)
Other Non-White Heritage	0	0.314	-0.479	-0.484	-0.407	0.268	0.279
binary (1=yes)	(0)	(0.537)	(0.282)	(0.283)	(0.300)	(0.211)	(0.215)
Hispanic	0	-0.0923	0.629**	0.614*	0.588*	-0.119	-0.135
binary (1=yes)	(0)	(0.303)	(0.292)	(0.298)	(0.292)	(0.299)	(0.304)
Ever Applied for Disability	1.740***	0.775***	0.965***	0.960***	0.948***	0.842***	0.842***
binary (1=yes)	(0.494)	(0.201)	(0.195)	(0.196)	(0.194)	(0.135)	(0.137)
Constant	11.12	38.01	-49.99	-49.65	-48.90	-0.122	0.730
	(63.90)	(68.84)	(43.94)	(44.01)	(43.54)	(30.42)	(30.80)
Observations	23	47	34	34	34	129	129
R-squared	0.637	0.494	0.763	0.763	0.768	0.339	0.340

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

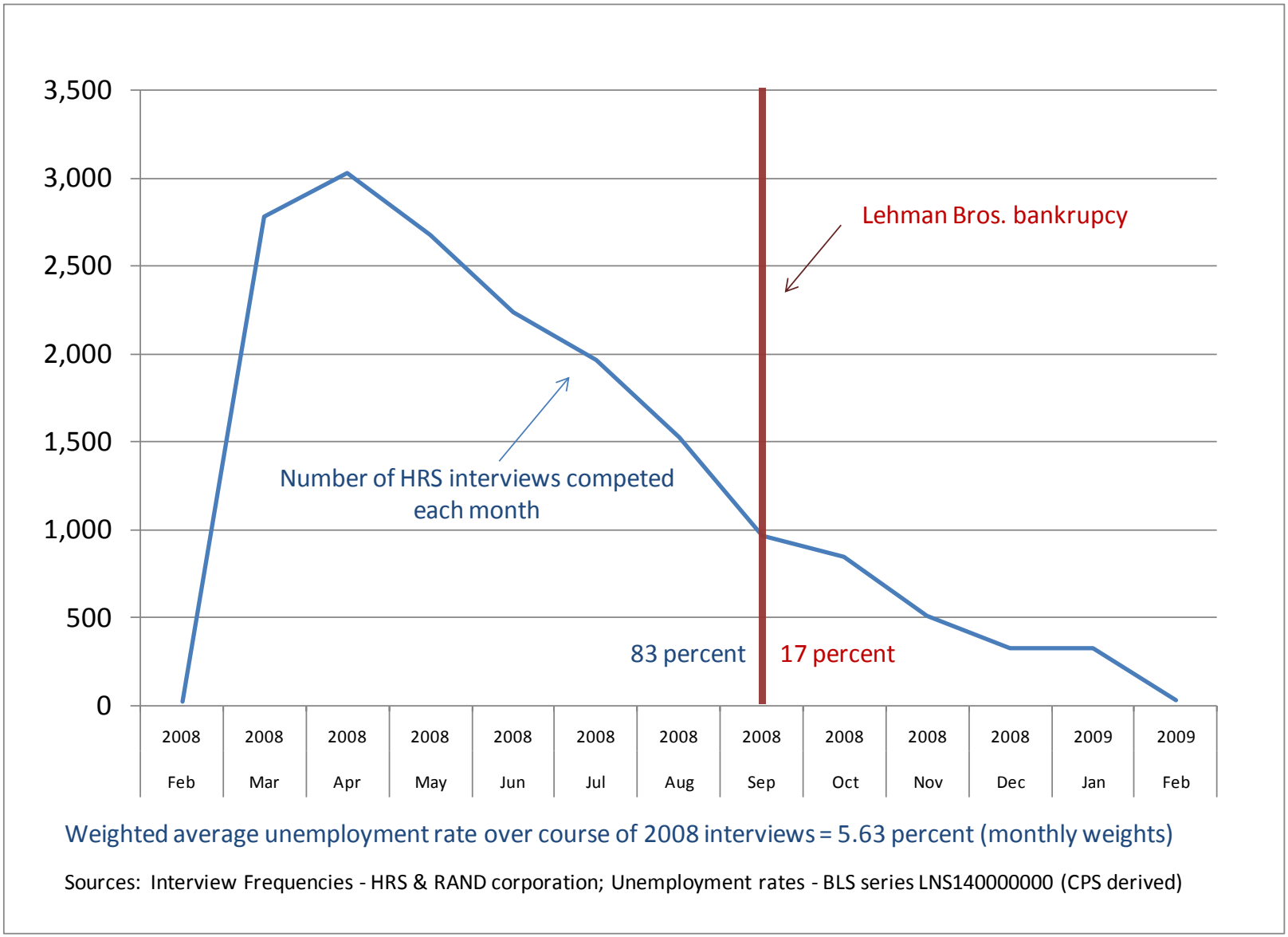
Table 6g: Regressions employing 2006-08 Change in Retirement Satisfaction as Dependent

Dependent:	2006-08 Change in Retirement Satisfaction (1=very, 2=somewhat, 3=not at all)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Financial Literacy Measures:							
Numeracy (pcnts) 2002	0						
4 questions: (0-4)	(0)						
Social Security Knowledge (2004)		0.0708					
2 questions: (0-2)		(0.185)					
Economics & Investing (2008)			-0.205				
8 questions: (0-8)			(0.140)				
" Diversification subsection				-0.0462			
5 :8 questions: (0-5)				(0.195)			
" Other questions					-0.270		
3:8 questions: (0-5)					(0.171)		
Financial Literacy (2002, 2006)						0.0927	
3 questions: repeated in 2002, 2006 (0-6)						(0.0604)	
" Percentage question (medical frame)							**0.0314**
learning (0=never, 1=correct02; 2=correct06; 3=both)							(0.0879)
" Division question (lottery frame)							0.107*
learning (0=never, 1=correct02; 2=correct06; 3=both)							(0.0577)
" Compound interest (savings frame)							0.0292
learning (0=never, 1=correct02; 2=correct06; 3=both)							(0.0695)
Included Controls:							
Defined Contribution (plan1)	-0	-0.717**	0.429	0.239	0.362	0.210	0.227*
binary	(0)	(0.300)	(0.278)	(0.276)	(0.252)	(0.128)	(0.130)
Risk Aversion	0	0.0186	-0.205	-0.0892	-0.0574	0.0365	0.0468
(1=least risk adverse, ..., 4=most risk adverse)	(0)	(0.144)	(0.268)	(0.304)	(0.244)	(0.0689)	(0.0693)
Level of Education	0	-0.00324	-0.0843	-0.189	-0.0964	0.0572	0.0565
(1<HS,2=GED, 3=HS, 4=some college 5=BA+)	(0)	(0.0995)	(0.169)	(0.168)	(0.162)	(0.0562)	(0.0564)
Birth Year	-0	-0.0210	0.106*	0.113*	0.132**	-0.0449**	-0.0465**
continuous	(0)	(0.0506)	(0.0525)	(0.0605)	(0.0519)	(0.0221)	(0.0223)
Female	-1	0.219	-0.0996	0.0370	-0.272	-0.0456	-0.0871
binary, (1=yes)	(0)	(0.321)	(0.352)	(0.387)	(0.384)	(0.150)	(0.154)
African Heritage	0	-0.449	-0.666*	-0.547	-0.367	0.0179	-0.00784
binary (1=yes)	(0)	(0.381)	(0.347)	(0.438)	(0.330)	(0.281)	(0.283)
Other Non-White Heritage	0	0.311	0	0	0	0.0531	0.0469
binary (1=yes)	(0)	(0.688)	(0)	(0)	(0)	(0.226)	(0.229)
Hispanic	0	-0.355	0	0	0	-0.0563	-0.0279
binary (1=yes)	(0)	(0.386)	(0)	(0)	(0)	(0.390)	(0.391)
Ever Applied for Disability	0	0.241	-0.224	-0.0665	0.0187	0.186	0.203
binary (1=yes)	(0)	(0.275)	(0.427)	(0.493)	(0.398)	(0.162)	(0.164)
Constant	1	41.07	-203.9*	-218.8*	-255.7**	86.47**	89.48**
	(0)	(98.09)	(101.4)	(117.2)	(100.1)	(42.93)	(43.35)
Observations	14	36	20	20	20	89	89
R-squared	1.000	0.336	0.567	0.485	0.578	0.111	0.131

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Figure 1: Timing of HRS 2008 wave interviews, financial crisis and average unemployment rate as weighted by interview frequencies



Using the Right Yardstick: Assessing Financial Literacy Measures by way of Financial Well-Being

Center for Financial Security Working Paper 2011-CFS.2

Center for Financial Security
School of Human Ecology
University of Wisconsin-Madison
Sterling Hall Mailroom B605
475 North Charter Street
Madison, Wisconsin 53706
608.262.6766
cfs@mailplus.wisc.edu

cfs.wisc.edu