

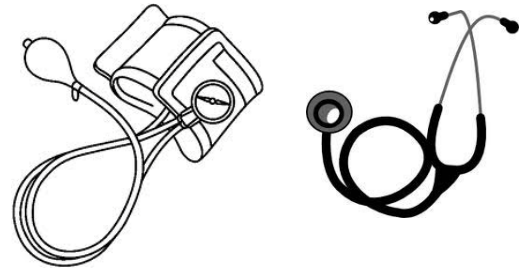
Finding A Yardstick: Field Testing Outcome Measures for Community-based Financial Coaching and Capability Programs¹

August 2013

Introduction

Despite a strong emphasis on building people's capability to manage their household finances, the "field" lacks a clear set of standards for measuring the condition of any individual's financial status. Community-based programs currently use a wide range of measures, each collected and calculated in a slightly different way. One perspective is that focusing on measurement distracts from service delivery and, at best, is a secondary priority. But this view is shortsighted.

Measurement Tools



Imagine, for example, how medical professionals would manage health care without a stethoscope and a blood pressure cuff. Imagine further if we lacked standards for pulse rates by gender and age or maximum (systolic) and minimum (diastolic) blood pressure. These metrics are widely used in medicine today but took many years to evolve. In fact, although blood pressure was understood in the 1700s, it could not be easily and accurately measured until the early 1900s.² Standards for appropriate levels for these measures continue to evolve to this day, but having reliable and valid ways of measuring blood pressure and pulse are fundamental to the health care field. This report is intended to focus on the issue of measurement within the specific application of financial coaching programs broadly defined.³

In 2011, the Center for Financial Security launched the *Financial Coaching Outcome Measures Project* to test a set of standardized measures for the financial coaching field. Four nonprofit community-based organizations collected data on client outcomes and shared the results to form a consistent database of similar measures. The goal of this report is to describe a small set of standardized measures that could be worthy of consideration as "benchmarks" for the field. The goal of this project is to develop measures that satisfy standards of social science while also being attuned to the practical issues of data collection and analysis. To that end, this document serves as a starting place for a discussion about outcomes measurement standards for the financial coaching field. Ideally, the recommendations from this project can be applied to financial capability interventions broadly, increasing our collective understanding of how, and why, programs work.

Why Standardize?

Community-based programs have developed existing data points based on their unique interests, goals, and demands from funders. This project seeks to propose a small set of measures that organizations can readily integrate into or alongside their existing client tracking systems. By working together in this way, the financial coaching field can significantly improve its capacity to demonstrate client outcomes to funders, policymakers, and other stakeholders. The current lack of uniformity in measures is a barrier to

¹Authors: J. Michael Collins (jmcollins@wisc.edu) and Collin O'Rourke. This project was generously supported by the Annie E. Casey Foundation. The authors thank Bon Secours, the Financial Clinic, LISC Chicago, and the University of Wisconsin-Extension for their efforts in collecting and sharing data on client outcomes.

² <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1543468>

³ There remain a spectrum of financial coaching models and programs, as described in <http://fyi.uwex.edu/financialcoaching/>

more rigorous program assessment. Consistent outcome measures are a prerequisite to larger-scale and more coordinated assessment efforts. Benefits of standardizing measures include:

- **Achieving Scale:** It is often difficult for individual coaching programs to amass the amount of data needed for rigorous program assessment. By aggregating data from multiple programs, the financial coaching field can work together to demonstrate its effectiveness.
- **Studying Models:** Because the coaching field encompasses a wide variety of approaches, having a common set of measures facilitates studies where outcomes across programs can be assessed.
- **Assuring Quality:** Standardized measures will ideally be statistically reliable and valid. Collecting individual-level financial data is difficult to do in a manner that is consistent yet also captures changes over time. No one organization will need to invest in developing measures.
- **Efficiency:** The adoption of standardized measures could ease some of the burden organizations face in collecting client data. Organizations may collect dozens of measures on an individual client, and agreement on a standardized set of measures could allow organizations to reduce the number of data points they track.

The development of low-cost information technology and customized databases has resulted in a significant increase in nonprofit organizations' capacity to collect customized data points. Programs can—and many do—collect hundreds of data points on a single client during the intake process and over the course of their interactions with clients. To the extent funders and other leaders in the field can agree to focus on a small number of indicators that are collected in a standardized way, these measures could provide more credible evidence that financial coaching and related activities are successful.⁴

Section I: Background on Metrics

Purposes

Unlike other fields such as public health, nutrition, child maltreatment, or food security, the financial capability field does not have a widely recognized set of measures. For decades, surveys have measured income, assets, debt, and aspects of spending and savings behaviors, as well as negative events like job losses and bankruptcy. Nonetheless, financial knowledge and basic financial management have only been measured sporadically for about the past decade. These measures have typically been included in general social or economic surveys, and have not been part of data collected from programs in the field. The current need is for a set of measures that can contribute to the following:

1. **Diagnosis:** for the purpose of triaging services and matching clients to appropriate resources
2. **Baseline status:** in order to document the initial or pre-service condition of clients
3. **Outcome measurement:** in order to re-assess clients' status at baseline or prior periods relative to their intermediate or post-service status.

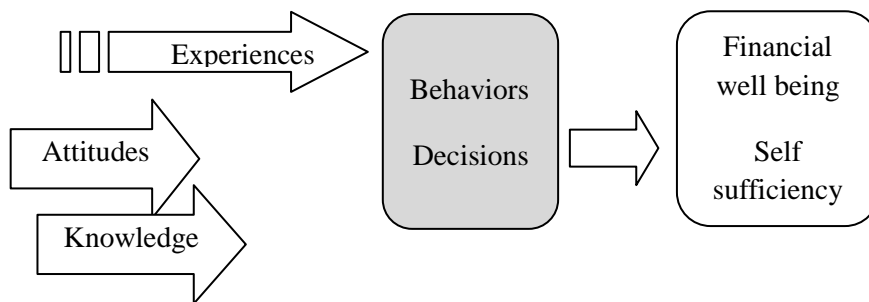
Ideally, the same set of measures can serve all three functions, although data used for diagnosis might include more administrative information related to matriculation of clients in a program, as well as extensive demographic and other static measures that will not change over time (e.g., race, age, gender). Outcome measures will be repeated at least once, and perhaps multiple times. Because not all clients will cooperate with follow-up efforts, the data points used for measuring outcomes over time need to be simple and efficient to administer.

⁴ Clearly, there are ramifications to highlighting any particular indicators over others. Each organization has invested in its own systems and will be resistant to changing them, especially when the systems were developed as part of funder-driven mandates. Thus, it is important to again underscore that a focus on a key set of indicators is not a substitute for existing measures.

Behavior, Knowledge, Attitudes, Well-being, and Other Constructs

Often the focus of financial capability programs is to influence behaviors, especially those related to paying bills in a timely way, maintaining a budget, and building targeted savings for investments in human capital and other activities that may lead to economic self-sufficiency. Some programs also view building the capability of clients to manage household finances as a goal, which may be measured through knowledge scales—either objective quiz style questions, or subjective assessments of knowledge levels. Less often measured are issues such as confidence and self-efficacy, which may drive behaviors and influence how people treat financial decisions. Finally, there are issues of stress, family relations, health, and other aspects of life in families and households that are broader than financial management but may be directly influenced by financial conditions and decision-making.

Ultimately, behaviors (e.g., paying bills on time, borrowing, saving) are observable and offer a reasonably clear link to financial capability, stability, and self-sufficiency. Knowledge and attitudes also likely contribute to behavior. Of course, human behavior is complicated and is subject to individual biases, mistakes, procrastination, and external social pressures. Prior experiences also influence future behaviors. Even clients with extensive financial capability supports may fail to change behaviors over time. Thus, attitudes and knowledge may be measured as potential influences or precursors to behavioral changes. The figure below represents these mechanisms. Currently, measures of well-being or self-sufficiency are not well-defined, thus placing the primary focus on behavioral measures.



What is the Ultimate Outcome of Interest?

In 2011, we conducted a focus group at the National Community Tax Coalition national meeting with about 20 leaders in the financial capability field. That session and follow-up telephone interviews generated the following—very global—list of potential outcomes measures for financial coaching:⁵

- **Savings:**
 - Dollars in savings, net savings, targeted savings (e.g., education or retirement), or rainy day fund.
- **Banked:**
 - Having a transaction account or reloadable debit card (being “banked”)
 - Use of alternative credit financial products such as payday or pawnshop loans
- **Debt:**
 - Dollars of total debt, credit card debt, or other consumer or revolving debt (not mortgages or car loans)

⁵ Not on this list are issues related to food security, housing stability and quality, school access and related services. Although clearly important, the full scope of measures became far too broad.

- **Credit:**
 - Credit score, credit report number of delinquent payments, credit report number of judgments (more severe) or accounts in collections (less severe)
- **Psychology:**
 - Self-reported confidence in one’s ability to manage own finances
 - Having a focused financial goal, specificity of goal, length of goal
- **Economic Self Sufficiency:**
 - Income amount, changes in wage rate, changes in hours worked, being employed
 - Accessing benefits such as SNAP and Medicaid; accessing subsidies for child care, housing and transportation; obtaining child support
- **Access to Services:**
 - Referrals to legal services for landlord issues, domestic violence or credit problems
 - Referrals to mental health and human services for children and family supportive services

Most practitioners appear to have two reactions to this list: (1) the list is too long, and (2) the list is still missing an important measure that is critical to their particular program or model. This obviously presents a fundamental conflict. The table below summarizes the logic of each set of measurement “constructs” in order to help narrow the field.

Construct	Critical Analysis
Savings (household level)	There are many forms and types of savings. For very-low-income families, savings can mean not consuming critical items including housing and food. Savings for schooling or retirement may not be rational given existing subsidy programs. Emergency funds are by design highly variable (they are spent down as needed) and may take non-traditional forms. Net savings becomes very complicated to calculate due to the wide variety of assets classes and forms of debt. The asset and debt modules alone on the Federal Reserve’s official surveys of wealth take 30 minutes or longer to complete. For many low-income and younger households, “investing” in schooling or other ways of building future income means a negative net worth is a potentially positive outcome. Overall, there is a high degree of ambiguity when interpreting data on assets and debt.
Banking (household level)	Implicit in this construct is that being banked is always positive. Individuals with low balances or highly variable cash flows may find checking accounts too costly or inconvenient. The proliferation of debit and pre-paid means “banking services” is a moving target. Using alternative sources of credit to tap liquidity may be the lowest cost—or last resort—option. Generally, banking status is most correlated with having a reliable income source and a stable household situation. These measures typically rely on self-reports.
Debt (household level)	Like savings, debt comes in many forms. It is difficult to distinguish good and bad debt, since most debt can be substituted across categories (e.g., using an auto title loan to pay off a credit card, or vice versa). Increasing debt levels are often associated with increasing savings, income, and credit scores. These measures generally rely on self-reports, although documents and credit reports can help verify.
Credit (individual level)	One unambiguously negative measure is delinquencies, especially recent delinquencies. Credit scores capture current delinquencies, as well as past behaviors (including long past behaviors); scores are not available for all individuals. Public records remain associated with clients long after their financial behaviors begin to improve. Unlike other measures, credit items are by individual client, not the overall household. These measures are essentially owned by

proprietary firms. Payment behavior seems to be a good proxy for regular financial management behavior and an indicator of financial capability.

Psychology (household level) These measures target higher level outcomes related to well-being, but are not well defined. Because these are self-reported, question wording and response rates are highly variable. No national survey benchmarks or standards exist. These measures require client cooperation but are not dependent on external data sources. These measures appear most related to a goal-based coaching approach.

Economic Sufficiency (household level) These measures are more well-defined in the workforce development field and among poverty researchers. Typically, this information is dependent on external data sources gathered in cooperation with public agencies. These measures are more related to case management than a goal-based coaching approach.

Human Services (household level) These measures are hard to collect and interpret. Referrals depend on the nature of clients' circumstances, the quality of the caseworker, and services available in local communities.

Financial coaching, like some other financial capability approaches, is focused on improving clients' ability to manage their own finances over time. Coaching clients often need support for basic financial management tasks, especially paying attention to regular monthly obligations, dealing with unexpected drops in income or expenses, and engaging in longer term planning and behaviors that are easy to put off. As such, the key constructs identified include **credit** and **psychology**.

In this report, we have a mix of credit report items, credit scores and self-reported indicators of confidence and goal-driven behavior. We have not focused on savings or net worth, due to both the difficulty of collecting and nebulous value of this information. In the marketplace, especially for low-income households, credit scores are a widely used metric. Scores of various kinds are used to evaluate loan applications, but also for employment, apartment leases, utilities access, insurance, and a range of other goods and services. As a result, credit score is a key benchmark for the other measures in this report. Unfortunately, credit scores are owned by the agencies that calculate them (esp. FICO, TransUnion, Experian, and the factors that determine scores are not easily determined with any precision.⁶ Some portion of clients, especially younger and lower-income clients, will not have a credit score, even if they have a credit report. We therefore resort to a range of alternative benchmarks to triangulate estimates for various measures.

Methods of Collection: Self Report vs. Administrative Data

Financial capability programs largely use self-reported intake (and later follow up) survey data, administrative data collected during service delivery, data from financial statements or income tax forms, and, more recently, credit report data. There are hundreds of variations in terms of self-report questions and administrative data points. Indeed, there seems to be an explosion of variables measured by community based organizations.

⁶ We observe—and in fact are told by FICO—that timely payments make up about 1/3rd of the score. The amount of credit limits used, total number of accounts, and age of accounts are all important factors. Nonetheless, exactly how scores are calculated remains vague. We cannot, for example, say that one 30 day late payment causes a 20 point drop in score for any given client.

Every form of data has benefits, but each also has limitations, as summarized in the following table:

	Pros	Cons
Survey	Flexible, especially when collected at intake or as part of services	Self-report bias; lack of cooperation, especially for follow-up; can be costly to collect, requiring incentives.
Credit score	Single metric; widely used in industry to allocate and price credit; well understood benchmarks	Proprietary formula based on credit report for a fee. Some clients have 'thin' files and no score. Slow to change even as behaviors improve.
Credit report items	Number of accounts, delinquencies and balance ratios available from credit report; more likely to move than score	Some clients lack reports; benchmarks not well agreed upon; privacy and consent barriers.
Bank balances	Directly manage assets; more accurate than self-report(?)	Very "noisy." Only a snapshot of one point in time; regulatory, privacy and client consent barriers.
Budgeting data	Self-reported in concert with transaction records and relevant documents for verification	Complicated and time consuming; one point in time. Challenging to collect longitudinally.
Public records	Legal filings generally posted within 1-2 months of judgment, foreclosure, or bankruptcy	Relatively low probability events. Data quality and access is uneven across communities.
Public agency data	Public benefits use, earned income reported under FICA, education, unemployment, and other issues are related to household financial status.	Requires agreements with state agencies; consent and human subjects limits. Generally need large-scale projects.

What do we mean by statistical reliability and validity?⁷

In testing the pilot measures, we used the available data, including data from credit reports, to assess different aspects of the measures' validity and reliability. Statistical validity and reliability are distinct concepts used to assess the quality of different measures used in research.

Validity: Does each item measure what we intend to measure? How true is the measure relative to "reality" or "truth?" (The problem is we often lack a benchmark for truth)

- Example: Are respondents' self-reported answers to the question, "How would you rate your current credit record?" accurate relative to the data in their actual credit record?

Reliability: Does a measure provide the same results each time it is used? Does the underlying item we are measuring change within a single client? A measure can be reliable but not valid if it provides consistent data that also happens to be inaccurate.

- Example: Are respondents' self reported estimates of savings account balances always 10% higher than actual balances? This might be a reliable measure. While not valid, the measure is consistent and can be used to directionally capture savings levels.

⁷ <http://www.socialresearchmethods.net/kb/> is an excellent source for further information about validity and reliability.

The overall goal of developing reliable and valid measures is to produce the highest quality data possible, so that any conclusions drawn from the data are sound. Research on the validity and reliability of measures of financial behavior are scarce, and the lack of rigorously tested measures raises questions about the conclusions drawn from these measures.

Why Not Collect More Data and Test More Outcomes?

There are of course practical reasons why we want the field to measure and collect only the best and most useful information; given time and resources constraints, we want to maximize the value of these efforts. There is also an important statistical reason for minimizing the number of measures collected. Generally, we must pick a level of significance for statistical tests (e.g. a 90, 95%, or 99% confidence level). Using a 95% confidence level, from any random draw of data, we would predict that about 1 in 20 times we would find a statistically significant result. As we add more and more outcomes, there becomes an increasing chance that any one outcome is statistically significant. This leads to “false positives,” or Type I errors. The best approach is to determine what outcomes matter from the outset, and explicitly test them. Collecting more measures and hoping that a few are significant is unacceptable.

One approach is also to take a set of measures and collapse them into a single score (for example, an APGAR test for newborns). In this report we will focus on taking measures and collapsing them into single numerical indices, but given the finite number of measures, these can also be used individually.

What Core Questions Have Been Tested?

The first step in this process involved developing a small set of standardized measures to pilot test. CFS worked with project partners at Bon Secours of Baltimore, MD; the Financial Clinic of New York City; UW-Cooperative Extension, and LISC Chicago to develop a short list of measures that capture outcomes of interest to the financial coaching field and are practical to collect. The process was iterative, with CFS refining the measures based on a series of feedback. CFS drew on input from these partners, existing surveys, and the Center’s own research on outcome measures to develop the following set of questions:

1. Over the last 3 months, have you followed a personal budget, spending plan, or financial plan?
* Yes * No * **Don’t know**

2. Do you currently have at least one financial goal?
* Yes (go to 2a) * No (go to 3)

 ➔ What is your main financial goal? [write-in response] _____

3. How confident are you in your ability to achieve a financial goal you set for yourself today?
* Not at all confident * Less than confident * Somewhat confident
* Relatively confident * Very confident (5 POINT SCALE)

4. In the last 3 months, did you use an automatic deposit or transfer to put money away for a future use such as saving for retirement or education?
* Yes * No * **Don’t know**

5. Have you set aside emergency or rainy day funds that would cover your expenses for 3 months, in case of sickness, job loss, economic downturn, or other emergencies?
* Yes * No * **Don’t know**

6. Over the past 3 months, would you say your household's spending was less than, more than, or about equal to your income? Please do not include large purchases such as a house or car.
* Less than * More than * Equal to * **Don't know** (3 POINTS)
7. In the last 3 months, have you paid a late fee on a loan or bill?
* Yes * No * **Don't know**
8. How would you rate your current credit record?
* Very bad * Bad * About average * Good * Very good * **Don't know** (4 POINTS)

These questions were selected based on feasibility, prior usage, and how well the measures target the key outcomes presumed in financial coaching. The rationale and prior usage for each of these questions is summarized the following table.

Question	Justification	Use in Other Studies
1	Having and following a plan suggests self-control and/or the intention to exercise control.	Health and Retirement Study (Retirement Planning Module)
2	Having a financial goal (ideally, a specific goal and plan of action that one can clearly articulate) facilitates behavior change.	CFS use in field studies
3	Confidence and self-actualization are needed to change behavior.	CFS use in field studies
4	Creating automatic mechanisms allows people to avoid exercising self-control; self-imposed constraints indicate behavioral goals	CFS use in field studies
5	Contingency funds provide liquidity and signal planning and foresight.	FINRA Financial Capability Survey
6	Spending more than income is a potential trigger of instability and growing debt levels.	FINRA Financial Capability Survey
7	Late fees signal money management problems and will hurt credit standing.	
8	Self-reported assessments of credit are proxies for actual credit history.	Perry (2008) Journal of Consumer Affairs

A Note about Question Design

The list above has some unique aspects from a survey design perspective.

Don't Know. Six of the 8 questions include "Don't know." This was a deliberate and perhaps debatable choice. In many cases (Q1, Q4, Q5 and Q7), a "don't know" response likely means "no" and is coded as such. In other cases (Q6 and Q8), "don't know" indicates a lack of financial awareness that might be a negative signal. In general, it is better to force clients to answer the question.

Open Text Response. Question 2 asked clients to write a financial goal. This was costly in terms of survey time and space, but also a deliberate choice. The rationale was that writing a goal would serve to verify whether the client really had a specific financial goal. On the other hand, coding and analyzing textual data is difficult and perhaps impractical going forward. We focus on the clients who wrote any goal, and the length in number of characters.

Item Scales. Questions 3, 6 and 8 use scales, ranging from a 3 to 5 point for each item. This was based on prior studies and may have added needless variation and complexity to the survey. Generally we would

want the minimum number of items possible. We also want to avoid a “middle” choice as found in odd numbered scales as clients will tend to use that option. An even number forces a response.

Section II: Field Testing the Measures

In order to test the standard 8-question survey in the field, we worked with four organizations that provided data, as summarized in the following table. The Bon Secours, Financial Clinic, and UW-Extension data were generally combined into a cross-sectional dataset to create composite scales and estimate reliability. The UW-Extension data included a small number of clients with pre and post service measures, allowing for a panel dataset that could be used to estimate changes in items before and after services. Bon Secours and the Financial Clinic provided credit score and report items, allowing us to test the validity of self-reported measures relative to credit items. The LISC Chicago data were entirely different, with more clients, a subset of whom form a repeated panel. The LISC dataset include credit data and a small number (4) of questions comparable to questions used in the Bon Secours, the Financial Clinic and UW-Extension data. We present these data in various combinations in the discussion that follows and the attached appendix figures.

Data	Time	Measures	Use
<i>Bon Secours (Baltimore)</i>	fall 2012 to spring 2013	8Q Survey; Credit records (debt, delinquencies, judgments, credit scores)	X-Sectional Reliability (combined); Validity
<i>The Financial Clinic (New York City)</i>	fall 2012 to spring 2013	8Q Survey; Credit records (debt, delinquencies, judgments, credit scores)	X-Sectional Reliability (combined); Validity
<i>University of Wisconsin-Extension (UW-Extension)</i>	fall 2012 to spring 2013	8Q Survey	X-Sectional Reliability (combined); Small <u>panel</u> to examine pre-post outcomes
<i>LISC Chicago (always used alone)</i>	January 2011 to December 2012	Full intake (100+ measures), Budgets, Credit records (debt, delinquencies, judgments, credit scores); Survey Qs (4 in common)	X-Sectional reliability and validity; <u>panel</u> to examine pre-post outcomes; Analysis of Credit Scores

Which Questions Perform “Best”?

The Appendix includes 21 tables and 7 figures. Rather than summarize each table, this report will discuss general themes, referring to the Appendix for background.

Internal Reliability

First, we explore the interrelationships among measures. By testing these interrelationships, we learn how well the eight questions work together to provide an assessment of clients’ financial status. To test the internal consistency of the survey measures, we find Chronbach alpha’s around 0.5-0.6 for the baseline survey responses. This test statistic measures the inter-correlation of a set of variables and is the ratio of a single variable’s variance to variance for the whole set of variables. This test is used in psychology to make sure various question responses measure a single unobserved factor (such as IQ and ability level). A general convention is that a higher alpha is better, with 0.8 being a benchmark level. The 0.5-0.6 levels shown here are weak.

Narrowing the 8 questions to 6 actually increases the alpha slightly; the fact that the alpha does not drop suggests that the 8 measures may include duplicated constructs and is longer than needed. Table 19 provides more detail on the items and each proposed scale. Table 18 presents a factor analysis, another commonly used statistical technique used to diagnose composite scales. Here, we look for Eigen values greater than 1, which is most related to a set of “factors” labeled as #1 in the table (factor 1 – factor here simply means numerical inter-relationships or correlations devoid of any meaning behind the question content). Late fees has a predicted negative relationship and much of the “loading” (think of this as the influence of a question relative to others) is on automatic deposits and emergency savings. Overall, these scales are weak as an overall composite.

Validity

We test whether the eight questions are predictive in cross section of credit report data points. In this case, the credit report data are treated as “true,” so we are testing the validity of the self-reported measures.

We begin by assessing the 8 self reported measures. Here the objective is to first assess each item individually and then as a composite scale. Using cross-sectional data (meaning one point in time) from Bon Secours and the Financial Clinic on about 100 clients, we can see how the 8 questions relate to credit score, total debt, and delinquencies. In general, a composite score suggests strong relationships to credit items, especially credit scores. A composite scale, either calculated on a 0-1 scale, a 1-6 scale, or a 1-10 scale, is related credit score, as shown in Tables 1-3. Table 4 then breaks out each measure discretely. Here we see that late fees, spending more than income, and confidence in achieving goals each has statistically significant relationship to credit score. These relationships also appear to hold when individually analyzing each dependent variable. Debt levels are “noisy” and because of the small sample show few relationships. 30-day delinquencies show that delinquencies are related to not having an emergency fund (predictably) and also having a goal (not an obvious prediction). This is likely because people who have more trouble paying bills may be more motivated to focus on their financial goals.

The LISC Chicago data contain a different set of questions, but have some similar credit measures, as well as savings and net wealth. Here we use a cross-sectional dataset of around 1,300 clients with credit reports and are able to control for factors such as age, income, education, gender, race, and service type (due to more data and more clients included). Table 5 shows that self reports of having accounts in collections are related to credit scores, but in some uneven ways. Reporting late fees are related to higher scores, as is having a checking account. Reporting late fees are also associated with higher savings levels and net worth. Net worth (column 4) appears to be a noisy, highly variable measure based on the size of the estimated coefficients. Having a written goal is related to lower wealth levels, although this accounts for the length of the written goal, which is positive. Overall, these results highlight the complexity of interpreting savings and wealth as an outcome. Tables 6 and 7 combine measures from the LISC data into scales. The first is a shorter 4-item scale, the second a longer 6-item scale. These questions were selected based on their similarity to those in the 8 question survey:

Composite Scale: Short

1. Budget: Do you have a budget (a written spending plan) for all your monthly expenses?
2. Spend More: Total spending < Total income (calculated from budget)
3. Save: Do you set aside money for savings on a regular basis?
4. Late/NSF: Do you bounce checks frequently (at least once a month for the past three months)?

Composite Scale: Long

1. Budget: Do you have a budget (a written spending plan) for all your monthly expenses?
2. Save: Do you set aside money for savings on a regular basis?

3. Collections: Are collection agencies presently contacting you about unsettled claims?
4. Bank: Do you presently have a checking account with a bank or credit union?
5. Goal: Big Vision—Write down the client’s major goal
6. Goal Length: Length of Goal

The shorter list allows testing of concepts similar to the 8 question survey; the longer list allows us to check for other constructs we could be missing. Directionally the shorter question list appears to work in a similar fashion to the 8 question scale related to credit score. It also appears to be related to savings balance, a metric we lack in the other dataset. This is a re-assuring finding.

Figure 1 shows how the composite scale relates to credit scores, suggesting a positive relationship. Figure 4 shows a similar pattern (although confidence is a component, so this is a bit formulaic).

Tables 8-11 provide means and correlations of the data as background. In general, correlations support the findings of the regressions. Table 12 takes advantage of the approximately 250 LISC clients who appear in the data several times (most are observed twice, with about 1/3rd observed 3 or more times). Simple correlations of the 4-item scale show strong relationships within clients across periods—people seem to be consistent in their answers over time. Of course, this prompts a question about outcomes—do these measures *move* as clients are exposed to coaching services?

Longitudinal Measures and Moving Towards “Outcomes”

UW-Extension administered the baseline survey on paper, and follow-up surveys have been administered three months after participants completed the baseline surveys. The follow-up surveys are administered either on paper through the mail or via email. Through June 15, 2012, 412 individuals completed a baseline survey, with 87 individuals completing a follow-up survey. We use these data to present changes in client outcomes over time. Table 13 shows average responses in each period (baseline or pre-coaching services, and then follow-up, 3 months later). Some clients are still receiving services in this follow-up period.

The refined 6 item scale increased 10% (as a marginal effect over baseline). Use of a budget or spending plan increased by 46%. Using automatic/direct deposit increased 28%. Having emergency savings fund increased by 41%. Reports of spending more than income decreased by 30%. There was no effect on reports of late fees and about a 6% increase in confidence in achieving goals. While we lack a comparison group these movements are re-assuring that (1) these measures will shift over time and (2) that coaching interventions can influence these measures.

These measures are all self-reported. The LISC data offer repeated measures within clients each 6 months, with most clients observed twice. These data allow us to create a longitudinal panel and track changes over time. LISC Chicago provides a variety of services including job readiness, tax preparation, and public benefits screening. Through the Family Financial Tracking system, we can observe the primary services sought at intake. About a third sought financial counseling; we propose that these clients are more likely to receive more intensive financial coaching services; clients who primarily sought other services can serve as a comparison group. In practice this is a vague distinction, but may present an opportunity to broaden this outcome analysis. Table 14 shows that these estimates are significant at an 11% level—close to standard levels—but small in size. The 4-item financial capability composite scale in the LISC data shows a modest increase each additional time period. The sample is small, so the significance level of 10% is reasonable to use. Table 15 includes credit report items and has a larger sample since these could be collected without client cooperation (based on prior consent). Controlling for client factors, form of counseling (phone or in person) and time spent with clients, this estimate suggests

an increase in credit score (although not significant statistically) and an increase in savings, as well as declines in delinquencies (again, non-significant).

Finding a Counterfactual (control group)

The challenge ahead, if indeed a set of valid and reliable measures can be developed, is to find a way to develop an unbiased comparison group. Typically, we would attempt to randomly assign clients to services, but often low take up, consent, and compliance rates result in small (and biased) samples. Other sampling methods include waiting lists, cohort control groups, and matching clients across comparable (but distinct) programs. The more consistently measures are used, the more opportunities for comparisons will be created.

Examining Credit Report Outcomes

The final analysis in this report re-focused on credit reports. We generally believe that credit scores and data in credit reports are important constructs. These metrics both represent important personal behaviors and are valued in the marketplace for decision making. Unfortunately, credit scores can be costly to obtain. They are also slow to move (similar to a school GPA, a negative prior history weighs the score down). Scores are also missing for some clients—in the data used here, about 20% of clients with a credit record do not have a credit score. All of these limitations are important to consider before making credit scores a standard metric for the field.

Table 16 conducts further analysis with the LISC panel data to predict follow up (period 2, or t1) credit scores based on baseline (period 1 or t1) responses to self-reported questions. The goal here is to see how these questions might predict credit scores 6 months in the future. The 5 questions examined have some relationship to future scores, especially problems with payments (although in the opposite direction predicted), being banked and number of items in collections. Overall, the model in Table 16 can only explain about 23% of the variation in future scores. Thus, the *predictive* value is low.

Table 17a focuses instead on cross-sectional data and what factors predict whether a client who has a credit report also has a credit score. Here we see that predicting the presence of a score is challenging. Graphically Figure 7 shows that having 2 credit cards increases the average rate of having a score relative to no cards, but the relationship flattens out after 2 active cards are owned by the client. Higher balances are predictive, but a 1,000 increase in credit balances is only estimated to produce a 2.4-2.5 increase in the probability of having a credit score. Demographic factors such as age, education level, and gender are far stronger predictors.

Table 17a examines clients with scores to try and estimate in cross section each client's contemporaneous credit score. Delinquencies are the dominant item, with some effect of revolving balance limits. Notably, client self-reports of credit quality (bad—good) are quite predictive. Using a 5 point scale, each increment of 1 rating on average is related to a 41 point increase in score. Figure 5 shows that self-assessed credit score is related to actual credit score. The relationship appears consistent across credit score values, with perhaps more variation at the tails (low and high score clients).

Recommendations

Following the assessment of measures presented above, we recommend a refined set of self-reported measures combined with selected items from the credit report. Credit score remains an important measure, but the limitations in terms of coverage, cost, and access present serious challenges.

Self-Reports

The following table summarizes the proposed and modified self-reported items.

Question	Concept	Evaluation
1	Budgeting	Modify. Language related to budget, spending plan and financial planning may be too broad. Key construct is planning ahead.
2	Goal	Perhaps retain, but drop written portion unless useful programmatically. Goals are a key activity of coaching but have little diagnostic value.
3	Confidence	Keep. Appears more predictive of credit and savings behaviors than reporting a goal.
4	Auto Deposit/Save	Keep. Language may need to be refined (currently restricted savings examples).
5	Emergency Fund	Keep. Language is convoluted. Time period of 3 months may be excessive. Not clear how spending for an emergency might be handled.
6	Spending	Modify. Currently 3 possible responses. Key factor appears to be spending less than income. Change to yes-no.
7	Late fee	Keep. Time period could be modified, but fees are strongly predictive.
8	Credit rating	Retain. Appears to have little relationship to other behaviors, but is correlated with current credit score (0.63)

Recommended Financial Capability Scale (0-8 point scale)

1. Over the last 3 months, have you followed a personal budget, spending plan, or financial plan?
(Yes=1, No =0)
2. How confident are you in your ability to achieve a financial goal you set for yourself today?
* Not at all confident (0) * Somewhat confident (1) * Very confident (2)
3. Have you set aside funds that would cover your expenses for 3 months if you or someone in your family lost a job, got sick or had another other emergency?
(Yes=1, No=0)
4. Do you currently have an automatic deposit or electronic transfer set up to put money away for a future use such as saving for retirement or education?
(Yes=1, No =0)
5. Over the past 3 months, would you say your household's spending on living expenses was less than its total income?
(Yes=1, No or DK=0) #
6. In the last 3 months, have you paid a late fee on a loan or bill?
(Yes=1, No =0)

Other Useful Questions:

1. How would you rate your current credit record? #, ##
* Very bad * Bad * About average * Good * Very good * Don't know

"Don't know" is an important indicator on its own and should be separated before using the scale.

The 5 responses can be used as a proxy for actual credit quality.

2. Do you currently have at least one financial goal? Yes – No (no "unsure" or "don't know" options)

Credit Report Items

If credit score can be obtained, even given limitations in coverage, it remains a clear measure with accepted benchmarks for quality (for example, 580, 620, 680, 720 and 750). Even if credit score can be obtained, other items from the credit report appear valuable, including:

- # 30 day delinquencies
- # items in collections
- # items judgments (public records section)
- A. \$ revolving credit limit (mainly credit cards, not HELOCs)
- B. \$ revolving credit outstanding (mainly credit cards, not HELOCs)
- Ratio: B/A

Given the cost and difficulty of pulling credit reports on behalf of clients, one alternative might be to develop online tools to help clients pull and extract key items from the annualcreditreport.com website. This would require ongoing client cooperation but could be a promising option for some programs. Coding a software program to "read" reports saved from the annualcreditreport.com website is not likely to be an extensive project, although one could imagine complex extensions of such a third party tool.

Demographics

Organizations already collect extensive background data on clients. A few key considerations that should be standardized along with any data collection system for coaching include:

- Age of client
- Presence of spouse/partner
- Number of dependents
- Employment status (full, part, unemployed, disability)
- Gender
- Education (HS, some college, 4+ degree)

Discussion

Feedback from practitioners is essential, as any future recommendations must be both practical and useful for financial coaching organizations. These discussion questions are designed to begin that dialogue.

Outcome measures

1. Regardless of the findings presented in the previous sections, how well does the 8-question survey capture outcomes of interest?
2. Setting and accomplishing personal goals is a central coaching strategy. Nonetheless, most clients indicate that they have a goal, and goal length or detail is not predictive. How might

data on goals be used? Are there other ways to measure goals or is confidence in goal achievement sufficient?

3. How does the value of credit report data compare to the costs of collecting such data? Recently, access to full credit reports with scores has become more difficult. Is the score worth obtaining? Would a tool to analyze (scrub) annualcreditreport.com reports be valuable?
4. How useful is it to benchmark outcomes against national datasets (e.g., FINRA's National Financial Capability Study)?
5. What is the appropriate timeframe for pre-post measurement? At what point might we predict effects to fade?

Data collection

1. To what extent do clients and coaches view the data collection process as a burden? What are the least intrusive ways to collect high quality data?
2. What do organizations recommend in terms of how to administer a paper-based survey (e.g., timing, whether clients complete alone or with the coach)?
3. How would organizations handle data entry and reporting for a paper-based survey in the future (e.g., enter into the organization's existing database, treat separately from existing data)?
4. How often can and should organizations collect data on client outcomes? What are the key challenges to tracking clients over time, and how can these challenges be addressed?
5. What data points are most feasible to collect from credit reports? How readily can credit data be pulled at multiple points over time?
6. How many clients can organizations collect data on annually, and what percentage of clients will likely complete surveys over multiple points in time?

Collaboration

1. What support is needed to facilitate regular and comprehensive reporting of results?
2. Will leaders in the field, including funders, accept a narrower and more standardized list of measures?
3. How interested are organizations in sharing their data with researchers so the data become part of a centralized database? What barriers and concerns exist to increased data sharing? What do organizations need and expect in return for sharing data?
4. How can human subjects be protected? Are organizations seeking written consent for use of data for research and evaluation?
5. What other administrative data could be obtained? State earnings records? Benefits access? Public records (legal filings)? Bankruptcy/foreclosures? School performance?

Selected LISC Questions: Client Forms

(Shaded areas used in scales)

Products/Practices (Yes No)

- Do you have a budget (a written spending plan) for all your monthly expenses?
- Over the past three months, have you been able to pay your bills on time?
- Over the past three months, have you had to borrow from friends or family to pay for basic necessities like food or rent?
- Do you presently have a checking account with a bank or credit union?
- Do you bounce checks frequently (at least once a month for the past three months)?
- Have you ever had a checking account?
- Do you presently have a savings account with a bank or credit union?
- Do you set aside money for savings on a regular basis?
- Do you presently have one or more active credit cards?

Over the past three months, have you...

- paid the minimum owed on all your cards
- paid less than the minimum, or nothing
- paid the entire balance on all your cards
- paid more than the minimum, but not the entire balance
- Have you ever had a credit card?
- Are you in a Debt Management Plan or working with a Debt Settlement or Credit Repair company?
- Did you file a tax return in the last tax season?

Health Insurance Status (primary insurance only)

- private insurance through a household member's employer;
- private insurance (not through a household member's employer)
- insured through government program;
- no insurance at all

Red Flags

- Have your wages been garnished in the past year, or are you in danger of having your wages garnished?
- Are you in bankruptcy now (i.e. your debt has not yet been fully discharged), or are you in the process of filing for bankruptcy?
- If your household rents, have you been evicted in the last year, or are you in danger of being evicted?
- If your household owns, has your mortgage lender started foreclosure proceedings against you?
- Have any of your utilities been disconnected in the past year, or are you presently in danger of having your utilities disconnected?
- Has your car(s) been repossessed in the past year, or is it presently in danger of being repossessed?
- Are collection agencies presently contacting you about unsettled claims?

Big Vision

Write down the client's major goal

Little Visions (can be multiple)

Write down the client's minor goals/steps to achieve goal

Budget

- Total spending < Total income (calculated)

UW Extension Survey Questions

Please take a moment to complete this survey about your personal finances. Results from the survey will help improve this program and provide better services in the future. Thank you!

1. Over the last 3 months, have you followed a personal budget, spending plan, or financial plan?

Yes No Don't know

2. Do you currently have at least one financial goal?

Yes
 No → Go to Question 3

2a. What is your main financial goal?

3. How confident are you in your ability to achieve a financial goal you set for yourself today?

Not at all confident Less than confident Somewhat confident Relatively confident Very confident

4. In the last 3 months, did you use an automatic deposit or transfer to put money away for a future use such as saving for retirement or education?

Yes No Don't know

5. Have you set aside emergency or rainy day funds that would cover your expenses for 3 months, in case of sickness, job loss, economic downturn, or other emergencies?

Yes No Don't know

6. Over the past 3 months, would you say your household's spending was less than, more than, or about equal to your income? Please do not include large purchases such as a house or car.

Less than More than Equal to Don't know

7. In the last 3 months, have you paid a late fee on a loan or bill?

Yes No Don't know

8. How would you rate your current credit record?

Very bad Bad About average Good Very good Don't know