Chapter 3
Methodology

Abstract We took a set of the psychometric instruments reviewed in Chap. 2 and gave them to a sample of 1,580 small business owners. The majority of businesses had 0–5 employees, had been in operation for more than 3 years, and had $10,000 USD or less in monthly sales. The sample had nearly an equal mix of males and females, typically between the ages of 25 and 54, and over three quarters had at least completed high-school studies. These entrepreneurs were selected because they had loans at one of six participating banks and microfinance institutions across Peru, Kenya, Colombia, and South Africa. Most loans were between $800 and $3,000, though the banks in Kenya & South Africa included clients with loans from $20,000 up to $100,000. Each client was given the series of assessments by representatives of the financial institution. The clients already had their loans for at least six months, and therefore the clients knew that their answers would not directly impact their loan (i.e. a low-stakes setting). This is useful for research, though not ideal for implementation as psychometric tools for credit scoring would be used in practice in a high-stakes setting, where applicants will try to manipulate their responses to get approval of their loan application. In evaluating the contribution of psychometric instruments to better identify high-potential entrepreneurs and direct finance to their ventures, there are two principal outcomes of interest: Business performance (best represented with company profits); and loan repayment (i.e., did the person default). We collected loan repayment history from the financial institution, and profit levels as reported by the entrepreneurs, to compare to responses on the psychometric assessments.

Empirical Strategy: Design Overview

The results reviewed in the previous chapter suggest that there are some dimensions measurable with psychometric instruments that have, at least in some cases, important relationships with entrepreneurial outcomes. However, as noted in many

B. Klinger et al., *Enterprising Psychometrics and Poverty Reduction*,
SpringerBriefs in Psychology: SpringerBriefs in Innovations in Poverty Reduction,
DOI 10.1007/978-1-4614-7227-8_3,

reviews of this body of literature, there are shortcomings with the studies. Many times, the entrepreneurial outcomes that are available for study are not ideal. For example, the studies may only know if the individual is self-employed or not but do not have data on how well the individual's business is actually doing. In this study, we benefit from well-defined measures of business and loan performance.

Another key challenge is that the samples typically available for such studies are drawn for convenience's sake. See, for example, the large number of samples drawing on classes of graduate students in developed countries, to whom researchers have easy access. As noted in a recent special issue of the *Journal of Economic Psychology* focused on the entrepreneurial personality research, "the papers in the special issue also make clear that to answer these questions, more data are needed," and moreover, "once we have such data- whether the results we find for industrialized countries can be extrapolated to transition economies and developing countries" (Caliendo and Kritikos 2012). The need for more data and data drawn from emerging markets is a key contribution of this work.

Most importantly for the question as to whether or not such instruments could increase access to finance, none of these past studies have directly examined loan repayment. In the present study, we will have the benefit of clearly defined outcome variables including loan repayment itself, as well as a highly relevant sample to the question of increasing access to finance: samples of existing small business borrowers from multiple emerging markets.

How can we figure out if we can use psychometric tests to distinguish high-risk from low-risk entrepreneurs in a meaningful way?

If we were interested in precisely estimating the causal relationship between psychometric characteristics and entrepreneurial outcomes, we would have to actually change these characteristics among a randomly selected treatment group and compare them to a control group. In other words, exogenously change intelligence or personality and observe the impact on entrepreneurial outcomes. This is not possible, as the types of traits measured by these tests like personality and intelligence are the outcome of a long process of influences, even genetic factors. They typically do not change much once an individual reaches adulthood (Costa and McCrae 1994), and even if they do, interventions to alter personality are neither well established nor feasible in most research setups.

So randomly varying the traits is not possible. But, in the case of examining psychology's potential contribution to unlocking access to finance, we are not really interested in isolating causal relationships. Many inputs into traditional credit scoring models like an address are not chosen because they necessarily have a causal impact on risk but simply because they have a correlation that can provide predictive power. Similarly, much of the research in industrial and organizational psychology for personnel selection is concerned exclusively with predictive validity, not causality (Almlund et al. 2011). Analogously, we are interested in evaluating tools for screening applicants for finance and identifying high-potential entrepreneurs. Other methods of doing this use proxies, such as the number of dependents, and to test whether these can be replaced with psychometric tools, we must simply examine the power and stability of the relationship between the dimensions we can measure with psychometric instruments and the outcomes we want to predict.

At a first approximation, the psychometric dimensions we seek to measure are stable over time among adults. This allows us to measure them and compare to historical and concurrent outcomes. Therefore, the approach taken herein is to apply psychometric measurements and compare them to current business characteristics and recent loan repayment performance.

This approach does face a number of challenges. Responses to the self-reported psychometric questions used here could be biased, with the respondents attempting to give more socially desirable answers. In this case, clients were explicitly told that their responses would have no effect on their relationship with the financial institution, reducing their desire to give socially desirable responses. Moreover, Hough et al. (1990) and Hogan (1991) show that even in high-stakes settings, respondents rarely manipulate their answers to these types of psychometric assessments unless explicitly instructed to do so.

In actuality, the bigger problem may in fact be the exact opposite: Clients may not manipulate their answers enough. The ultimate goal is to evaluate the potential for these types of tests to be used as screening devices to allocate finance and assistance to entrepreneurs. This means that when entrepreneurs complete the psychometric assessments, they would be under high incentives to give socially desirable answers and "game" the test. To determine if these types of psychometric questions can be implemented in such a high-stakes setting, it would be desirable to replicate that high-stakes situation as closely as possible. In the case of intelligence and skills, there is less of a concern of faking for social desirability because it is not possible to fake on such questions. However, there is an impact of effort on such tests because complex thinking is not automatic and requires effort (Schmeichel et al. 2003), and this "low-stakes" situation may reduce effort and therefore affect results of those questions.

A cleaner method than testing entrepreneurs in a low-stakes setting and looking at their history would be to test the entrepreneurs in a "high-stakes" setting to mimic the incentives in place if the tool were implemented and then to follow them subsequently to address the potential for reverse causality. Such high-stakes up-front testing will be pursued in future studies but has the drawback that a great deal of time must pass between testing, providing financing and then having loans mature and business performance unfold. Testing applicants and looking at current business performance and loan repayment history, while not perfect, has the advantage of providing information more rapidly. Moreover, the comparative results across the dimensions investigated here still contain valuable information because the incentives and timeframe are consistent across the entire sample. For example, all participants have the same motivation in their performance of the Ravens Progressive Matrices and digit span recall tasks, meeting what is termed "standardization for effort" (Almlund et al. 2011).

Testing concurrently to measuring outcomes also limits the types of psychological and cognitive dimensions that can be considered, in favor of the most stable. But for future work, we could extend the focus beyond the more stable dimensions like intelligence and personality and also examine more variable/malleable psychometric dimensions. These dimensions could be tested, and even combined with interventions to improve them in a targeted way. Moreover, this testing and targeted training around malleable dimensions would allow for closer evaluation of causal relationships between the dimensions and outcomes of interest, because randomly selected participants could have the dimensions altered with the training. See Glaub et al. (forthcoming) for a study using this methodology: a randomized control trial of personal initiative training program on entrepreneurs in Africa.

High-stakes ex ante testing and inclusion of stable and malleable traits will be pursued in future research. However as a first examination, testing in a low-stakes situation and comparing responses to historical data is still quite revealing, particularly since the outcome variable and sample available for the present study is a major improvement over what is available in much of the literature to date.

Sample

To evaluate the potential contribution of psychometric tools to entrepreneurial evaluation, we partnered with six financial institutions in four countries:

– Bank #1: A small microfinance organization in semi-urban and rural Peru with average loans of $2,000 to $3,000
– Bank #2: A large commercial bank in Kenya providing with average loan size of $2,000
– Bank #3: A large microfinance organization in Colombia with an average loan size of $800
– Bank #4: A very large commercial bank in South Africa providing loans of $20,000 to $100,000
– Bank #5: A medium-sized commercial bank in Kenya providing loans from $5,000 to 25,000
– Bank #6: A large microfinance organization in Lima, Peru, with an average loan size of $1,000

Following are some summary statistics of the sample.

The majority of entrepreneurs tested were between the ages of 25 and 45 (Fig. 3.1, Table 3.1). The overall sample has a larger number of female than male entrepreneurs, though it can be clearly seen that this is due to the large percentage of female clients of the smaller, Latin American micro-lending institutions (Fig. 3.2, Table 3.2). Microfinance institutions traditionally target female borrowers as they are viewed as lower credit risks and are frequently engaged in small-scale business activities lacking in access to credit, but as finance sizes grow, the gender breakdown of clients begins to skew more heavily towards male borrowers. This concentration of female borrowers only in lower sizes of finance is increasingly being called the female “microfinance ghetto.” In terms of education level, the majority of respondents have at least secondary education, with 40 % having post-secondary training of some type (Fig. 3.3, Table 3.3).

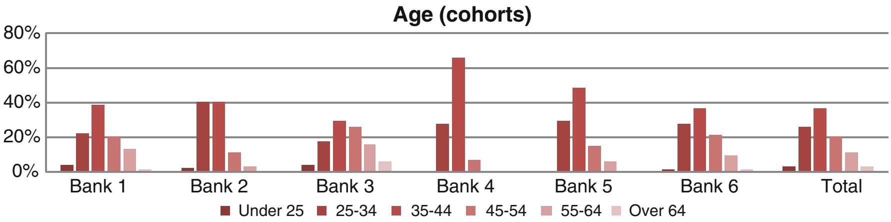

Fig. 3.1 Age distribution by bank in the sample

Table 3.1 Age distribution by bank in the sample

	Partner						
Age (cohorts)	Bank 1 (%)	Bank 2 (%)	Bank 3 (%)	Bank 4 (%)	Bank 5 (%)	Bank 6 (%)	Total (%)
Under 25	4	3	4	0	0	2	3
25–34	23	40	18	28	30	27	26
35–44	38	41	30	66	49	37	37
45–54	20	11	26	7	15	22	20
55–64	13	4	16	0	6	10	11
Over 64	1	1	6	0	0	2	3
Total	100	100	100	100	100	100	100

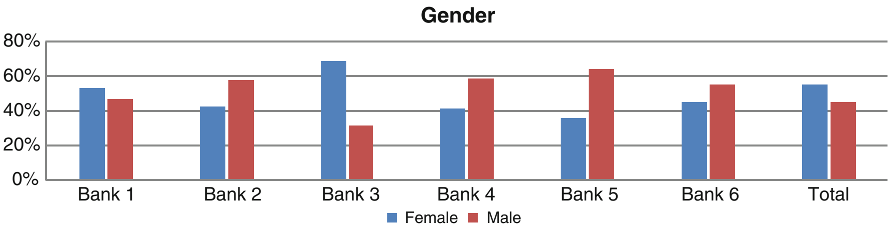

Fig. 3.2 Gender composition by site

Table 3.2 Gender composition by site

	Partner						
Gender	Bank 1 (%)	Bank 2 (%)	Bank 3 (%)	Bank 4 (%)	Bank 5 (%)	Bank 6 (%)	Total (%)
Female	53	43	68	41	36	45	55
Male	47	57	32	59	64	55	45
Total	100	100	100	100	100	100	100

In terms of the businesses themselves, business size unsurprisingly relates closely to the size of financing offered by the financial institution (Fig. 3.4, Table 3.4). Over half the sample consists of businesses earning \$1,000 or less per month in sales revenues and 90 % earning less than \$120,000 per annum with 5 or fewer employees

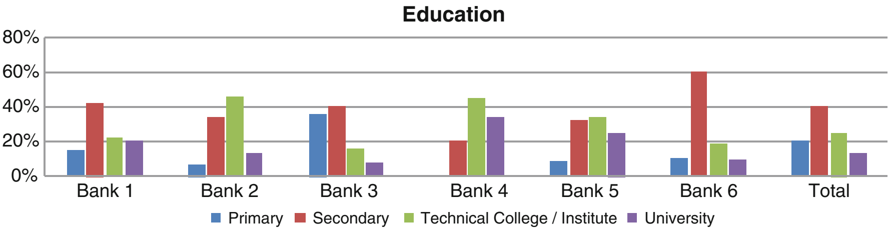

Fig. 3.3 Education in the sample

Table 3.3 Education in the sample

	Partner						
Education	Bank 1 (%)	Bank 2 (%)	Bank 3 (%)	Bank 4 (%)	Bank 5 (%)	Bank 6 (%)	Total (%)
Primary	15	7	36	0	9	11	21
Secondary	42	34	40	21	32	61	40
Technical college/ institute	22	46	16	45	34	19	26
University	21	13	8	34	26	10	13
Total	100	100	100	100	100	100	100

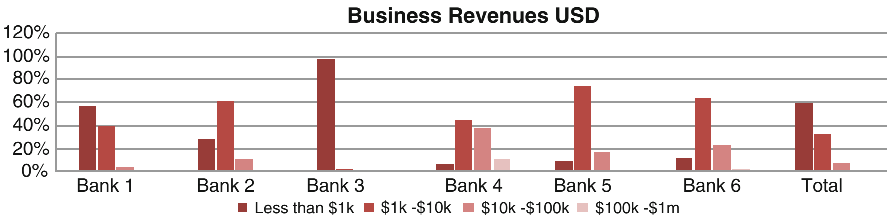

Fig. 3.4 Business revenue in the sample

Table 3.4 Business revenue in the sample

	Partner						
Business revenues USD	Bank 1 (%)	Bank 2 (%)	Bank 3 (%)	Bank 4 (%)	Bank 5 (%)	Bank 6 (%)	Total (%)
Less than $1k	57	28	97	7	9	12	59
$1k–$10k	40	61	2	45	74	64	33
$10k–$100k	3	11	0	38	17	23	7
$100k–$1m	1	0	0	10	0	2	1
$1m–$10m	0	0	0	0	0	0	0
Total	100	100	100	100	100	100	100

(Fig. 3.5, Table 3.5). Though small, these businesses are definitely not start-ups, as over two-thirds have been in operation for three years or more (Fig. 3.6, Table 3.6). This is in keeping with common patterns globally where small business lending is restricted to only established firms that can show some operational track record.

Fig. 3.5 Numbers of employees

Table 3.5 Numbers of employees

	Partner						
Employees	Bank 1 (%)	Bank 2 (%)	Bank 3 (%)	Bank 4 (%)	Bank 5 (%)	Bank 6 (%)	Total (%)
0	42	10	60	3	2	24	38
1–5	48	62	37	45	38	68	48
5–10	7	18	2	17	34	4	9
10–20	1	6	1	28	11	4	4
20–50	1	2	0	7	13	1	2
50–100	0	2	0	0	0	0	0
More than 100	0	0	0	0	2	0	0
Total	100	100	100	100	100	100	100

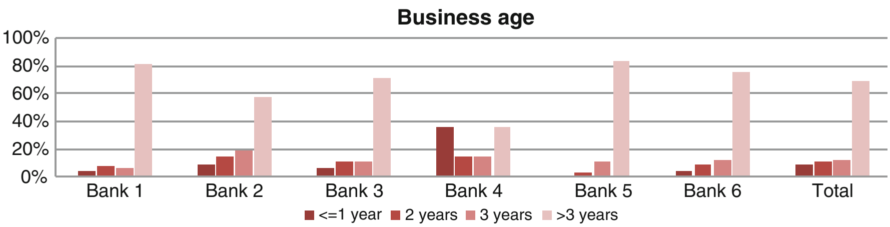

Fig. 3.6 Business age

Table 3.6 Business age

	Partner						
Business age	Bank 1 (%)	Bank 2 (%)	Bank 3 (%)	Bank 4 (%)	Bank 5 (%)	Bank 6 (%)	Total (%)
<=1 year	5	9	7	35	1	4	9
2 years	8	15	11	15	3	8	11
3 years	6	18	10	14	11	12	12
>3 years	81	58	71	35	84	76	69
Total	100	100	100	100	100	100	100

Outcome Measures

In evaluating the contribution of psychometric instruments to better identify high-potential entrepreneurs and direct finance to their ventures, there are two principal outcomes of interest: business performance (best represented with company profits) and loan repayment (i.e., did the person default).

Loan repayment is a particularly interesting outcome variable, because it is traditionally considered to be composed of two elements: ability to repay and willingness to repay.

Ability to repay will be driven by the ability of the entrepreneur. Though it is also influenced by the dynamics of the business, industry, and country in which the entrepreneur operates. Better entrepreneurs will presumably select better industries and will better adjust to and profit from changes to their environment.

Willingness to repay may be due to simple strategic calculations of the borrower (the costs of default are lower than the value of the capital retained) or driven by other individual level differences such as the level of commitment and honesty of the borrower.

Unfortunately it is usually not possible to know if an individual defaulter did not repay because of ability or willingness, but it is important to keep in mind that both factors may be contributing. And from the lender's perspective, it is not as important which of the two is the cause. What they care about is only if default can collectively be predicted and avoided with psychometric-enabled credit scoring.

Unlike default, which is independently and externally measured, business profits in this study are self-reported by the entrepreneurs completing the assessment. Therefore, the figures could be subject to misreporting. On the one hand, there is an incentive to understate profits, out of the worry that figures will be reported to the tax authorities and most small businesses underreport taxable profits. On the other hand, there could be an incentive to overstate profits, to appear more successful to the bank in case the business owner wanted to apply for another loan in the future. To minimize this risk, all entrepreneurs answered these and all other questions on the assessment on their own, outside of the view of bank officials. Moreover, they were told explicitly that their responses to financial questions would neither be shared with the financial institution nor to the government.

Business profits are self-reported and have been converted to monthly US dollar amounts. In terms of default, we adopt a definition of 30 days or more in arrears, that is, if the business owner missed a complete monthly payment cycle at any point in the past 6 months and therefore for some time owed the lender two or more payments. Because defaulters tend to make up a small percentage of total bank clients, we pursued stratified random sampling and over-sampled clients with repayment problems. This stratification was achieved with varying degrees of success, as can be seen in the summary statistics below, but resulted in an overall sample of 1,580 small business owners, just under 30 % of which had an arrears incident in recent history and were therefore labeled as "bads" to use the standard terminology of credit scoring (Tables 3.7 and 3.8).

Table 3.7 Sample size and composition by bank

Default rate at 30 days	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6	Total
Goods	240	175	371	155	64	105	1,110
Bads	173	167	68	11	23	28	470
Default rate (%)	41.9	48.8	15.5	6.6	26.4	21.1	29.7

Table 3.8 Average business profits by bank

	Partner	Mean
Business profits USD	Bank 1	1,267
	Bank 2	1,846
	Bank 3	830
	Bank 4	7,362
	Bank 5	3,606
	Bank 6	1,880

Appendix 2 features a table with summary statistics of the psychometric assessments detailed above: The Big Five, integrity, digit span, and Ravens Progressive Matrices scores. All variables have been standardized, other than digit span which is shown in number of digits, to facilitate comparisons of economic impact in the regression results.

Generally, we can see that the mean scores across the Big Five are within one standard deviation of the overall sample mean but with stronger differences between banks in the neuroticism, extroversion, and conscientiousness scales. The integrity and intelligence scales feature some differences, with banks 2, 4, and 5 clients scoring higher on both digit span and Ravens than the others. Interestingly, these are also the banks serving larger SMEs with higher levels of profits and numbers of employees (see Tables 3.5 and 3.8).

Procedure

The authors approached numerous financial institutions across Africa and Latin America to participate in this research project, eventually obtaining the participation of the six banks described above. Additional banks agreed to participate and launched testing but withdrew from the project after administering very few assessments.

For participating banks, the researchers first held a series of workshops with senior management and stakeholders explaining the goals of the testing and designed a rollout plan. While each plan varied slightly by institution, they all followed the same general setup.

First, loan officers were presented the project, to better understand their borrowers so that in the future, they could make lending decisions with more accurate and useful information about applicants. The loan officers were not mislead in any way about the nature of the testing and were instructed in detail as to the importance of engaged and active participation by their clients: The assessments had to be filled out in as "high stakes" a setting as possible, meaning the business owners were trying to "do well" on the assessment rather than completing it as a meaningless market survey. As explained below, this is important in order to simulate as closely as possible the incentives in place if this tool were implemented in practice.

Second, loan officers were trained on how to use the assessment platform. All the assessments were loaded onto a computer-based survey platform to ensure uniform implementation of interactive assessment components such as the digit span recall exercise. Officers were taught how to use the hardware, launch the assessment, and save the results upon completion.

Most importantly, loan officers were trained on how to present the assessment to clients. There was an introductory script provided to all participants, as well as an informed consent form that was signed (both available upon request). Officers were instructed to give clients any help requested with using the computer hardware (e.g., the mouse) and software (e.g., how to progress from one question to the next). In addition, they were instructed to provide any help requested by the applicant regarding vocabulary comprehension, but limited to the meaning of a word or phrase and not extending to suggested answers to questions. Officers were explicitly instructed to give no opinions on the question answers, no help on the questions with correct and incorrect answers such as the digit span and Ravens Progressive Matrices, and to ensure that clients did not use a pen and paper while taking the digit span test.

After training, loan officers tested samples of clients. In order to participate, a client had to have a loan product with the bank for at least 6 months and had to be both the legal counterpart for the loan and the owner/manager of the business. Cases where a business was co-managed by multiple owners or the loan was in the name of an owner not active in the daily management and strategic decision-making of the business were excluded from the sample. Researchers explained these characteristics to loan officers, who in some cases also received lists of testable clients from management. Loan officers were given monthly targets for numbers of clients tested, with separate targets for clients in good standing and clients in arrears. Researchers selected a small subset of tested clients from each of the four largest samples (banks 1 through 4) and directly contacted the clients to confirm that they did complete the assessment (to ensure officers weren't completing the tests themselves). Altogether, over 1,500 clients were tested across these six institutions, with response rates from 45 % to 80 % depending on the institution.

In order to give clients the incentive to actively participate, clients were informed prior to testing that if they participated, they would have the opportunity to win one of the netbook computers used for testing. Moreover, to simulate higher stakes, applicants were informed that their chances of winning would increase if they did "better" on the assessment. The assessment questions with correct/incorrect answers were graded and each participant had their name added to the draw one extra time for each correct answer. Though some clients requested feedback on their assessment scores, they were told from the outset that this would not be possible, as the measurement and interpretation of personality constructs is a specialized function of psychologists.

Clients were informed that their answers would be confidential, particularly for financial questions where this was repeated in the text of each question. The answers would not be shared neither with the bank (to avoid over-reporting to seem like "better" clients for future loans) nor the tax authority (to avoid underreporting to avoid taxes). The financial question about profit levels was not asked in the first wave of assessments and only subsequently added, reducing the number of observations against that outcome variable as compared to default behavior.