Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Same Finances, Different Scores: Unexplained Dispersion in the CFPB Financial Well-Being Scale

Shay Tsaban

shay.tsaban@gmail.com

Peres Academic Center

Research Article

Keywords: financial well-being, measurement validity, household finance, scale validation, consumer financial protection, survey measurement

Posted Date: September 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10542396/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The Consumer Financial Protection Bureau's Financial Well-Being Scale is the principal instrument for describing the financial condition of American households. Its scores compare demographic groups and help direct financial education. This study asks how much of the score a household's observed financial position accounts for. In the Bureau's own National Financial Well-Being Survey (n = 6,389) and its Making Ends Meet Survey fielded in January 2025 (n = 2,623), thirty-two indicators of financial position together with demographic controls explain just over half the variance in the score. Since the instrument is itself highly reliable, roughly forty percent of the variance it reliably carries remains unaccounted for. Households in the same decile of predicted financial position receive scores that differ by seventeen to twenty-eight points on a hundred-point scale. The pattern survives a saturated categorical specification and gradient boosting, so it is not an artefact of assuming linearity. The unexplained component is not noise either, since it tracks temperament and general life satisfaction. Its consequences fall unevenly across the comparisons the scale supports. Income and education gaps are almost entirely attributable to financial position, whereas about forty percent of the age gap is not, an asymmetry that appears in both surveys. Comparisons drawn from the score therefore carry different meanings, and different degrees of confidence, according to which groups they compare.

Introduction

Household financial condition is now measured, tracked and compared using standardised subjective instruments. The Consumer Financial Protection Bureau introduced its Financial Well-Being Scale to give researchers, practitioners and public agencies a common metric (Consumer Financial Protection Bureau, 2015). The scale has since become a fixture of household finance research, and the Bureau uses it to report the financial condition of the American population and of subgroups within it (Consumer Financial Protection Bureau, 2017b). Its appeal lies in producing a single number comparable across people and over time.

That comparability carries an assumption. When two groups receive different average scores, readers take the difference to be a difference in financial well-being, a construct the Bureau defines by reference to a person's financial situation and financial capability (Consumer Financial Protection Bureau, 2015). Group differences in the score are therefore treated as evidence about circumstances rather than about persons. Reports of lower scores among younger adults, or among some racial and ethnic groups, are read as statements about the financial position those groups occupy.

The validation literature has not examined that assumption directly. Validation work on the scale has concentrated on its internal structure, its performance in new populations, and its association with financial outcomes (Chen et al., 2026; Totenhagen et al., 2026; Khashadourian & Harrison, 2024). These are the right questions for establishing whether an instrument measures something coherently and predicts what it should. They do not establish how much of the score a household's observed financial

position accounts for, nor what occupies the remainder. An instrument can rank groups correctly on average while assigning very different scores to households that are financially alike.

The distinction matters because of the work the scale performs. When a score compares two demographic groups, or identifies who should receive financial education, a broadly correct ranking is not sufficient. The reported difference must also reflect the financial circumstances the instrument is meant to capture. If a substantial share of the difference between two groups reflects something other than financial position, then a comparison presented as a statement about circumstances is partly a statement about something else.

This study estimates that share, and two questions organise the analysis. First, how much of the variance in the score does a household's observed financial position account for, and how much dispersion remains among households whose financial positions are alike? Second, does the unexplained component fall evenly across the group comparisons the scale supports, or does it concentrate in some and not others?

The analysis uses the Bureau's own public data. The National Financial Well-Being Survey provides the ten-item scale together with an unusually detailed record of household financial circumstances and a set of dispositional, affective and health measures. The Making Ends Meet Survey, fielded in January 2025, provides an independent replication using the five-item version of the scale in a different sample drawn on a different frame.

The paper makes three contributions. The first is to quantify how much of the score financial position accounts for, which proves to be close to half in both surveys. The second is to show that unreliability cannot explain the remainder, since the instrument's own reliability sets a ceiling on explainable variance that the models fall well short of. The third is to show that the remainder falls unequally across group comparisons, so that the interpretive risk depends on which comparison a reader draws. Table 1 and Figure 1 carry the central evidence, and a reader can check every claim below against them.

The study responds to a call now being made within the field itself for closer attention to how measures in consumer finance are developed, chosen and applied (Warmath et al., 2026a; Warmath et al., 2026b). The most widely used measure is a reasonable place to begin, and the data its own developer released are the appropriate evidence.

The paper proceeds as follows. The next section positions the study within the validation literature on the scale and specifies what the analysis adds. The third section describes the two surveys, the measures and the analytic strategy. The fourth reports results, beginning with the group comparisons. The fifth discusses how the score should be read and how future comparisons should be reported, and states the study's limitations.

Background

Objective and subjective measurement of household financial condition

Research on household financial condition has long faced a choice between objective and subjective measurement. Objective approaches record what a household has and owes. They support indices of material hardship, of economic insecurity, and of financial fragility, each of which has been shown to capture something distinct about household circumstances (Iceland & Bauman, 2007; Osberg & Sharpe, 2014; Lusardi et al., 2011). Their appeal is that they do not depend on how a respondent feels about the situation being recorded.

Objective measures also have known limits. They are demanding to collect and sensitive to recall error, and on their own they do not indicate whether a given balance sheet leaves a household secure. Two households with identical savings may face very different obligations and very different prospects. The same amount of debt can be manageable or ruinous depending on income stability, which is itself difficult to observe (Gathergood, 2012).

Subjective approaches ask households directly. The InCharge Financial Distress and Financial Well-Being Scale was an early instrument of this kind (Prawitz et al., 2006). Later work established perceived financial well-being as a construct in its own right, distinguishing current money management stress from expected future financial security, and connected it to overall well-being (Netemeyer et al., 2018; Brüggen et al., 2017). The Bureau's scale belongs to this tradition and was developed to give the field a common and publicly available instrument.

Attempts to bridge the two traditions show how hard the reconciliation is. Comerton-Forde et al. (2022) construct parallel reported and observed financial well-being scales for customers of a large Australian bank, using survey responses matched to that bank's own records. Building both scales from the same households, rather than relying on either alone, acknowledges that the two do not substitute for one another. Whether the choice among available scales changes the substantive picture has itself become a question in the literature (Heck et al., 2024).

The instrument and the construct it targets

The Bureau defines financial well-being as a state in which a person can meet current and ongoing obligations, feel secure in their financial future, and make choices that allow enjoyment of life (Consumer Financial Protection Bureau, 2015). The scale operationalises that definition through ten items answered on five-point response scales, scored using an item response theory model that returns a value on a nominal zero to one hundred range (Consumer Financial Protection Bureau, 2017a). An abbreviated five-item version exists for surveys with limited space.

The construct is deliberately indirect in what it measures. Rather than recording assets or income, it targets the security and freedom of choice that a financial situation and a person's financial capability together produce. That framing is deliberate, and it responds to a long-standing observation that

objective financial indicators alone map poorly onto how households experience their circumstances (Netemeyer et al., 2018; Brüggen et al., 2017).

The framing also creates the interpretive question this study addresses. If the definition ties the construct to the financial situation, then that situation should substantially determine the score. If instead circumstance and stable characteristics of the person jointly determine the score, then two households in identical circumstances will receive different scores, and comparisons across groups that differ in those characteristics will not be comparisons of circumstance alone.

What validation has and has not established

Researchers have examined the scale repeatedly and generally favourably. Work in this journal has assessed the psychometric properties of financial well-being measures in diverse samples and across cultural contexts. Similar scrutiny now covers related instruments in consumer finance (Totenhagen et al., 2026; Chen et al., 2026; Reiter et al., 2025; Sorgente et al., 2025). This body of work establishes that the instrument behaves coherently and generalises reasonably well.

A smaller literature has compared the scale against objective benchmarks. Khashadourian and Harrison (2024) construct a typology of four financial well-being categories from household financial ratios, derive it through discriminant analysis under a model they call the equilibrium model of the household, and report that higher scale scores are associated with higher category ranks. They note that most financial well-being scales rest on perceptions because objective markers have not reliably captured the construct, and they argue that claims about the scale measuring something beyond traditional financial measures may be imprecise. Comerton-Forde et al. (2022) take a different route, building parallel reported and observed financial well-being scales from survey responses matched to bank records for customers of a large Australian bank.

These studies establish agreement in the aggregate, which is a genuine finding that this study does not dispute. Rank agreement between a score and a categorical summary of financial position is nonetheless compatible with wide dispersion within categories. A measure can order groups correctly while assigning scores that differ substantially among households placed in the same financial category. The two facts are not in tension, and both need to be established before anyone can read the score confidently as a statement about circumstances.

This study therefore takes up the question that agreement analyses leave open. Rather than asking whether the score and financial position agree on average, it asks how much of the score financial position accounts for, and characterises what remains.

Grounds for expecting a systematic remainder

There are grounds for expecting a substantial and systematic remainder. Subjective assessments of a domain are shaped by stable characteristics of the assessor as well as by the domain itself. Life satisfaction judgments show considerable stability and return toward person-specific levels after circumstances change (Fujita & Diener, 2005; Diener et al., 2009; Lyubomirsky et al., 2005). Three decades of research on subjective well-being have established that such judgments depend jointly on circumstance and on the person making them (Diener et al., 1999). Dispositional optimism and negative affectivity shape how people evaluate their situations across domains, and optimism in particular has been linked to evaluations of well-being in several domains at once (Scheier & Carver, 1987; Watson & Clark, 1984; Conversano et al., 2010).

The financial domain supplies further reason to expect it. The relationship between money and evaluated well-being is neither simple nor a matter of absolute resources alone. Evaluations depend on relative standing and on comparison groups as well as on absolute resources (Boyce et al., 2010; Ferrer-i-Carbonell, 2005; Clark et al., 2008). The association between income and well-being differs depending on whether the outcome is an evaluation of life or an emotional state, and the shape of that association has been actively contested (Kahneman & Deaton, 2010; Killingsworth, 2021; Killingsworth et al., 2023). Reports about one's own circumstances are further shaped by how questions are posed and by what comparisons are made salient (Schwarz, 1999; Deaton, 2008).

There is also a causal channel running in the other direction. Financial strain consumes attention and impairs cognitive function, and debt is associated with psychological distress (Mani et al., 2013; Shah et al., 2012; Gathergood, 2012). A person-level component in the score may therefore be partly a consequence of financial circumstances rather than an independent disposition. The analysis below cannot separate these directions, and does not attempt to.

If these regularities operate on the financial domain, the score will carry a component attributable to the respondent rather than to the household balance sheet. Whether that component is small enough to ignore is an empirical question, and the analysis below answers it.

Comparison as the practical stake

The practical consequences of these interpretive questions emerge most clearly in group comparison. The Bureau reports scores by age, race and ethnicity, income and other characteristics, and such comparisons inform where attention and resources go (Consumer Financial Protection Bureau, 2017b). Comparable national reporting on household financial circumstances follows a similar pattern (Board of Governors of the Federal Reserve System, 2025). Financial education programmes are evaluated in part against measures of this kind, and the evidence on their downstream effects is mixed enough that measurement quality matters for what those evaluations can support (Fernandes et al., 2014; Kaiser et al., 2022). Financial literacy itself varies systematically across the same demographic groups the score compares, and varies unevenly by age and education (Lusardi & Mitchell, 2014; Lusardi et al., 2010; Lusardi et al., 2017).

Age deserves particular attention among these comparisons. The relationship between age and general well-being has attracted extensive study and is not monotone, with evidence for a U-shaped pattern across the life cycle in many countries (Blanchflower & Oswald, 2008). If a similar pattern operates on financial self-assessment independently of financial circumstances, then age comparisons using the score require more care than comparisons along dimensions where no such pattern appears.

Research published in this journal has examined financial well-being from several directions. These include the financial capability perspective, structural determinants, debt delinquency and financial stress, economic hardship, transitions into parenthood, and change during the COVID-19 period (Anvari-Clark & Ansong, 2022; Fan & Henager, 2021; Xiao & Kim, 2021; Sun et al., 2022; Sun & Wang, 2026; Kelley et al., 2022; Birkenmaier et al., 2022). That work relies on measures of the kind examined here. How much of the score is attributable to circumstance is therefore consequential for the field, not only for the instrument.

Recent work in this journal has argued for building equity considerations into construct, scale and model development (Caraballo, 2026). The analysis below bears directly on that argument. If the unexplained component of the score falls unevenly across groups, then the interpretive burden of a comparison depends on which groups it involves, and a single caveat applied uniformly to all comparisons will be wrong in both directions.

Data and Method

Two national surveys

The primary analysis uses the National Financial Well-Being Survey public use file, collected for the Bureau in late 2016 from a probability-based online panel of United States adults (Consumer Financial Protection Bureau, 2017c). The file contains 6,394 respondents. Five have no valid scale score, leaving 6,389 for analysis. The survey includes an oversample of adults aged 62 and older, and the file supplies a single analysis weight that accounts for unequal selection probabilities and adjusts the sample to population benchmarks. All estimates below are weighted accordingly.

The replication uses the Making Ends Meet Survey, Sample 6 Wave 1, fielded in January 2025 (Consumer Financial Protection Bureau, 2025). The file contains 2,630 respondents, of whom 2,623 have a valid score. This survey is drawn from a credit-record panel, which allows nonresponse adjustment at the individual level rather than only against external demographic benchmarks (Beaumont, 2008). The five-item version of the scale is used. Weight selection follows the published guidance, which distinguishes weights for items asked of all respondents from weights for items appearing only in the online instrument.

The two surveys differ in year, sampling frame, mode, instrument length and nonresponse adjustment. Agreement between them is therefore informative in a way that agreement between two analyses of one file would not be.

Measures

The outcome is the scale score as distributed in each file. In the 2016 file the ten items are also available individually, which permits direct estimation of the instrument’s reliability.

Financial position. The objective block comprises thirty-two indicators covering the domains the construct definition names. These include the ability to make ends meet, savings held, capacity to absorb an unexpected expense, capacity to cover costs, contact from a collections agency, income volatility, benefit receipt, household income, housing tenure and value, mortgage status, six indicators of material hardship, two indicators of credit denial, borrowing behaviour, and twelve indicators of adverse events experienced in the preceding year. The flags for legitimate skips and for genuine nonresponse described below add three further columns, so the fitted block carries thirty-five predictors. Predictor counts throughout are counts of fitted columns. In the 2025 file the corresponding block comprises fifty-eight indicators covering income and its variability, account balances, emergency savings, months of expenses covered, investment and retirement holdings, tenure and housing payments, rent delinquency and eviction threat, unexpected expenses, difficulty paying bills, benefit receipt and transfers.

Temperament, affect and health. Ten measures from the 2016 file are used, chosen to exclude anything financial or behavioural: two outlook items, positive emotionality, psychological distress, three self-control items, social connectedness, general health and memory difficulty. One nonresponse flag brings the fitted block to eleven columns. The exclusion is deliberate, because measures such as frugality, saving habit, planning horizon and goal confidence describe financial dispositions and behaviours, and cannot arbitrate between a circumstance account and a person account of the score. They are reported only as a sensitivity analysis.

Demographics. Age category, education, gender, race and ethnicity, marital status, household size and employment status, entered as indicator variables.

Discriminant measure. The 2016 file includes three general life satisfaction items in the tradition of the Satisfaction With Life Scale, averaged to form a measure of subjective well-being that is not domain-specific (Diener et al., 1985).

In the 2025 file the five scale items appear among the survey variables. They are excluded from every predictor block, since including any part of the instrument among its own predictors would be circular.

Common method variance

Outcome, financial position and temperament are all self-reported within a single instrument, which raises the possibility that common method variance inflates the associations reported below (Podsakoff et al., 2003). Two features of the design limit the concern. First, the paper’s central claim is about variance the model fails to explain, and common method variance would work against that claim by inflating explained variance rather than in favour of it. Second, the objective block consists largely of

concrete behavioural and event reports, such as whether a collections agency made contact, rather than of attitudinal items sharing a response format with the outcome. The concern applies more directly to the association between the residual and the temperament block, which is reported as descriptive and is not essential to the main result.

Handling of nonsubstantive codes and missing data

The 2016 file distinguishes several nonsubstantive codes, including one indicating that a respondent was correctly skipped because the item did not apply. This is structural rather than missing, and it is modelled as an explicit not-applicable indicator rather than imputed. Genuine nonresponse is handled by median substitution accompanied by a missing indicator, which preserves the analytic sample and makes the adjustment visible. The missing-data literature prefers multiple imputation to single substitution of this kind, and for estimating individual coefficients that preference is well founded (Schafer & Graham, 2002; Rubin, 1987). The quantity at issue here is explained variance rather than any particular coefficient, and the complete-case analysis reported below, which discards every case carrying genuine nonresponse, returns the same answer.

Both files supply weights constructed to correct for unequal selection and for differential nonresponse. Weighting of this kind adjusts for observable differences between respondents and nonrespondents, and its effectiveness depends on how strongly nonresponse is related to the variables used in the adjustment (Groves, 2006; Groves & Peytcheva, 2008). The 2025 survey is unusual in being drawn from a credit-record panel, which allows that adjustment to use observed characteristics of nonrespondents rather than external benchmarks alone.

Analytic strategy

Estimation is by weighted least squares with heteroskedasticity-robust standard errors. The analysis compares five specifications: financial position alone, temperament alone, demographics alone, financial position with demographics, and all three blocks together. A sixth adds the excluded block of financial attitudes and behaviours, and is reported as a sensitivity check rather than as a preferred model.

Weighted variances are computed in population form. The bias correction that some implementations apply by default is designed for reliability weights rather than survey weights, and it distorts estimates most inside small subgroups such as deciles, which is precisely where this analysis requires them.

Because a linear specification could understate what financial position explains, the main result is also estimated under a saturated categorical specification and under gradient boosting. Those comparisons are evaluated by ten-fold cross-validation, with dispersion computed from out-of-fold predictions, so that added flexibility cannot inflate fit mechanically. Confidence intervals for the main quantities come from 2,000 bootstrap replications. The public use file supplies a single analysis weight and no replicate weights, so the resampling treats observations as independent, and the intervals do not reflect any clustering in the panel from which the sample was drawn.

Because the blocks are correlated and differ in size, incremental R-squared understates the contribution of whichever block is entered last. Explained variance is therefore also apportioned by Shapley decomposition, which averages each block's marginal contribution over all orderings of entry and so divides shared variance evenly rather than assigning it by arbitrary precedence (Grömping, 2007; Grömping, 2006; Fryer et al., 2020).

To assess dispersion, households are sorted into deciles of predicted financial position and the realised scores within each decile examined. Results appear as the mean within-decile standard deviation relative to the total, and as the tenth to ninetieth percentile spread.

For the group comparisons, predicted financial position is estimated from the objective block alone. Including demographic controls in that prediction would remove age and racial differences by construction, since those variables would then appear on both sides of the comparison.

The interpretation of unexplained variance depends on how much of it could be measurement error. Two bounds are used. First, the instrument's reliability places a ceiling on the variance any set of predictors could explain, so reliable variance left unexplained can be computed directly. Second, attenuation from error in the predictors is bounded by asking what average reliability the objective block would need for the residual to be attributable to that source alone.

Results

Group comparisons carry different meanings

Table 1 reports, for each demographic variable, the difference between its highest and lowest scoring category, and the corresponding range across categories once the part of the score predicted by financial position is removed. The two ranges are taken separately, so the residual range need not fall between the same pair of categories as the raw gap. A variant holding that pair fixed is reported among the robustness checks.

Table 1
Group gaps in the score before and after financial position, 2016 survey. Confidence intervals are from 2,000 bootstrap replications.

Group variable	Raw gap (points)	Unexplained by position (points)	Share surviving	95% CI on share
Household income	15.9	1.8	11.5%	[8.8, 17.8]
Education	13.3	2.4	17.9%	[13.3, 23.3]
Age	10.0	4.1	41.1%	[36.5, 52.5]
Race and ethnicity	4.6	2.4	51.2%	[22.8, 83.4]

The income and education gradients behave as the construct definition implies. Households with higher incomes and more education report higher financial well-being, and almost all of that difference is accounted for by their financial positions. Only 11.5 percent of the income gap and 17.9 percent of the education gap remain once position is removed, and both intervals are narrow. For these comparisons the score functions as intended.

The age comparison behaves differently. The raw gap of 10.0 points falls to 4.1 points, so 41.1 percent survives, with a confidence interval running from 36.5 to 52.5 percent. The interval excludes the income and education estimates entirely. Something other than measured financial position accounts for a substantial part of the age gradient in the score, and the estimate is precise enough to rely on.

The racial comparison points the same way but is not precisely estimated. The point estimate is 51.2 percent, but the confidence interval runs from 22.8 to 83.4 percent. Two features of the data produce this. The raw gap is small at 4.6 points, so the ratio is sensitive to the numerator, and the racial and ethnic categories available in the public use file are coarse. The defensible conclusion is that these data cannot characterise the racial comparison with confidence, not that half of it is unexplained. The comparison is reported because its direction agrees with the age result, and because the imprecision is itself worth knowing to anyone contemplating it.

The asymmetry between the income and education comparisons on one side and the age comparison on the other is the paper's central practical finding. A caveat about the score's interpretation that is appropriate for the age comparison would be excessive for the income comparison, and a reassurance appropriate for the income comparison would be misleading for the age comparison.

The 2025 survey shows the same asymmetry. Conditioning on that file's objective block, 51.2 percent of the age gap survives against 6.9 percent of the education gap. The two surveys therefore agree on the ordering of the comparisons as well as on the overall dispersion, which matters because they differ in year, frame, mode and instrument length. That file carries no income measure comparable to the 2016 one. Its racial comparison gives 44.0 percent, which is reported for completeness and is subject to the same imprecision as its 2016 counterpart, for which no interval was computed.

Financial position explains about half the score

Table 2 reports the model comparisons that underlie the decomposition.

Table 2
Variance in the score explained, 2016 survey.

Model	Predictors	R^2	Adjusted R^2
M1 Financial position only	35	0.526	0.524
M2 Temperament only	11	0.351	0.350
M3 Demographics only	31	0.197	0.193
M4 Financial position and demographics	66	0.553	0.548
M5 M4 with temperament	77	0.614	0.610
M6 M5 with financial attitudes	93	0.657	0.652

Thirty-two indicators of financial position explain 52.6 percent of the variance in the score. Adding demographics raises this to 55.3 percent. Ten measures of temperament, affect and health explain 35.1 percent on their own, and raise the combined model to 61.4 percent.

Entered last, temperament adds 0.061 while financial position adds 0.190. These figures understate temperament, because entering a block last credits it with none of the variance it shares with blocks already present. The Shapley decomposition in Table 3 apportions that shared variance evenly.

Table 3
Shapley decomposition of explained variance, 2016 survey.

Block	Shapley R^2	Share of explained variance
Financial position	0.339	55.1%
Temperament, affect and health	0.187	30.4%
Demographics	0.089	14.5%

Financial position is the largest single contributor and remains so under every allocation rule considered. It does not, however, account for everything the score records. Roughly thirty percent of the explained variance is attributable to ten measures that contain no financial content.

Households with similar finances receive very different scores

Figure 1 Realised scale scores within deciles of predicted financial position, 2016 and 2025 surveys. Grey bars show the tenth to ninetieth percentile range of realised scores within each decile; the line marks the median.

Table 4
Score dispersion within deciles of predicted financial position.

Decile	2016 p10	2016 median	2016 p90	2016 spread	2025 p10	2025 median	2025 p90	2025 spread
1	23.0	37.0	49.0	26.0	19.0	31.0	45.0	26.0
2	33.0	45.0	55.0	22.0	24.0	36.0	48.9	24.9
3	38.0	48.0	56.0	18.0	26.0	40.0	51.0	25.0
4	42.0	51.0	59.0	17.0	29.0	43.0	53.0	24.0
5	47.0	56.0	68.0	21.0	33.0	45.0	55.0	22.0
6	50.0	60.0	72.0	22.0	36.1	49.0	58.0	21.9
7	52.0	62.0	75.0	23.0	40.1	51.0	65.0	24.9
8	53.0	63.0	76.0	23.0	44.0	54.0	67.9	23.9
9	55.0	66.0	83.0	28.0	48.0	60.0	72.0	24.0
10	58.0	69.0	86.0	28.0	52.2	65.0	82.0	29.8

The median rises steadily across deciles, reproducing the agreement earlier work has documented. The distributions around those medians overlap heavily. In the 2016 survey the total weighted standard deviation of the score is 13.74 points and the mean within-decile standard deviation is 9.27 points, so 67 percent of the dispersion survives conditioning on financial position. Within-decile spreads between the tenth and ninetieth percentiles range from 17.0 to 28.0 points.

The 2025 survey reproduces this closely. The total standard deviation is 14.60 points and the mean within-decile standard deviation is 9.51 points, leaving 65 percent. Spreads range from 21.9 to 29.8 points. The agreement is notable given that the two surveys differ in year, frame, mode and instrument length.

These dispersion figures follow closely from the model fit and are not independent evidence. For a linear model the ratio of residual to total standard deviation is the square root of one minus R-squared, which here gives 0.669 against the 0.675 observed. The decile presentation is reported for two other reasons. It expresses the residual on the scale a reader of the score actually sees, in points rather than as a share of variance, and it shows that the dispersion is present at every level of financial position rather than concentrated at one end of the distribution.

The remainder is not measurement error

The ten items of the 2016 instrument yield a Cronbach’s alpha of 0.916, once the six negatively worded items are reverse coded (Cronbach, 1951). The value is high and consistent with the Bureau’s own development work (Consumer Financial Protection Bureau, 2017a). The distributed score is produced by

an item response theory model rather than by summing items, so alpha computed on the items estimates the reliability of the item set rather than of the scored value itself. For a unidimensional set of this kind the two are closely related, and the argument below does not turn on the exact figure. The argument requires only that reliability not fall close to 0.553, the fit the models achieve, since the ceiling disappears at that point and not before.

Reliability places a ceiling on explainable variance, since no set of predictors can explain more of a measure than the measure reliably contains. Table 5 sets the observed fit against that ceiling.

Table 5
The measurement-error bound, 2016 survey.

Quantity	Value
Reliability of the score	0.916
R^2 of financial position and demographics	0.553
Reliable variance left unexplained	0.363
Share of reliable variance unexplained	40%

Roughly forty percent of the score's reliable variance is not accounted for by measured financial position. Because the instrument is reliable, this residual cannot be attributed to noise in the outcome. The fit entering that calculation is estimated in sample, which favours the model. Substituting the cross-validated fit of 0.540 obtained in the robustness checks below raises the unexplained share from 40 to 41 percent, so the figure given here is the more conservative of the two.

The alternative attribution is error in the predictors, which attenuates fit, and that possibility can be bounded approximately. Treating the objective block as a single composite, the classical attenuation relation makes the observable fit the product of that composite's reliability and the fit attainable without predictor error. Setting the second of these at the reliability ceiling of 0.916 and the first at the observed 0.553 puts the required reliability at 0.604. For the residual to be attributable to predictor error alone, then, some two fifths of the variance in items such as whether a household was contacted by a collections agency, or whether it could cover an unexpected expense, would have to be noise. Items of this kind are behavioural and concrete, and reliability that low is not plausible for them. The attenuation relation holds exactly only for a single predictor, so the figure is an approximation rather than a bound in the strict sense.

What the remainder tracks

If the unexplained component were noise, it would correlate with nothing. Table 6 reports the associations that characterise it.

Table 6
Associations of the unexplained component, 2016 survey.

Correlation	r
Raw score with general life satisfaction	0.417
Residual with general life satisfaction	0.269
Residual with predicted temperament	0.310

The part of the score not explained by financial position and demographics correlates 0.310 with the temperament block's prediction, so it carries systematic person-level content.

It also correlates 0.269 with general life satisfaction, which raises the discriminant concern. A domain-specific measure should show weaker association with a general well-being measure once the domain itself is held fixed (Campbell & Fiske, 1959; Borsboom et al., 2004). The observed value is only moderately below the raw correlation of 0.417, which indicates that holding financial circumstances fixed removes less of the shared general well-being content than a clean domain separation would imply.

Robustness

Functional form

The models above enter ordinal financial indicators linearly. If the true relationship between financial position and the score is non-linear or involves interactions, some of the unexplained dispersion would be misspecification rather than a property of the instrument. This is the most serious alternative explanation for the results, and it can be tested directly.

Three specifications of financial position were compared: the linear model used above, a saturated model in which every ordinal financial indicator enters as a full set of category indicators, and gradient boosting on the raw indicators, which can represent arbitrary non-linearities and interactions. Because added flexibility raises in-sample fit mechanically, all three are evaluated by ten-fold cross-validation, with dispersion computed from out-of-fold predictions. Table 7 reports these three alongside a fourth model that adds the temperament block, which is discussed after the table.

Table 7
Functional form, evaluated by ten-fold cross-validation, 2016 survey.

Specification	Predictors	Cross-validated R^2	Share of SD surviving
Linear, as used above	66	0.540	0.684
Saturated categorical	122	0.586	0.654
Gradient boosting	26 raw	0.581	0.660
Gradient boosting with temperament	36 raw	0.630	0.619

Flexible functional form does help, and the paper's linear specification understates what financial position can explain. Cross-validated fit rises from 0.540 to 0.586, and the share of dispersion surviving falls from 68.4 to 65.4 percent. The improvement is genuine but modest. Under the best-fitting specification available, roughly two thirds of the score's dispersion still survives conditioning on financial position, and cross-validated fit remains far below the reliability ceiling of 0.916. Taking the saturated model as the benchmark, reliable variance left unexplained is 0.330 rather than 0.363, which is 36 percent of reliable variance rather than 40 percent. The conclusion does not depend on the linear approximation.

Adding the temperament block to the flexible model raises cross-validated fit further, to 0.630, which is consistent with the Shapley result that temperament carries systematic content the financial block does not.

On the 2025 file the comparison runs the other way. Cross-validated fit is 0.521 for the linear specification and 0.497 for gradient boosting, so the smaller sample does not support the additional flexibility. The linear specification therefore loses nothing on that file.

Differential item functioning

If the instrument functions differently for older respondents, the age asymmetry would be a property of the measure rather than of how older households evaluate their circumstances. The two cannot be separated with these data, but the item-level pattern bears on the question. Regressing each of the ten items on the objective block and comparing respondents aged 62 and over with those under 62 gives standardised differences with a mean of 0.071 and a standard deviation across items of 0.106, ranging from − 0.149 to 0.183, with eight of the ten positive. No single item carries the age effect, which would be the clearest signature of item-level malfunction. The dispersion across items is nonetheless of the same order as the average difference, so differential item functioning cannot be excluded. A formal invariance analysis of the instrument across age groups would settle the question and lies outside what this study attempts (Vandenberg & Lance, 2000; Meredith, 1993).

Definition of the group statistic

The share surviving in Table 1 is a ratio of two ranges, each maximised over categories separately, which biases it upward. Holding the pair of categories fixed at the two that produce the raw gap gives 38.5 percent for age, 4.2 percent for income and 14.3 percent for education. Under either definition the age comparison stands far above the other two, and under the fixed-pair definition the income and education gaps are accounted for by financial position more completely than Table 1 indicates. For race and ethnicity the fixed-pair figure is negative, at -12.0 percent, because the categories with the highest and lowest raw scores are not those with the highest and lowest residuals. That reversal is further evidence that these data cannot characterise the racial comparison, and it is reported for that purpose rather than as an estimate.

Precision

Two thousand bootstrap replications place the explained variance at 0.553 with a 95 percent interval of [0.541, 0.577], and the share of dispersion surviving at 0.675 with an interval of [0.657, 0.686]. The central quantities are precisely estimated. Confidence intervals for the group comparisons appear in Table 1, and only the racial comparison is imprecise.

Reliability estimator

The ceiling argument depends on the reliability of the instrument. Coefficient alpha is 0.916, the Spearman-Brown split-half estimate averaged over 200 random splits is 0.918 with an interval of [0.889, 0.933], and McDonald's omega from a one-factor solution is 0.930. Alpha is the most conservative of the three, so using it makes the smallest claim about unexplained reliable variance available from these estimators.

Weighting, subsamples and other choices

Estimating without weights gives an R^2 of 0.560 and 65.7 percent of dispersion surviving. Forming deciles from the objective block alone, without demographics, gives 68.4 percent. Restricting to complete cases, where every objective item is either answered or a legitimate skip, retains 5,730 of 6,389 respondents and yields an R^2 of 0.556 against 0.553, with 67 percent of dispersion surviving in both.

The result is stable across subsamples. Among respondents under 62 the surviving share is 66.7 percent and among those 62 and over it is 69.8 percent. In the bottom third of the income distribution it is 69.0 percent and in the top third 71.5 percent. These four values bracket the full-sample figure of 67.5 percent, the lowest falling 0.8 points below it and the highest 4.0 points above.

Weight choice is immaterial in the 2025 survey. The surviving share is 65.1 percent under the paper weights, 65.4 percent under the online weights, 65.1 percent under the combined weights, and 68.0 percent unweighted.

Adding the excluded block of financial attitudes and behaviours raises explained variance to 65.7 percent (M6 in Table 2). This is reported for completeness rather than as a preferred specification, since those measures describe financial dispositions and cannot separate a circumstance account from a person account.

Discussion

What the score records

The results support a conclusion that is specific and deliberately limited. The scale is not invalid, and it is not dominated by temperament. Financial position is its largest determinant under every allocation rule applied here, and the ordering of deciles is monotone in both surveys. Findings of agreement between the score and objective financial classifications are consistent with what is reported here (Khashadourian & Harrison, 2024).

The results add the magnitude of what agreement leaves out. Around forty percent of the score’s reliable variance is unaccounted for by an extensive record of financial circumstances, and that remainder is systematically associated with temperament and with general life satisfaction. It is large enough that households in the same financial decile differ by twenty points or more between the tenth and ninetieth percentiles.

The natural reading is that the score is a joint product of circumstance and person. That is arguably faithful to the construct as defined, since the definition refers to security and freedom of choice rather than to balance sheet quantities alone. The difficulty lies in what the score is then asked to do. A group difference used to decide where financial education or assistance should be directed is actionable only so far as it reflects circumstances that an intervention could change. Whether or not any particular user of the score holds a circumstance-only interpretation of it, that is the interpretation the targeting use requires, and the results above show that it does not hold uniformly across comparisons.

Implications for group comparison

The asymmetry in Table 1 has direct consequences. Comparisons of the score across income and education levels can be read much as they usually are, since financial position accounts for almost all of the observed gaps and both estimates are precise. The age comparison does not permit the same reading. Roughly forty percent of the age gap is not attributable to measured financial position, and the interval around that estimate excludes the income and education values.

This does not establish what the surviving portion is. Adding the temperament block to the model reduces the residual age gap only from 4.1 to 3.6 points, so the dispositional measures at hand cannot supply the explanation. Older respondents may evaluate similar circumstances differently for reasons this study cannot adjudicate, including differences in reference points, expectations formed over longer horizons, or aspects of financial position that the survey does not record, such as accumulated wealth or pension entitlements. The last of these is the most likely and the least testable with these data. Reporting the question as open is more useful than resolving it by assertion.

Race and ethnicity require a different kind of care. The point estimate resembles the age result, but the confidence interval spans from roughly a fifth to more than four fifths of the gap. The appropriate conclusion is that these data cannot say how much of the racial gap in the score is financial position. The result is not a null finding, and it does not license reading the racial gap as circumstance. The question requires data carrying finer racial and ethnic categories, and a larger sample than the public use file provides.

The recommendation that follows is narrow. When the score is used to compare groups, the comparison should be reported alongside an estimate of how much of the gap survives conditioning on financial position, with an interval attached. That quantity is computable from the same data used to produce the comparison, it differs enough across comparisons that a uniform caveat will misstate it, and its

precision differs too. This is a concrete instance of the case for building equity considerations into how measures are developed and applied (Caraballo, 2026).

The recommendation presumes measures of financial position alongside the score, which the Bureau's own surveys carry and many smaller studies do not. Where the available measures are thinner the quantity remains computable, and its interpretation shifts in a known direction. A thinner block conditions on less, so the share of a gap it leaves unexplained is an upper bound on what a fuller record of financial position would leave. Reported together with the measures it conditions on, the share is informative even when those measures are few.

Relation to the measurement agenda in consumer finance

Recent work has argued that the development, selection and use of measures is a core element of consumer finance research that receives little attention, and has sought consensus recommendations for improving measure-related decisions (Warmath et al., 2026a). The present study is an application of that argument to the field's most widely used instrument, using the data its developer released.

It also suggests an addition to the standard validation repertoire. Validation in this area has concentrated on internal structure, on performance in new populations, on invariance across groups, and on association with outcomes (Totenhagen et al., 2026; Chen et al., 2026; Vandenberg & Lance, 2000; Meredith, 1993). These are the established components of construct validation and each answers a real question (Cronbach & Meehl, 1955; Campbell & Fiske, 1959). An instrument with the properties documented here would pass all of them.

A conditional dispersion check adds something they do not provide. It asks how much variation in a measure remains among units that are alike on the construct's defining inputs. Where those inputs are already recorded it requires no further data collection, it yields a single interpretable quantity, and it speaks directly to whether a score can be read as a statement about the circumstances it names. That question is central to any measure used to compare groups and direct resources, and it is among the least often asked.

Limitations

Seven limitations qualify the results, and they bear unevenly on the claims made above.

The age asymmetry is open to a measurement explanation. If the instrument functions differently across age groups, part of what is reported here as differential evaluation would instead be a property of the scale. The item-level evidence reported above does not point to any single item carrying the effect, but it does not exclude the possibility either, and a formal invariance analysis would be needed to do so.

Where race and ethnicity are concerned the estimate is not precise. The raw gap is small and the categories in the public use file are coarse, so the bootstrap interval on the share surviving is too wide to

support a claim in either direction. That comparison should be treated as unresolved rather than as evidence, and it would be better addressed with data carrying finer categories and a larger sample.

The objective block, however extensive, is not exhaustive. Neither survey records net worth, pension entitlements or detailed asset holdings. Some of the unexplained dispersion is therefore attributable to financial circumstances that were not measured, particularly for older households, whose position depends heavily on accumulated wealth. That omission is the most likely partial explanation for the age result, and it cannot be resolved with these data. The estimate of unexplained variance should accordingly be read as an upper bound on the non-financial component.

The design is cross-sectional and the analysis is descriptive. No causal claim is made about temperament influencing the score. The association reported is consistent with several causal structures, including circumstances shaping disposition over time, which the evidence on financial strain and cognitive function makes plausible (Mani et al., 2013).

The temperament block is limited to ten measures available in a survey not designed for personality assessment. A dedicated instrument would likely account for more of the residual, so the temperament share reported here is best read as a lower bound.

All measures come from a single self-reported instrument. The design features noted above limit the consequences for the central claim, but the associations involving the temperament block would be better estimated with independently sourced measures.

The reliability bound uses coefficient alpha, which is a lower bound on reliability under some conditions and can be inflated under others. Alternative reliability estimators would move the ceiling modestly, but not by enough to alter the conclusion, since the gap between the ceiling and the observed fit is large.

Conclusion

The CFPB Financial Well-Being Scale orders households by financial position in the way its developers intended, and the median score rises steadily across deciles of financial circumstance in two independent national surveys nine years apart. Within those deciles, households receive scores that differ by twenty points or more. About forty percent of the score’s reliable variance is not accounted for by an extensive record of financial position, and that remainder carries systematic association with temperament and with general life satisfaction. The pattern is not an artefact of functional form, since it survives a saturated specification and gradient boosting, and it is not an artefact of weighting, missing data treatment or subsample.

These results do not warrant abandoning the measure. They warrant reporting any comparison drawn from it alongside the share of the gap that financial position accounts for, together with the precision of that share. That share is close to complete for income and education, roughly sixty percent for age, and not determinable for race and ethnicity from these data. A measure used to direct attention and

resources toward the households that need them should be accompanied by an estimate of how much of what it reports is about circumstance.

Declarations

Conflict of interest: On behalf of all authors, the corresponding author states that there is no conflict of interest. The author has no relevant financial or non-financial interests to disclose.

Consent to participate: Not applicable. The analysis uses secondary de-identified data.

Code availability: All analysis code is available at https://osf.io/fnc6j/?view_only=f12c3787ea6649bc92f87cab9390eac4, an anonymised view-only link provided for peer review. The repository reproduces every number, table and figure reported here, and includes a script that downloads both public use files and verifies each against the checksum of the version analysed. The project will be made publicly available on acceptance.

References

1. Anvari-Clark, J., & Ansong, D. (2022). Predicting Financial Well-Being Using the Financial Capability Perspective: The Roles of Financial Shocks, Income Volatility, Financial Products, and Savings Behaviors. *Journal of Family and Economic Issues*, *43*(4), 730–743. https://doi.org/10.1007/s10834-022-09849-w
2. Beaumont, J. F. (2008). A new approach to weighting and inference in sample surveys. *Biometrika*, *95*(3), 539–553. https://doi.org/10.1093/biomet/asn028
3. Birkenmaier, J., Rothwell, D., Frey, J. J., & Coffey, D. S. (2022). Introduction to the Special Issue on Financial Capability and Asset Building for Family Financial Wellbeing. *Journal of Family and Economic Issues*, *43*(4), 647–653. https://doi.org/10.1007/s10834-022-09871-y
4. Blanchflower, D. G., & Oswald, A. J. (2008). Is well-being U-shaped over the life cycle? *Social Science & Medicine*, *66*(8), 1733–1749. https://doi.org/10.1016/j.socscimed.2008.01.030
5. Board of Governors of the Federal Reserve System (2025). *Economic well-being of U.S. households in 2024*. https://www.federalreserve.gov/consumerscommunities/shed.htm
6. Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295x.111.4.1061
7. Boyce, C. J., Brown, G. D., & Moore, S. C. (2010). Money and Happiness. *Psychological Science*, *21*(4), 471–475. https://doi.org/10.1177/0956797610362671
8. Brüggen, E. C., Hogreve, J., Holmlund, M., Kabadayi, S., & Löfgren, M. (2017). Financial well-being: A conceptualization and research agenda. *Journal of Business Research*, *79*, 228–237. https://doi.org/10.1016/j.jbusres.2017.03.013
9. Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

10. Caraballo, J. (2026). Adding Equity Considerations to Construct, Scale, and Model Development in the Field of Financial Education. *Journal of Family and Economic Issues*. https://doi.org/10.1007/s10834-026-10103-w
11. Chen, P. J., Lee, H., & Lee, J. (2026). Validating Financial Well-Being Scales Across Cultures: Evidence from Eastern Cultural Contexts. *Journal of Family and Economic Issues*. https://doi.org/10.1007/s10834-026-10080-0
12. Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature*, *46*(1), 95–144. https://doi.org/10.1257/jel.46.1.95
13. Comerton-Forde, C., de New, J., Salamanca, N., Ribar, D. C., Nicastro, A., & Ross, J. (2022). Measuring Financial Wellbeing with Self-Reported and Bank Record Data. *Economic Record*, *98*(321), 133–151. https://doi.org/10.1111/1475-4932.12664
14. Consumer Financial Protection Bureau (2015). *Measuring financial well-being: A guide to using the CFPB Financial Well-Being Scale*. https://www.consumerfinance.gov/data-research/research-reports/financial-well-being-scale/
15. Consumer Financial Protection Bureau (2017a). *CFPB Financial Well-Being Scale: Scale development technical report*. https://www.consumerfinance.gov/data-research/research-reports/financial-well-being-technical-report/
16. Consumer Financial Protection Bureau (2017b). *Financial well-being in America*. https://files.consumerfinance.gov/f/documents/201709_cfpb_financial-well-being-in-America.pdf
17. Consumer Financial Protection Bureau (2017c). *National Financial Well-Being Survey public use file user's guide*. https://www.consumerfinance.gov/data-research/financial-well-being-survey-data/
18. Consumer Financial Protection Bureau (2025). *Making Ends Meet Survey public use file user's guide*. https://www.consumerfinance.gov/data-research/making-ends-meet-survey-data/
19. Conversano, C., Rotondo, A., Lensi, E., Della Vista, O., Arpone, F., & Reda, M. A. (2010). Optimism and Its Impact on Mental and Physical Well-Being. *Clinical Practice & Epidemiology in Mental Health*, *06*(1), 25–29. https://doi.org/10.2174/1745017901006010025
20. Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/bf02310555
21. Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957
22. Deaton, A. (2008). Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll. *Journal of Economic Perspectives*, *22*(2), 53–72. https://doi.org/10.1257/jep.22.2.53
23. Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13
24. Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276–302. https://doi.org/10.1037/0033-2909.125.2.276

25. Diener, E., Lucas, R. E., & Scollon, C. N. (2009). Beyond the Hedonic Treadmill: Revising the Adaptation Theory of Well-Being. *Social Indicators Research Series*, 103–118. https://doi.org/10.1007/978-90-481-2350-6_5
26. Fan, L., & Henager, R. (2021). A Structural Determinants Framework for Financial Well-Being. *Journal of Family and Economic Issues*, *43*(2), 415–428. https://doi.org/10.1007/s10834-021-09798-w
27. Fernandes, D., Lynch, J. G., & Netemeyer, R. G. (2014). Financial Literacy, Financial Education, and Downstream Financial Behaviors. *Management Science*, *60*(8), 1861–1883. https://doi.org/10.1287/mnsc.2013.1849
28. Ferrer-i-Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics*, *89*(5–6), 997–1019. https://doi.org/10.1016/j.jpubeco.2004.06.003
29. Fryer, D., Strümke, I., & Nguyen, H. (2020). Shapley Value Confidence Intervals for Attributing Variance Explained. *Frontiers in Applied Mathematics and Statistics*, *6*. https://doi.org/10.3389/fams.2020.587199
30. Fujita, F., & Diener, E. (2005). Life Satisfaction Set Point: Stability and Change. *Journal of Personality and Social Psychology*, *88*(1), 158–164. https://doi.org/10.1037/0022-3514.88.1.158
31. Gathergood, J. (2012). Debt and Depression: Causal Links and Social Norm Effects. *The Economic Journal*, *122*(563), 1094–1114. https://doi.org/10.1111/j.1468-0297.2012.02519.x
32. Grömping, U. (2006). Relative Importance for Linear Regression inR: The Packagerelaimpo. *Journal of Statistical Software*, *17*(1). https://doi.org/10.18637/jss.v017.i01
33. Grömping, U. (2007). Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. *The American Statistician*, *61*(2), 139–147. https://doi.org/10.1198/000313007x188252
34. Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, *70*(5), 646–675. https://doi.org/10.1093/poq/nfl033
35. Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, *72*(2), 167–189. https://doi.org/10.1093/poq/nfn011
36. Heck, P., Ratcliffe, C., & Tibbitts, E. (2024). Consumer Financial Well-Being: Does Scale Choice Alter the Measure? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4683351
37. Iceland, J., & Bauman, K. J. (2007). Income poverty and material hardship: How strong is the association? *The Journal of Socio-Economics*, *36*(3), 376–396. https://doi.org/10.1016/j.socec.2006.12.003
38. Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, *107*(38), 16489–16493. https://doi.org/10.1073/pnas.1011492107
39. Kaiser, T., Lusardi, A., Menkhoff, L., & Urban, C. (2022). Financial education affects financial knowledge and downstream behaviors. *Journal of Financial Economics*, *145*(2), 255–272. https://doi.org/10.1016/j.jfineco.2021.09.022

40. Kelley, H. H., Lee, Y., LeBaron-Black, A., Dollahite, D. C., James, S., Marks, L. D., & Hall, T. (2022). Change in Financial Stress and Relational Wellbeing During COVID-19: Exacerbating and Alleviating Influences. *Journal of Family and Economic Issues*, *44*(1), 34–52. https://doi.org/10.1007/s10834-022-09822-7
41. Khashadourian, E., & Harrison, A. L. (2024). Perceptions or behavior? An evaluation of CFPB's financial well-being scale using household financial ratios. *Financial Planning Review*, *7*(3–4). https://doi.org/10.1002/cfp2.1194
42. Killingsworth, M. A. (2021). Experienced well-being rises with income, even above $75,000 per year. *Proceedings of the National Academy of Sciences*, *118*(4). https://doi.org/10.1073/pnas.2016976118
43. Killingsworth, M. A., Kahneman, D., & Mellers, B. (2023). Income and emotional well-being: A conflict resolved. *Proceedings of the National Academy of Sciences*, *120*(10). https://doi.org/10.1073/pnas.2208661120
44. Lusardi, A., Mitchell, O. S., & Curto, V. (2010). Financial Literacy among the Young. *Journal of Consumer Affairs*, *44*(2), 358–380. https://doi.org/10.1111/j.1745-6606.2010.01173.x
45. Lusardi, A., Schneider, D., & Tufano, P. (2011). Financially Fragile Households: Evidence and Implications. *Brookings Papers on Economic Activity*, *2011*(1), 83–134. https://doi.org/10.1353/eca.2011.0002
46. Lusardi, A., & Mitchell, O. S. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature*, *52*(1), 5–44. https://doi.org/10.1257/jel.52.1.5
47. Lusardi, A., Michaud, P. C., & Mitchell, O. S. (2017). Optimal Financial Knowledge and Wealth Inequality. *Journal of Political Economy*, *125*(2), 431–477. https://doi.org/10.1086/690950
48. Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing Happiness: The Architecture of Sustainable Change. *Review of General Psychology*, *9*(2), 111–131. https://doi.org/10.1037/1089-2680.9.2.111
49. Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty Impedes Cognitive Function. *Science*, *341*(6149), 976–980. https://doi.org/10.1126/science.1238041
50. Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/bf02294825
51. Netemeyer, R. G., Warmath, D., Fernandes, D., & Lynch, J. G. (2018). How Am I Doing? Perceived Financial Well-Being, Its Potential Antecedents, and Its Relation to Overall Well-Being. *Journal of Consumer Research*, *45*(1), 68–89. https://doi.org/10.1093/jcr/ucx109
52. Osberg, L., & Sharpe, A. (2014). Measuring Economic Insecurity in Rich and Poor Nations. *Review of Income and Wealth*, *60(S1)*. https://doi.org/10.1111/roiw.12114
53. Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879
54. Prawitz, A. D., Garman, E. T., Sorhaindo, B., O'Neill, B., Kim, J., & Drentea, P. (2006). InCharge Financial Distress/Financial Well-Being Scale. *PsycTESTS Dataset*. https://doi.org/10.1037/t60365-000

55. Reiter, M., Jurgenson, J. B., McCoy, M., Watkins, K., White, K., & Warmath, D. (2025). Evaluating the Klontz Money Script Inventory-Revised (KMSI-R): Factorial Validity, Internal Consistency, and Measurement Invariance with a Diverse Sample. *Journal of Family and Economic Issues*. https://doi.org/10.1007/s10834-025-10055-7
56. Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. *Wiley Series in Probability and Statistics*. https://doi.org/10.1002/9780470316696
57. Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989x.7.2.147
58. Scheier, M. E., & Carver, C. S. (1987). Dispositional Optimism and Physical Well-Being: The Influence of Generalized Outcome Expectancies on Health. *Journal of Personality*, *55*(2), 169–210. https://doi.org/10.1111/j.1467-6494.1987.tb00434.x
59. Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*(2), 93–105. https://doi.org/10.1037/0003-066x.54.2.93
60. Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some Consequences of Having Too Little. *Science*, *338*(6107), 682–685. https://doi.org/10.1126/science.1222426
61. Sorgente, A., Lanz, M., Bhatia, S., Crespo, C., Dunai, J., Fonseca, G., Lep, Ž., Martos, T., Negru–Subtirica, O., Portugal, A., Ranta, M., Relvas, A. P., Sallay, V., Singh, N., Sirsch, U., Vosylis, R., Zupančič, M., Shim, S., & Serido, J. (2025). Financial Identity Scale: Testing the International Validity of Its Variable-Centered and Person-Centered Models. *Journal of Family and Economic Issues*. https://doi.org/10.1007/s10834-025-10054-8
62. Sun, S., Chen, Y. C., Ansong, D., Huang, J., & Sherraden, M. S. (2022). Household Financial Capability and Economic Hardship: An Empirical Examination of the Financial Capability Framework. *Journal of Family and Economic Issues*, *43*(4), 716–729. https://doi.org/10.1007/s10834-022-09816-5
63. Sun, M., & Wang, S. (2026). Parenthood Transition and Financial Well-Being Across Gender and Occupational Class: A Longitudinal Study in the United Kingdom. *Journal of Family and Economic Issues*. https://doi.org/10.1007/s10834-026-10105-8
64. Totenhagen, C. J., Saxey, M. T., Love, H. A., Wilmarth, M. J., & Lucier-Greer, M. (2026). A Psychometric Evaluation of the Perceived Financial Well-Being Scale among Individuals in Diverse Couple Relationships. *Journal of Family and Economic Issues*. https://doi.org/10.1007/s10834-026-10099-3
65. Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002
66. Warmath, D., Borowiec, K., LeBaron-Black, A. B., Mottola, G., Reiter, M., Schindler, K., & Caraballo, J. (2026a). Improving Measurement in the Study of Consumer Finance: Consensus from the NEFE Scale Validation Project. *Journal of Family and Economic Issues*. https://doi.org/10.1007/s10834-026-10096-6

67. Warmath, D., LeBaron-Black, A. B., & Wilmarth, M. (2026b). Introduction to the Special Issue on Measurement. *Journal of Family and Economic Issues*. https://doi.org/10.1007/s10834-026-10102-x

68. Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, *96*(3), 465–490. https://doi.org/10.1037/0033-2909.96.3.465

69. Xiao, J. J., & Kim, K. T. (2021). The Able Worry More? Debt Delinquency, Financial Capability, and Financial Stress. *Journal of Family and Economic Issues*, *43*(1), 138–152. https://doi.org/10.1007/s10834-021-09767-3

Figures

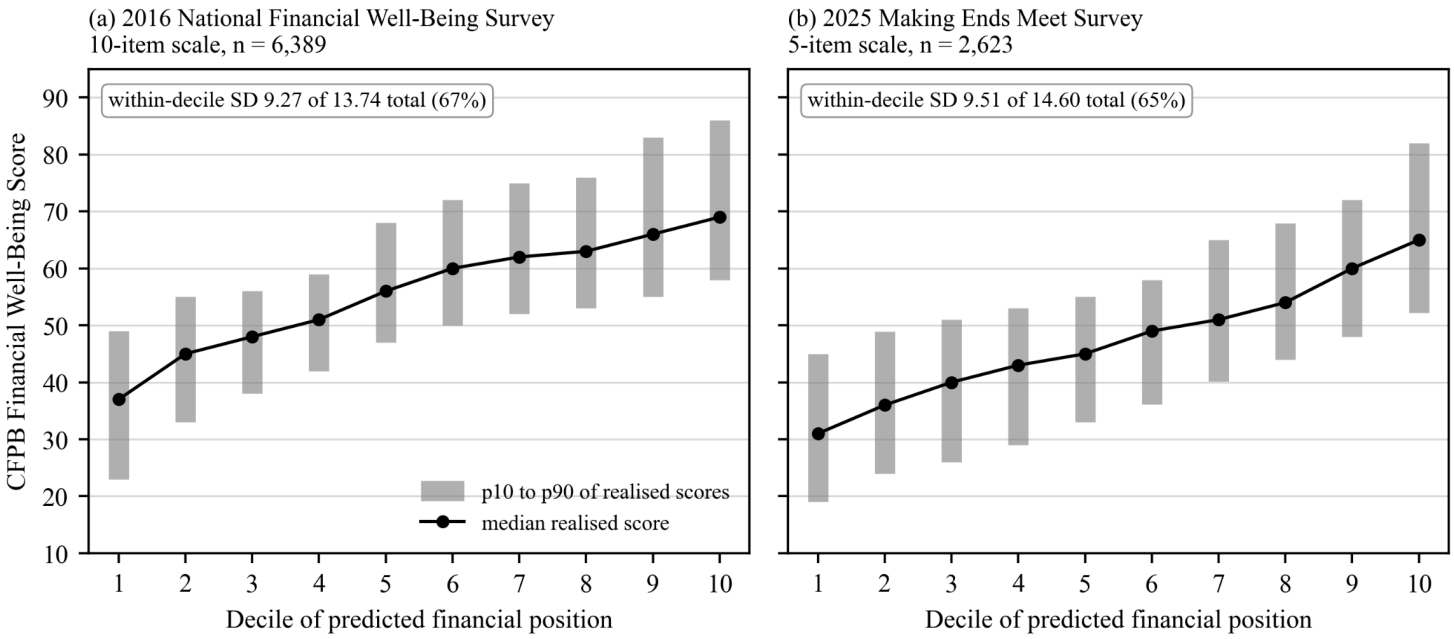


Figure 1

Realised scale scores within deciles of predicted financial position, 2016 and 2025 surveys. Grey bars show the tenth to ninetieth percentile range of realised scores within each decile; the line marks the median.