Inter-American Development Bank

INSTITUTIONS FOR DEVELOPMENT-Capital Markets and Financial Institutions Division

TECHNICAL NOTE

No. IDB-TN-532

Unlocking SME Finance in Argentina with Psychometrics

Bailey Klinger
Lucio Castro
Paula Szenkman
Asim Khwaja

April 2013

Unlocking SME Finance in Argentina with Psychometrics

Bailey Klinger
Lucio Castro
Paula Szenkman
Asim Khwaja

Inter-American Development Bank

2013

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Klinger, Bailey.
Unlocking sme finance in Argentina with psychometrics / Bailey Klinger, Lucio Castro, Paula Szenkman, Asim Khwaja.
p. cm. (IDB Technical Note ; 532)
Includes bibliographical references.
1. Small business—Finance—Argentina. I. Castro, Lucio. II. Szenkman, Paula. III. Khwaja, Asim. IV. Inter-American Development Bank. Capital Markets and Financial Institutions Division. V. Title. VI. Series.
IDB-TN-532

http://www.iadb.org

Abstract

Access to finance remains a challenge for Small and Medium Enterprises (SMEs) in Argentina. It is a particularly acute problem among newly created SMEs that have a scant credit record and frequently lack collateral. The authors of this paper carried out a pilot test of an innovative psychometric tool aimed at evaluating credit risk among SME clients of Banco Ciudad de Buenos Aires (BCBA), one of the top three public banks of Argentina. SME responses were compared to their historical repayment records to determine if psychometric-based credit scoring models could help ameliorate the constraints to SME finance. A scorecard based on that information and built using international data results in a Gini coefficient between 20 percent and 40 percent: those SMEs rejected by the psychometric-based scorecard with this Gini coefficient have a probability of defaulting that is three to four times greater than those accepted. Along with other policies to reduce information asymmetry in SME lending, such a tool could help relieve constraints to SME finance in Argentina.

Keywords: SME, credit, risk, scoring, Argentina, Peru, bank, finance, psychometrics
JEL Codes: N26, G21, G32, H32

Table of Contents

Figures and Tables

1. Introduction

Access to formal finance remains a challenge for the majority of the Small and Medium Enterprises (SMEs) of Argentina. According to the World Bank's 2010 Enterprise Survey (WBES), only one-third of Argentina's SMEs have access to banking services and use formal credit to fund their investments. Restricted access to finance is particularly acute among newly created SMEs, which tend to have a scant credit record and frequently lack collateral. Recurrent financial and currency crises coupled with historically high inflation rates have notoriously hindered the development of long-term financing in Argentina, particularly for SMEs.

As a response to this situation, the Center for the Implementation of Public Policies for Growth and Equity (CIPPEC), a public policy think tank based in Buenos Aires, and Entrepreneurial Finance Lab (EFL), a financial firm nurtured in Harvard University, launched a pilot test of an innovative tool in January 2012, with the support of the Better Conditions for Productivity (MAP) fund of the Inter-American Development Bank (IDB). This tool is aimed at evaluating credit risk among SME clients of Banco Ciudad de Buenos Aires (BCBA), one of the top three public financial institutions in Argentina.

Between May 2011 and February 2012, CIPPEC tested 255 SME clients with a collection of psychometric assessments; clients were randomly selected from a sample of 1,000 SMEs that had applied for a loan from BCBA and had been approved by the bank's screening and scoring system. To determine whether credit scoring models, including psychometric data, could help ameliorate the constraints to SME finance in Argentina, each SME owner's responses to the application were compared to that SME's repayment history with BCBA.

The results of this analysis were consistent with past studies showing that average responses to psychometric-based questions differ in Argentina, compared with other countries. However, the psychometric application's key indicators that have been statistically related to default in other countries remain so in Argentina. A scorecard based on these questions and built using international data produced a Gini coefficient between 20 and 40 percent in the BCBA data.

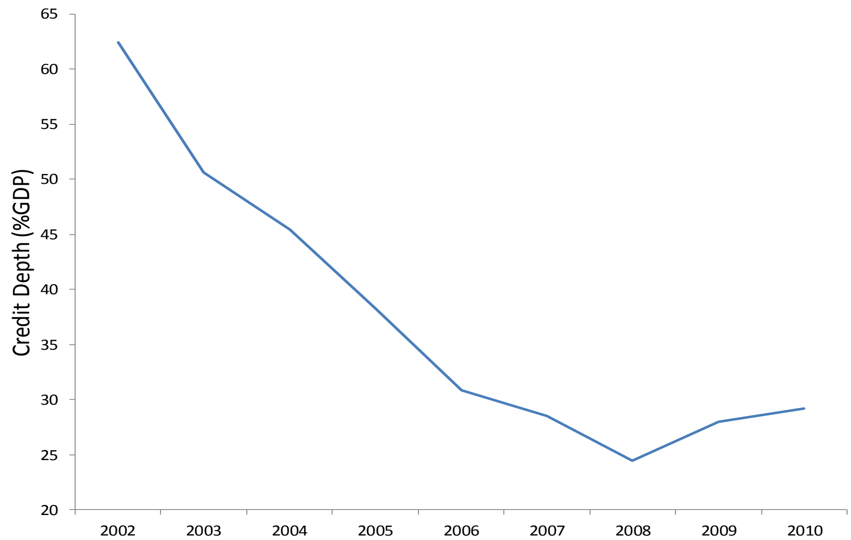
This is powerful risk separation: SMEs that are rejected by a psychometric-based scorecard with this Gini coefficient are three to four times more likely to default than those accepted. Such information would give banks such as BCBA greater confidence in the SMEs it lends to, and would allow SMEs that previously could not provide sufficient reliable information to control risk to be considered for loans. Along with other policies to reduce information asymmetry in SME lending, such a tool could help relieve the constraints to SME finance in Argentina, thus stimulating economic growth.

This paper is structured as follows. Section 1 describes the characteristics of Argentina's financial system, with particular attention to constraints that SMEs experience in accessing formal finance. Section 2 analyzes existing credit risk systems for commercial loans in Argentina and provides a glimpse of BCBA's recent history and performance, as well as a brief description of its analysis systems for lending to private companies. Section 3 describes some common challenges to SME risk analysis in Argentina and other developing countries and proposes a solution: applying psychometric tools to evaluate potential credit performance of SMEs. Section 4 provides a short description of the pilot trial of a psychometric tool developed by EFL involving a representative sample of BCBA's SME clients. Section 5 concludes by presenting the main results and policy recommendations of this pilot.

2. Access to Finance in Argentina

Despite achieving rapid and sustained economic growth in the aftermath of the 2002 crisis, Argentina remains a relatively small and underdeveloped financial market. Over the last decade, robust domestic investment and consumption have not been accompanied by a parallel expansion in domestic credit as a percentage of GDP (Figure 1).

Figure 1. Domestic Credit as Percent of GDP in Argentina, 2002–10

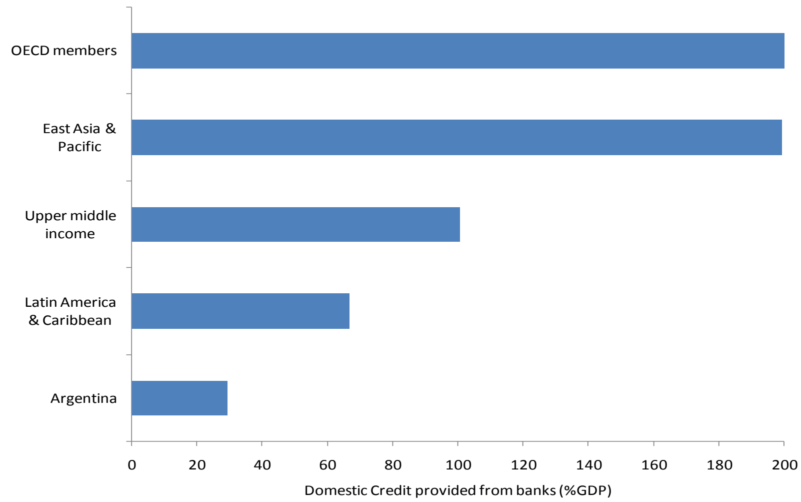


Source: Authors' calculations based on WDI (2012).

The sluggish development of Argentina's financial system is largely a result of the twin financial and currency crises suffered in 2002. Other contributing factors involve the country's difficulties to entice foreign investment (partly as a result of the still unresolved 2002 external debt default), the negative effects of double-digit inflation rates on financial intermediation, and the imposition of distortive taxes on financial transactions, and, more recently, on capital and exchange-rate controls (Castro, 2010).

Accordingly, financial depth in Argentina, measured by domestic credit as percent of GDP, is less than half that of the rest of the Latin American region (30 percent versus 65 percent), three times less than the depth of economies with a similar income per capita, and much less than that of the OECD and East Asia and Pacific (Figure 2). In turn, bank deposits from the private sector stand at 17 percent of GDP and most deposits are short term and denominated in local currency (EIU, 2010).

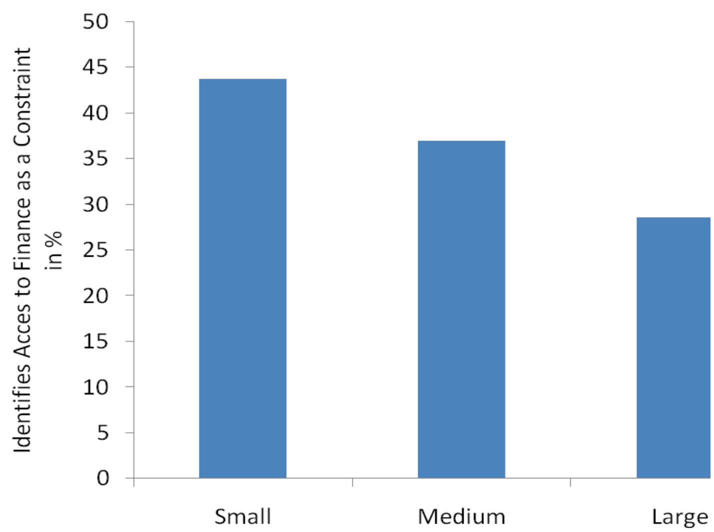
Figure 2. Financial Depth as Percentage of GDP, 2010



Source: Authors' calculations based on WDI (2012).

Access to banking finance and use of it is particularly low among SMEs. While fewer than 30 percent of the largest companies point to such access as a major constraint to investment, 40 percent of SMEs identify financing as one of their most serious obstacles (Figure 3).

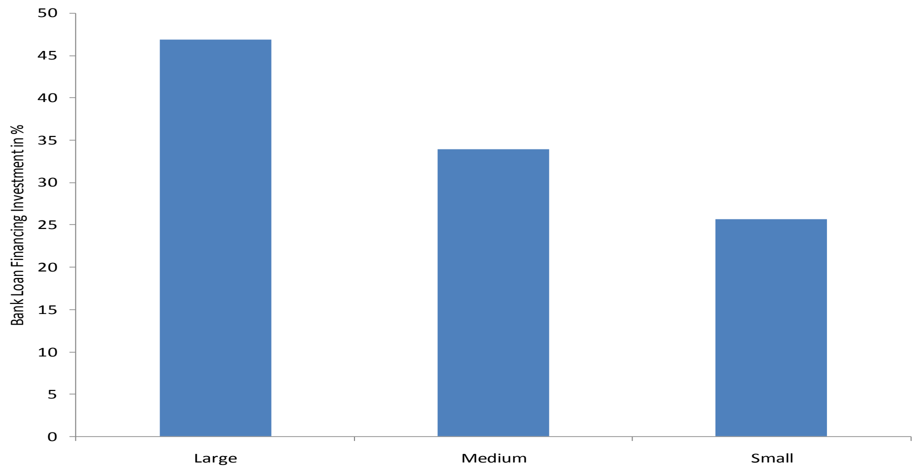
Figure 3. Percentage of Firms Identifying Access to Finance as a Major Constraint, 2010



Source: Authors' calculations based on WBES (2010).

Figure 4 shows that, while half of the large firms draw on some form of banking finance, only 25 to 35 percent of SMEs use formal loans. Hence, for SMEs, internal funds account for two-thirds of investment financing.

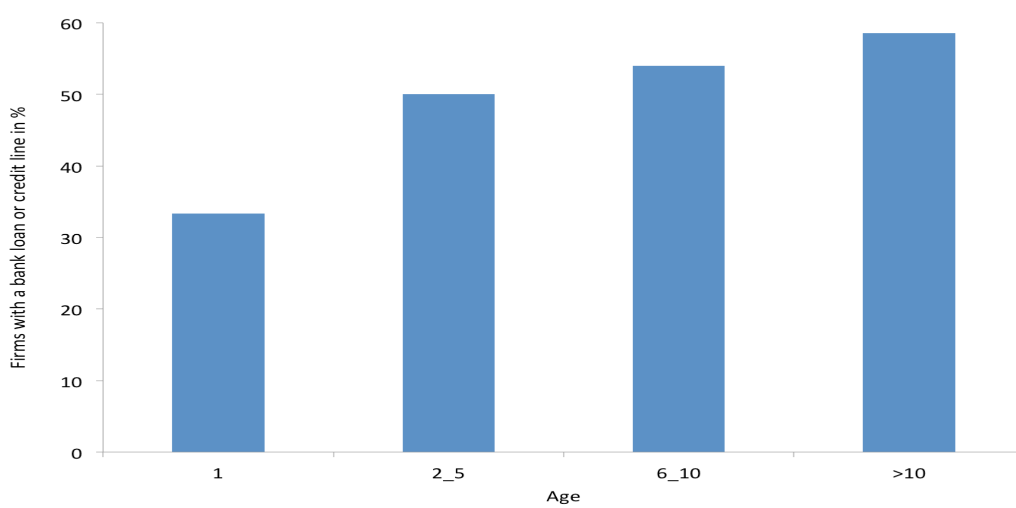
Figure 4. Firms by Size Using Bank Loans to Finance Investment, 2010 *(in percent)*



Source: Authors' calculations based on WBES (2010).

Restricted access to finance is particularly severe among newly created firms. Figure 5 indicates that while 60 percent of companies with 10 or more years have a loan or a credit line, only 30 percent of enterprises with a year or less are able to obtain formal financing.

Figure 5. Firms with a Bank Loan or Credit Line by Age, 2010



Source: Authors' calculations based on WBES (2010).

Two recent nationwide firm-level surveys have shown that the most significant impediments for use of banking finance among private companies, particularly SMEs, relate to macroeconomic uncertainty, tax changes, the regulatory framework, and high financial costs (OPyME, 2009; BCRA, 2010a). Furthermore, distribution of credit to the private corporate sector exhibits a high territorial concentration. For instance, in the Autonomous City of Buenos Aires (or CABA, according to its Spanish acronym) 43 percent of all lending goes to private companies. The surrounding Pampeana and Centro regions represent an additional 43 percent of total loans granted to private enterprises (BCRA, 2010b).

3. Credit Risk Systems in Argentina

This section presents an overview of regulation, overseeing, and functioning of credit risk systems for commercial loans in Argentina. In addition, a brief description of BCBA's recent history and performance is given, along with an abridged depiction of BCBA's credit risk assessments for loans to private companies.

3.1. Credit Risk for Commercial Loans

The Central Bank of the Argentine Republic (BCRA) oversees regulation and operation of Argentina's financial sector. The BCRA establishes guidelines for classifying debtors considering their credit quality and compliance with commitments. These guidelines lead to the classification of commercial loans into five main categories: normal, special follow-up, substandard, high insolvency risk, and unrecoverable (BCRA, 2010a).

Evaluating commercial loans usually involves considering liquidity, financial structure, payment behavior, governance and management standards, information technology systems, outlook for the borrower firm's main activity sector, the firm's legal status, and the existence of refinancing procedures or write-offs. To facilitate lending to SMEs, the BCRA establishes that commercial loans under 750,000 Argentine pesos (about US$190,000 dollars at the official exchange rate[1]) can be classified, at the lender's option, using the more flexible parameters

[1] Reference exchange rate: 3,946. September 2010. http://www.bcra.gov.ar/estadis/es010100.asp?descri=19&fecha=Fecha_Ref&campo=Tip_Camb_Ref

employed for consumption loans. The BCRA's top authorities must approve ratings of borrowers with amounts exceeding 2.5 percent of the bank's regulatory capital (Bleger, 2011).

The BCRA also requires commercial banks to keep a record of each credit application, along with basic information for the applicant company (e.g., balance sheet, business plan, and other financial data). Commercial banks also must set up a permanent payment and compliance monitoring system.[2]

Argentine banks employ different credit risk methods, which can be automatic, semi-automatic, or manual. Some banking institutions utilize econometric techniques to estimate credit scores and default probabilities. Other financial institutions employ risk matrices based on companies' financial information, while the rest rely on the analysis of credit officials based on simple rules of thumb to determine whether to approve a loan.

The BCRA also requires that commercial banks keep records of each credit application, including basic information for the applicant company (e.g., balance sheet, business plan, and other financial data). Commercial banks also must set up a permanent payment and compliance monitoring system.[3]

3.2. Banco Ciudad de Buenos Aires

BCBA is a large state-run bank in Buenos Aires, Argentina. It is a semi-autonomous financial institution of the Government of the CABA. Founded in 1878, BCBA has 23 branches scattered across the CABA, the adjoining Greater Buenos Aires, and the city of La Plata.

In 2007, BCBA was suffering severe financial losses. The share of nonperforming loans had reached a historical peak after the financial and currency crises of 2002; BCBA was not even among Argentina's top 20 financial institutions. Early in 2008, Federico Sturzenegger, a renowned local academic, was appointed BCBA's executive chairman by the CABA government. He launched a series of radical reforms aimed at improving the bank's profitability and efficiency (Musacchio et al, 2011).

Largely as a result of these efforts, BCBA is now Argentina's ninth largest bank and the city's third largest public bank, as measured by total deposits, total loans, and net assets as of

---

[2] http://www.bcra.gov.ar/pdfs/texord/t-cladeu.pdf
[3] Ibid.

December 2011. The proportion of nonperforming loans in total lending is at a minimum (1.1 percent), even below the financial sector's average (1.2 percent). Furthermore, BCBA's financial performance currently outshines the rest of the country's national and provincial public financial institutions (ABA, 2011).

Presently, BCBA directs more than half of its lending to the private nonfinancial corporate sector and the rest to personal loans and mortgages (ABA, 2011). Since 2008, BCBA's lending to the private sector has increased by 198 percent, and the amount of loans granted to SMEs has risen four times in nominal terms.[4]

3.3. BCBA's Credit Risk Analysis for Commercial Loans

Following BCRA's regulations, BCBA utilizes a risk classification matrix for commercial loans over US$34,000. The matrix is a semi-automatic expert score that comprises a series of financial ratios based on historical financial information: liquidity, profits, solvency, cash flows, current debt, and sales. Additionally, some qualitative variables are considered by the matrix: balance sheet information, actual commercial performance, and payment records, which are mostly drawn from public and private credit bureaus.

To determine risk associated with each loan application, the BCBA risk matrix includes 70 percent assigned to these qualitative variables, with the other 30 percent assigned to quantitative indicators. It is noteworthy that the matrix provides only a general guideline for credit approval; the final decision remains a prerogative of BCBA's risk analysis department. Furthermore, credits over US$170,000 are granted only with the approval of BCBA's board of directors. Presently, BCBA is developing a new credit risk system based on a more sophisticated quantitative scoring, which is expected to replace the existing expert score matrix. The new system is also aimed at automating most of the commercial credit applications and approval processes.

[4] http://www.bancociudad.com.ar/institucional/mision,-vision-y-valores

4. Barriers in Risk Assessment for SMEs

Credit risk analyses in Argentina reflect the typical approach and offer numerous advantages, but continue to face some challenges. These challenges also exist in other countries around the world, resulting in higher costs, higher risk, and less lending for financial institutions, in addition to restricting financial access for SMEs. This section summarizes these challenges and proposes a solution.

4.1. Shortcomings of the Current Approach

Financial ratios, which are often difficult to verify, can be either incorrect or nonexistent for smaller firms. Moreover, they reflect past financial performance rather than future performance, which is the key driver of risk for loan repayment. This fact is particularly important in countries with strong economic cycles, as is true of Argentina: two firms may perform well in good times, but which of the two will perform well when economic conditions change? This question is difficult to answer solely looking at historical financial information.

Rules of thumb and subjective assessments by credit officers can incorporate a wider set of information for risk assessment; unlike financial ratios, these can be performed for any business (to the extent that they can provide information requested by an officer). However, these types of assessments are often only as good as the officers performing them. Experienced officers may be finely tuned credit risk analysts who, after years of experience, can incorporate many subtle signals of borrower quality, but it is expensive for analysts to accumulate this experience. Also, when an individual officer leaves the bank, the experience is gone as well. A lender's need for personnel training and experience, along with losses from turnover, create limits on how quickly SME lending can grow; it takes a long time and a large investment to build and retain the required army of experts. Moreover, the high costs per evaluation mean that only large loans are profitable, limiting the extent to which banks can serve smaller businesses. Banks can rely on less experienced (and less costly) loan officers, but they face an increase in risk.

Thus, banks have less information available to evaluate risk and are limited in their ability to gather more information. This leads to a difficult combination of challenges: how can financial institutions adequately evaluate risk for information-scarce SMEs, and do so while keeping transaction costs low?

4.2. Potential Contribution of Psychometrics

Industrial and organizational psychology has been working for decades on a problem with similar characteristics: how to screen people applying for jobs. Firms must decide which individuals to hire, often based on little available information. Moreover, particularly for entry-level positions, firms must evaluate a large number of applicants in a low-cost way. In response to this challenge, industrial psychology has developed a series of assessments of individuals that are predictive of a person's future success in a job. This field, which assesses skills, abilities, personalities, and intelligence, is known as *psychometrics*, whose literal meaning is "measurement of the soul."

There is also a large established industry using psychometric testing for preemployment screening. According to a 2001 survey by the American Management Association, 29 percent of employers use psychological assessments of employees for selection and development. Psychometric testing of job applicants is expanding quickly because of its effectiveness. A meta-analysis published by the American Psychological Association found that tests of general cognitive ability have the highest validity of all selection methods. These tests outperform screening based on personal interviews, biographical data, and past education and experience. The combination of a cognitive ability test and an integrity/honesty test has one of the highest composite validities of any combination of techniques (Schmidt and Hunter, 1998).

To understand these tools' applicability to financial services, it is relevant to note that job selection is a high-stakes environment where some applicants have very powerful incentives to attempt to "game" the psychometric instruments so as to appear highly qualified for a certain job. Because of these incentives, the instruments used for preemployment screening have been designed to prevent cheating and have proven highly effective, despite an applicant's less honorable attempts (Hogan, Barrett, and Hogan, 2007).

In addition to their practical use in preemployment screening, psychometric instruments have been extensively used to study the characteristics of good entrepreneurs. These studies typically seek to identify differences between people who become entrepreneurs as opposed to managers or other salaried employees. The literature dates back to McClelland's (1961) seminal work, and has considered a wide variety of psychometric tools over the years. Baum, Frese, and Baron (2006) and Chell (2008) provide both systematic literature reviews and meta-analyses.

Given that psychometric tools have proven to be effective screening tools in the human resources field despite the high stakes and incentives to cheat the system, and given that these tools have also been used to distinguish entrepreneurial characteristics among individuals, the questions arise: could psychometric instruments help resolve the shortcomings of the traditional approach to credit risk analysis for SME borrowers, and could they increase the currently restricted access to finance among Argentina's SMEs? To answer these questions, a pilot trial of a collection of these tools with BCBA was performed, which is described in the following section.

5. Pilot Trial Implementation

Between May 2011 and February 2012, a pilot trial was conducted with BCBA. CIPPEC gave an assortment of psychometric assessments to a representative sample of existing SME clients. This section describes the strategies to select participants, the pilot implementation, and the methodology for evaluating the potential for such screening tools to improve financial results for BCBA and financial access for SMEs in Argentina.

5.1. Methodology

The wide variety of psychometric assessments that were implemented were based on literature reviews and new studies by EFL. The EFL tool was administered to 255 SMEs, randomly selected from a sample of 1,000 SMEs that had applied to a loan from BCBA and were approved by the bank's screening and scoring system. The sample included loans that were current, under revision, and underperforming. Further, 20 of the 255 SME loans experienced delays of at least 30 days in their payments. Finally, the sample comprised SME clients from the BCBA's 23 branches, and the distribution of cases was proportional to each branch's SME portfolio.

It was decided to assess existing SME clients because testing new clients and then observing their ex post repayment performance (in some ways, a cleaner test) would mean waiting a significant amount of time before loans matured (often many years); also, such an assessment would require enormous sample sizes because stratification on arrears status is not possible ex ante (one does not know who will default). Since the dimensions evaluated by these assessments are stable over time among adults (Costa and McCrae 1994), and the sample is made up of adults with an average age of 40, it is possible to test applicants today on the basis of repayment history, discarding the possibility of reverse causality (those in arrears causing different responses on the psychometric instrument).

5.2. Implementation

In collaboration with the BCBA, CIPPEC carried out a three-fold implementation strategy for the trial pilot of the EFL assessment tool; this strategy aimed to (a) recruit and contact each SME's "decision maker"; (b) administer the tool; and (c) monitor an adequate data collection.

(a) Recruitment

The BCBA authorities appointed a team of SME credit officials to work under a coordinator with CIPPEC's team. Each official was assigned a recruitment target, based on the sampling strategy. Thus, officials from branches with the largest SME portfolios were assigned the largest recruiting targets.

The recruitment procedure commenced with a random selection of SMEs. Participating officials contacted a firm and invited the owner/manager (or other key decision maker) to voluntarily participate in the pilot. Each official read the selected client a standardized message, which was designed exclusively for this purpose.[5] If the client agreed to complete the test, the official prepared a contact list containing relevant information from all participants and provided it to CIPPEC. Finally, CIPPEC contacted potential participants in order to schedule an appointment for survey administration.

At this point, the following adverse situations took place: (i) the client constantly postponed the appointment; or (ii) the client refused to participate, despite having agreed to be tested. In such cases, CIPPEC informed the BCBA coordinator, who instructed the officers to replace that client for sampling.

(b) Test Administration

Three portable computers were especially acquired to administer EFL's test. The laptop's portability gave surveyors the flexibility to schedule simultaneous appointments and cover an ample geographical area. Clients were offered two alternative appointment location options: a

[5] The complete text of the recruitment message is presented in Annex 1.

corresponding BCBA branch (each branch had a meeting room available for this purpose); or the SME's own facilities.

Once an appointment had been scheduled, a surveyor visited the SME, briefly introduced the project and offered basic instructions. Each participant was provided with a unique identification number (ID) that had to be typed at the commencement of the test. Once the test was completed, the results, ID, and date were stored in the laptop. While taking the test, no extra assistance or external disruptions were allowed. Surveyors were present to prevent cheating or any external influence during the test. They also assisted the participant with any technical issues that might arise.

(c) Data Monitoring

Every week the laptops were synchronized in order to send test results to EFL. Subsequently, in order to check the correct reception of data, CIPPEC sent a report with the number of tests completed and corresponding IDs. A similar report was provided to the BCBA, stating which firms effectively completed the test, which ones refused to participate or presented difficulties in scheduling an appointment, and which firms needed replacement. Any distinctive situation that had occurred was also reported. These weekly reports helped ensure a correct completion of the survey during the 10 months of the fieldwork.

Once the tests were completed, CIPPEC prepared a list with the 255 SME clients that had effectively completed the assessment and provided the list to the BCBA, which was in charge of building a microdataset regarding credit repayment history. On February 15, 2012, CIPPEC received the complete credit history of the 255 firms.

The ID number created by CIPPEC ensured an appropriate matching between a given client's repayment data and the test results, while protecting the client's identity and complying with the BCBA's confidentiality requirements. The final result consisted of repayment data for each SME from the date of approval of the credit to the date of dataset construction.

(d) Training Sessions

A first training meeting with officials from three selected branches took place on April 26, 2011. At this stage, CIPPEC and BCBA also carried out a four-week pilot with a small sample of 20

SMEs, aimed at refining and adjusting the implementation strategy. For instance, some adaptation of the vocabulary used in some of the test questions was required.

In June 2011, four training sessions were carried out with the credit officials selected by BCBA. The sessions were structured around three topics: (a) project's motivation and objective; (b) EFL's test; and (b) implementation strategy. A standardized text was also prepared for credit officials to use when calling the selected SMEs. Given that some of the test questions address sensitive personal issues, the text stressed two points: (a) test results were completely confidential; and (b) results would not influence the client's standing with the bank. In addition, officials mentioned that participants would be entered into a lottery for three laptop computers. The complete text of the standardized message can be found in Annex 1.

Given their proximity and personal knowledge of the selected companies, BCBA officials played a crucial role during pilot implementation. They were also instrumental in identifying key decision makers within participating SMEs. Overall, fluent communication and solid coordination between BCBA officials and CIPPEC was critical to ensure a successful implementation.

6. Results

Psychometric principles have been successfully applied to credit screening across Africa. To assess the potential impact of these tools in Argentina, the present study compares the distributions on a subset of psychometric indices between Argentina and other countries, and then analyzes the overall predictive power on the BCBA sample of an internationally derived psychometrics-based scorecard.

6.1. Cross-Country Results

There is a debate in the literature about the consistency of psychometric instruments across countries and cultures. Strong cultural differences between countries are often reflected in different average answers and scores on a wide variety of psychometric tools (see, for example, the World Values Survey (available at WVS at www.worldvaluessurvey.org). Yet significant evidence exists that across countries as diverse as Canada, China, Israel, Japan, the Philippines, Poland, and South Africa, major psychometric constructs such as the "Big Five" personality structure (the predominant personality model in psychology) remain a consistent construct

(McCrae et al., 1998). The present study first compares data collected across five countries: Argentina, Kenya, Nigeria, Peru, and South Africa. This sample represents a wide variety—in terms of culture and business size—of businessowners (see Table 1).

Table 1. Cross-Country Summary Statistics for Participating Firms in EFL Pilots

Country	Median monthly Sales of entrepreneurs (USD)	Average loan size (USD)	Sample size
Argentina	$100,000 - $1m	$50,000	256
Kenya	$1000 - $10,000	$5,500	4931
Nigeria	$10,000 - $100,000	$22,000	5561
Peru	Less than $1000	$3,000	126
South Africa	$10,000 - $100,000	$150,000*	91

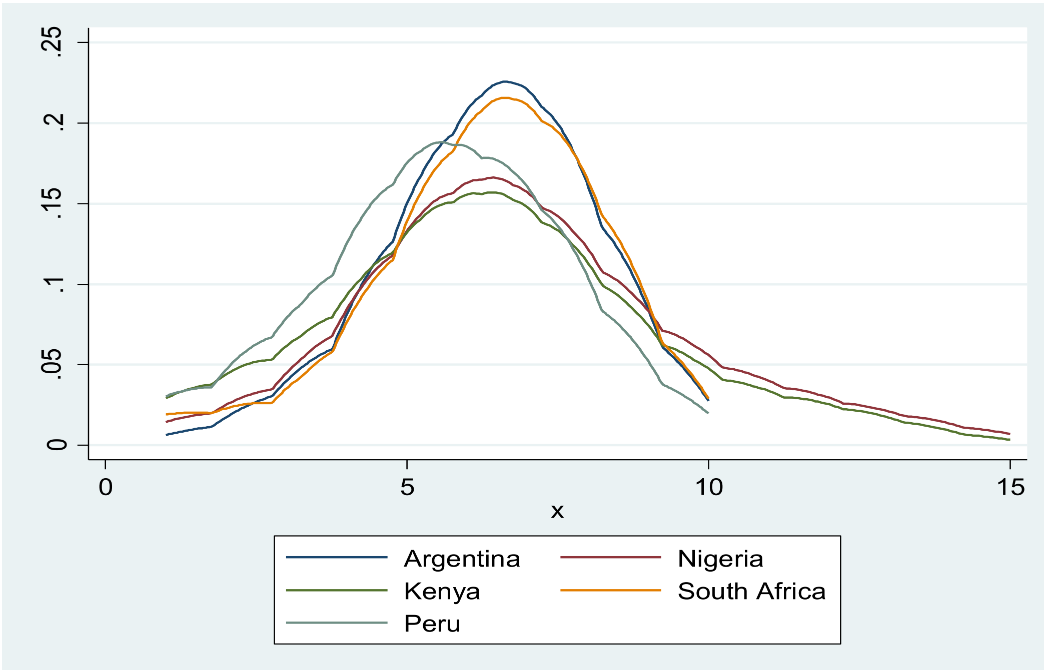
*In the case of South Africa, these are combination loans plus equity investments. See www.businesspartners.co.za for more information.
Source: Authors' calculations based on data from financial institutions.

Could measurements of things such as personality traits and intelligence in Africa be relevant for a country like Argentina? The following is a discussion of the comparative distributions for a variety of personal characteristics across these countries.

First is fluid intelligence. Intelligence is often thought of as a person's general ability to think and act rationally. The American Psychological Association formed a task force on intelligence research and defined it as the "ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, and to overcome obstacles by taking thought" (Neisser et al., 1996). Fluid intelligence describes the ability, regardless of prior skills or experience, to grasp relationships between abstract concepts. Fluid intelligence is usually measured by tests of problem-solving ability, pattern recognition, and working memory (Snow and Yalow, 1982).

Intelligence has been repeatedly shown to be the best single predictor of job performance in both entry-level and advanced vocations (Hunter, 1986; Ree et al., 1994). Because pattern recognition plays a strong role in profitable entrepreneurship (Baron, 2006) and because successful entrepreneurs must be a "jack of all trades" rather than a specialist (Lazear, 2005,) fluid intelligence could play an even stronger role than it plays in general vocational performance.

Figure 6. Digit Span Recall Results by Country



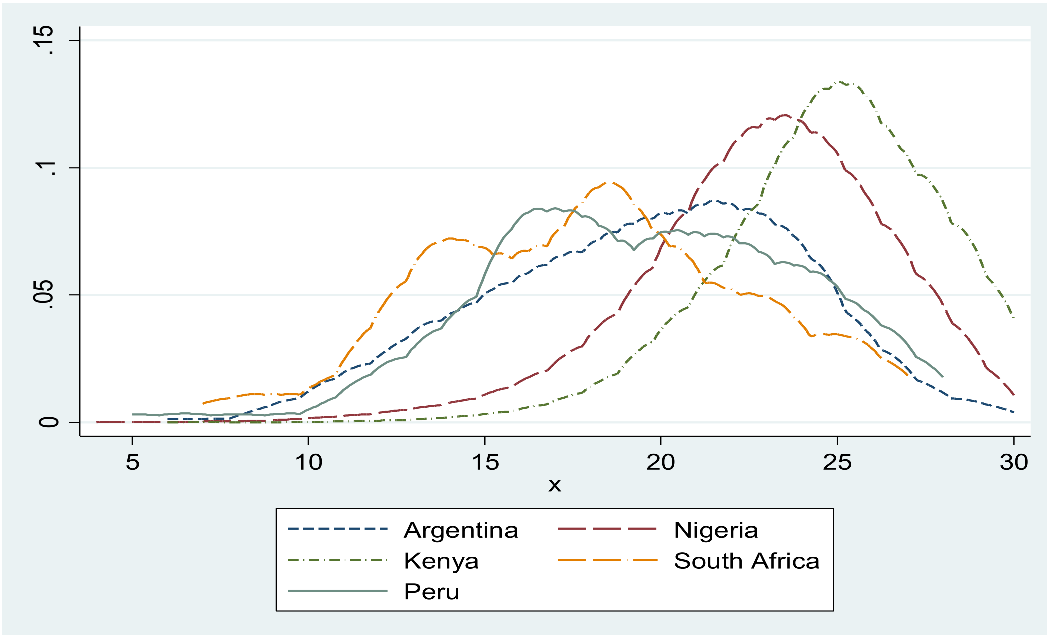
Source: Authors' calculations based on pilot implementation data.

In Figure 6, unlike the remainder of psychological indices shown below, the cross-country results exactly match with business size. That is, in countries where our data is for smaller firms with lower revenues (see Table 1), average digit span recall scores are lower, and in countries where our data is for firms with larger revenues, average digit span recall scores are higher. For Argentina and South Africa, the two countries with the largest businesses in this data, distribution of digit span recall scores is almost identical. The smallest borrowers, in this case microfinance clients in Peru, have the lowest distribution of digit span recall scores. This relationship seems to overpower other factors that may be more cultural, given that purely cultural factors would suggest that Latin American and African countries are more similar to one another than to countries of other regions. This positive relationship between revenue levels and digit span recall holds within each country as well as across countries: a regression of business revenues on digit span recall controlling for country shows a positive and significant relationship at the 1 percent significance level. In other words, digit span scores distinguish high from low revenue businesses within each country, not just between them. This finding suggests that in the case of fluid intelligence, at least, larger businesses tend to be run by higher-scoring entrepreneurs, implying that measuring intelligence by a digit span recall test could distinguish between more and less successful entrepreneurs, and that cultural factors across countries might be less relevant to success. It is not that African entrepreneurs are overall higher or lower scorers than Latin American entrepreneurs, or vice versa.

The second set of psychometric dimensions examined is a subset of the five-factor or Big Five personality model (Barrick and Mount, 1991), which is the dominant model of personality structure today. There is much literature examining the relationship between personality dimensions and entrepreneurship, going back to McClelland (1961). Strong relationships have been found between Big Five dimensions and the likelihood of becoming an entrepreneur (Seibert and Zhao, 2006), as well as entrepreneurial business survival (Ciavarella et al., 2004).

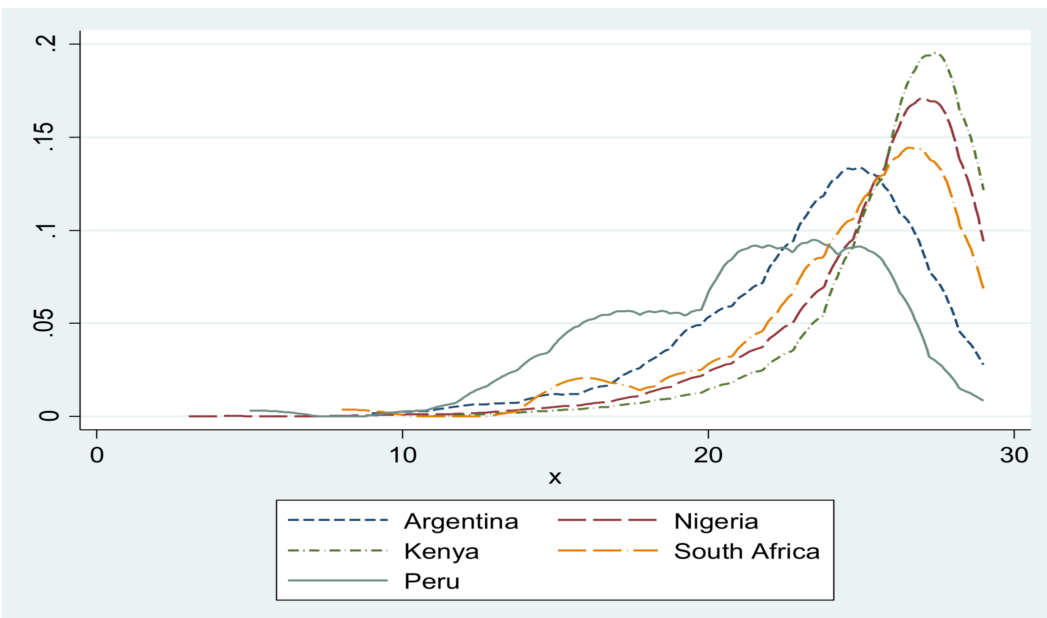
However, compared to nonverbal intelligence tests, verbal questions related to psychological traits and beliefs could produce greater variation in scores due to cultural context. Evidence for this conclusion is provided below (Figures 7–10), comparing the distribution of raw scores for four of the Big Five personality traits: conscientiousness (the degree to which a person seems conscientious, conforming, and dependable); extraversion (the degree to which a person seems socially self-confident, leader-like, driven, and energetic); emotional stability (the degree to which a person appears calm, self-accepting, and steady in the face of pressure); and openness (the degree to which a person is perceived as bright, creative, and interested in intellectual matters). Agreeableness could not be included because of the instrument's space constraints. The x-axis is the standardized score (in all cases, a higher score means that the measured trait is higher or stronger), and the y-axis is the frequency of that score in the population indicated by the color/pattern of the line.

Figure 7. Conscientiousness Cross-Country



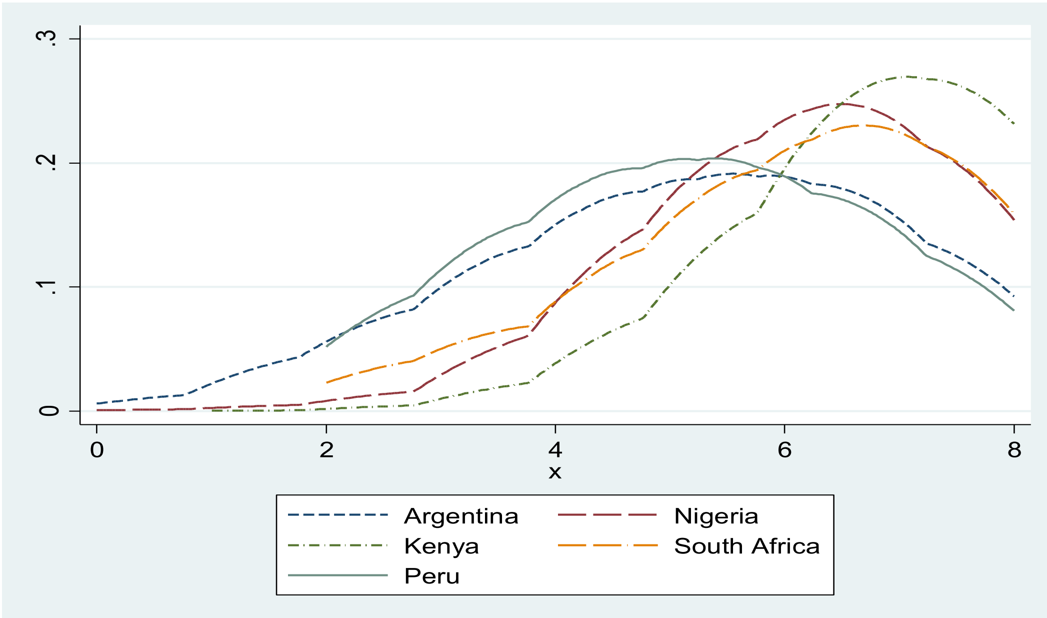
Source: Authors' calculations based on pilot implementation data.

Figure 8. Extraversion Cross-Country



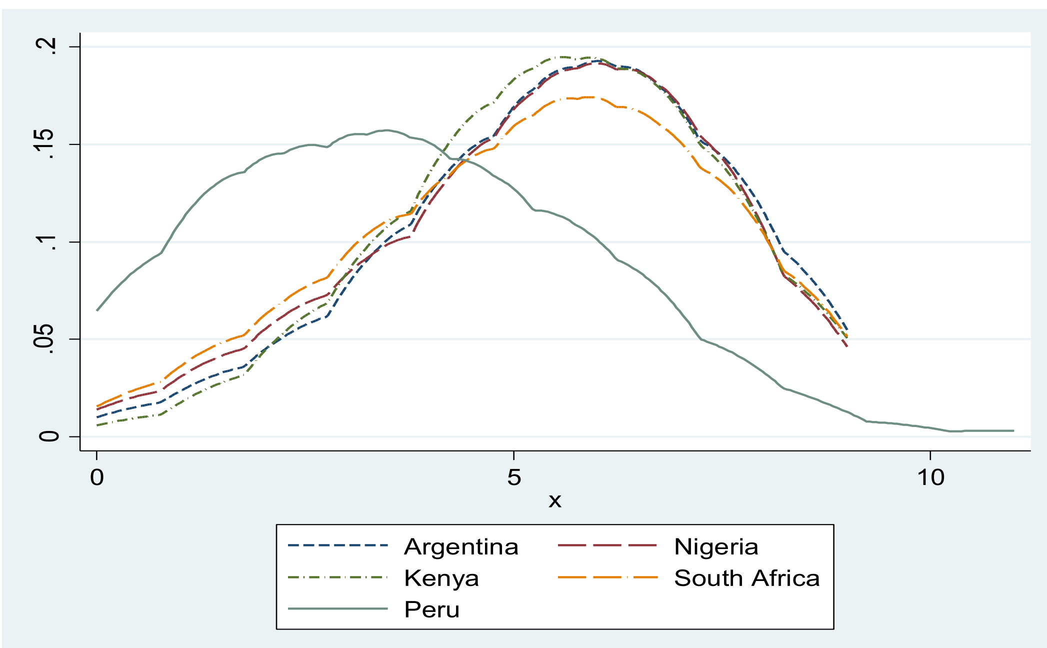
Source: Authors' calculations based on pilot implementation data.

Figure 9. Openness Cross-Country



Source: Authors' calculations based on pilot implementation data.

Figure 10. Emotional Stability Cross-Country



Source: Authors' calculations based on pilot implementation data.

Unlike the case of digit span recall, these personality dimensions show no clear pattern relating to business size. Moreover, differences do seem to exist across regions. For example, in the case of conscientiousness, a t-test for equality of means shows only one country, Peru, with a statistically indistinguishable mean from Argentina (i.e., the Latin American countries are statistically indistinguishable), and the average for Argentina is significantly different from all three African countries at the 1 percent level. In terms of levels, values are lower for Peru and South Africa, but higher for Kenya and Nigeria. In the case of openness, Peru and Argentina have a similar distribution of scores, which are lower than those of the African countries. In the case of extraversion, Argentina lies between Peru and the African countries.

Overall, these results suggest strong differences in distribution of personality assessments between countries, and these differences do not directly correlate with business size. While regional patterns are not exact, Latin American countries and African countries tend to be more similar to other countries within their region than to countries in other regions, with some exceptions. Exceptions to these regional patterns occur, however; for example, Peru is an outlier on one of the Big Five dimensions (emotional stability), and Argentina is more similar to the African countries in that regard. But overall, these results suggest that values or traits revealed by measuring psychometric dimensions are influenced by language and culture, and might not migrate easily from country to country.

6.2. Predicting Credit Risk

But what are the implications for entrepreneurial performance and credit risk? For a psychometrics-based credit scoring tool, it is not necessary that the measured traits have equal meanings and distributions across countries. It is more important that the same dimensions predict risk within each context. For example, in determining whether height is a good predictor of success at basketball, it does not matter that average height is lower in one country and higher in another; what matters is that within all countries, height is important for basketball players, so basketball abilities can be predicted by comparing players' heights within each country.

An emerging study by Patterson et al. (2012) examines an issue that is similar, but related more to managerial effectiveness than to entrepreneurial success. The authors studied whether the importance of key skills and attributes such as taking action, making decisions, following through, developing relationships, and having drive and ambition vary by national culture (divergence) or are universally similar (convergence). Despite some research showing cultural differences in management styles and behaviors, proponents of the convergence hypothesis suggest that globalization and the prevalence of multinational corporations can cause cultural norms and values to converge. The authors tested managers across 30 countries and found support for the convergence hypothesis: that, indeed, the importance of these skills and attributes for successful management are not shaped by culture, but are similar across all 30 countries.

To the extent that managerial and entrepreneurial tasks are similar, these findings support the hypothesis that the particular traits affecting entrepreneurial performance (and thereby, credit risk) could also converge across cultures, allowing similar instruments to be used in multiple countries. Just as companies are increasingly working across borders, small businesses and the firms that train and invest in them are also increasingly transnational. Nongovernmental organizations, such as TechnoServe and Endeavor, identify, select, and train entrepreneurs across nearly all regions of the world. The emerging industry of "impact investing," which focuses on bottom-of-pyramid businesses in developing countries and their environmental and social impacts, is also applying selection methodologies and sending investment officers across a wide variety of countries (see, for example, members of the Aspen Institute's ANDE network). The ability of individuals to source deals and select entrepreneurs across multiple countries and cultures suggests that at least some of the drivers of entrepreneurial ability are similar.

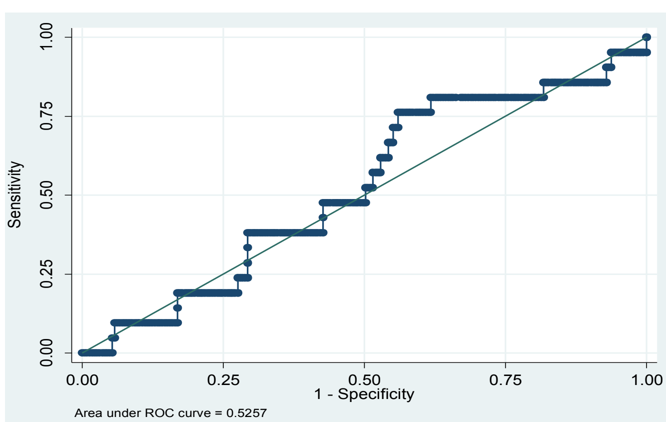
Credit scoring models are also used to order risk among individual applicants within a particular country or for a particular bank, market segment, or product. Therefore, whether the findings of a psychometric-based credit application can be generalized across countries and cultures merely requires that the same questions be able to distinguish default risk within different groups. The average answers to questions traditionally used to evaluate conscientiousness might be true more often in one culture than in another, but a scorecard based on those questions would still be effective if, within each of the two cultures, entrepreneurs that have higher credit risk systematically tend to give similar answers when compared to entrepreneurs with lower credit risk. Thus, psychometrics-based credit scoring is a purely empirical exercise.

Can the same default-predicting model work in Argentina that works in other countries? The present sample of Argentine SME borrowers contains 255 entrepreneurs, including 21 who entered into arrears (30 days or more) at some point in the past year. It is possible to score these

255 entrepreneurs using psychometrics-based models that were created using data from other countries. The sample from Argentina is too small to create a model based on the Argentinean data. But applying a model made from another country's data is a stronger test, because it is completely out of sample and out of time, so you have even greater confidence that the statistical power will continue when implemented. For small samples, a model based on the country's own data is likely to be over-fit, meaning that it may seem to have high predictive power in-sample, but once you implement it, that power would fall. If a scorecard built on international data, with no customization to Argentina, helps predict likelihood of default within Argentina, then underlying relationships are very strong, so strong that they apply even when measured and modeled on a completely different country. This means that once you have a sample size large enough to build a customized model with that country's own data without it being over-fit, it would be even more powerful

The present study first applies a model made using the data from the Peruvian microfinance institution. The model's predictive power is measured by the area under the "receiver operating characteristic curve" (AUC), a common metric in credit scoring. The other common metric in industry is the Gini coefficient, which is simply two times AUC minus one. For more details on how to calculate an AUC or a Gini, see Annex 2. The present model achieved an AUC of .5257, or a Gini coefficient of 5.1 percent. This result is low, indicating that although it relates to a country from the same region, either the differences in business sizes or cultural differences between Peru and Argentina make that model nontransferable.

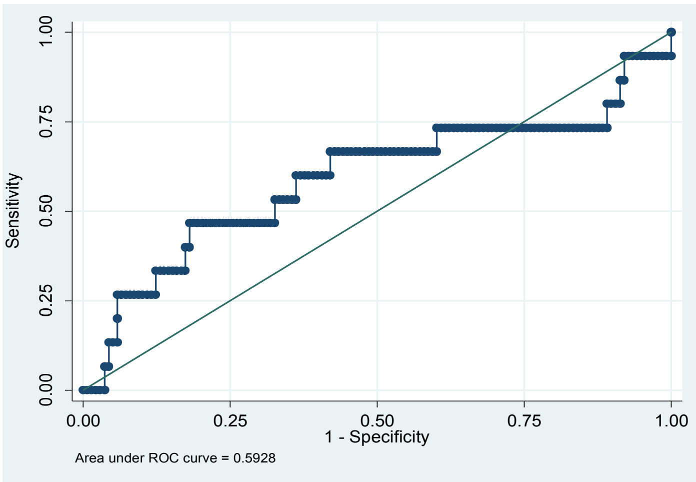
Figure 11. ROC Curve, Peru-based Model



Source: Authors' calculations based on pilot implementation data.

Applying a model based on the Kenyan data, taken from larger-scale entrepreneurs than the Peruvian data, stronger results are obtained (see Figure 12): an AUC of .6 and a Gini coefficient of 18.6 percent. Though still low, as a Gini coefficient approaches 20 percent, it is considered a relevant addition to the decision-making process, particularly when the Gini can be complemented with other information used in traditional application scoring.

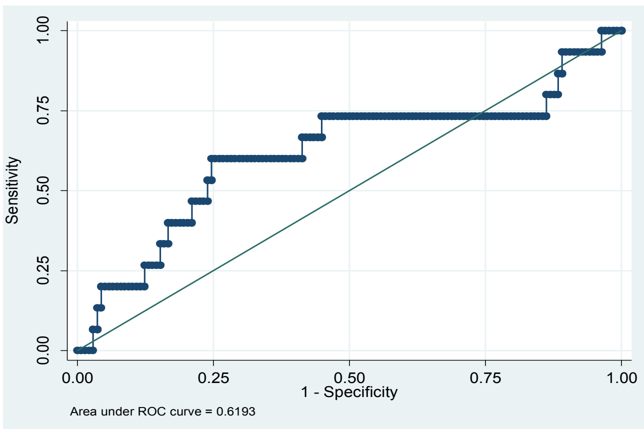
Figure 12. ROC Curve, Kenya-based Model



Source: Authors' calculations based on pilot implementation data.

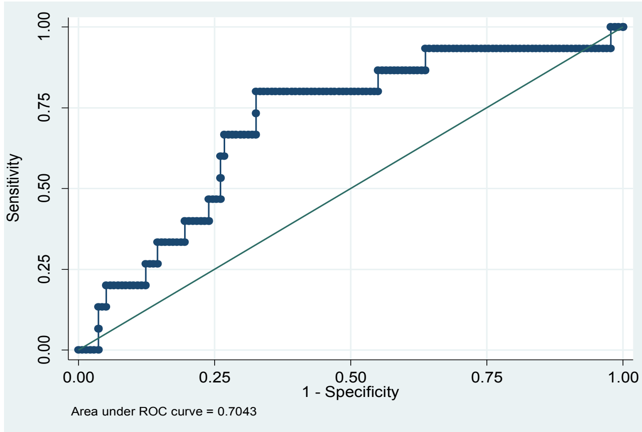
Figures 13 and 14 illustrate the results for scores based on data from Nigeria (AUC.62, Gini 24 percent) and South Africa (AUC.7, Gini 40 percent).

Figure 13. ROC Curve, Nigeria-based Model



Source: Authors' calculations based on pilot implementation data.

Figure 14. ROC Curve, South Africa-based Model



Source: Authors' calculations based on pilot implementation data.

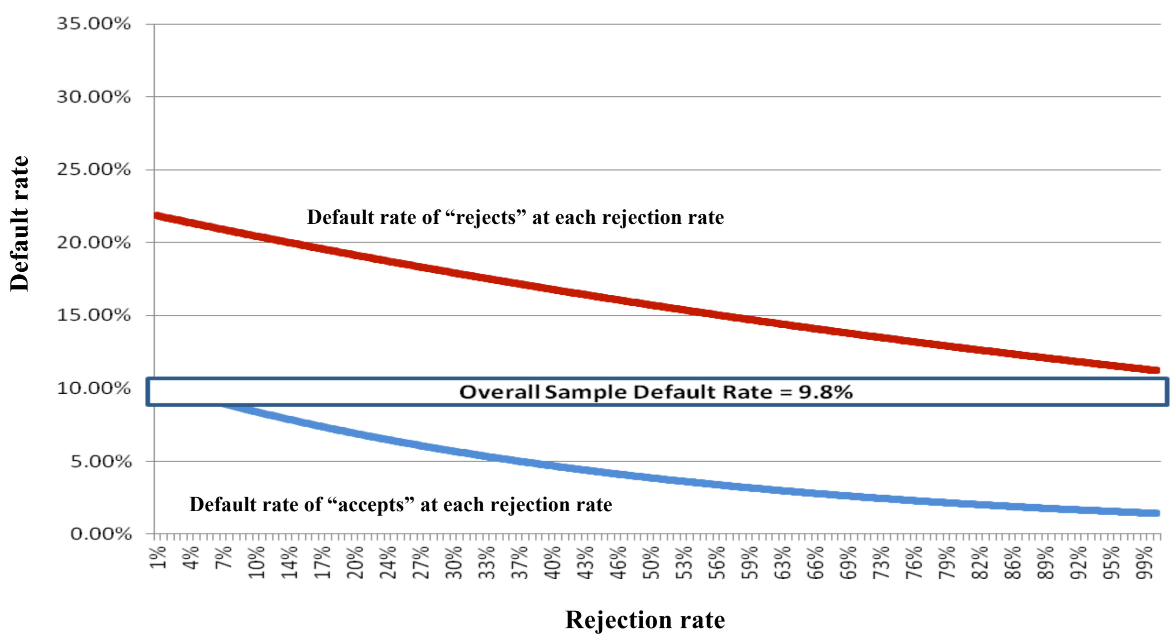
These are rather powerful results, particularly considering how statistically robust this test is: taking a model based on psychometric application responses and repayment data only from either South Africa or Nigeria, and applying that model to a completely different bank, economy, country, culture, and region with no adaptation. Despite this high statistical bar, the resulting scores surpass it and powerfully distinguish risk.

Figure 15 helps interpret what a model with a Gini of 40 percent could provide to Argentina's Banco Ciudad. The sample of clients tested had an overall default rate of 9.8 percent. It is not possible to predict the reduction in default rate, using this model, because the model gives a continuous score, not a "yes" or "no" decision. To make such a decision, one must select the cut-off score, that is, a score below which clients are not accepted. Our model gives an infinite number of potential default rates, including the current 9.8 percent (if everyone in the sample is accepted) to 0 percent (if everyone in the sample is rejected). The important feature of a good model is that it gives a lower default rate among accepted clients than for rejected clients.

This tradeoff is illustrated below. The x-axis shows potential rejection rates of the model, from 0 percent to 100 percent. The y-axis gives the default rate. At any particular rejection rate, default rates are shown for those who would be rejected (i.e., have an EFL score below the cutoff) and for those who would be accepted (i.e., have an EFL score above the cutoff). If a model had no predictive power, then these two curves would be flat, equal to 9.8 percent: the model would reject good clients and defaulters in equal proportion. Instead, what can be observed is that the model divides the sample into groups with strongly different repayment performance at all rejection rates.

For example, if the cutoff were set to reject the bottom 10 percent of scorers on the EFL score, it would eliminate a group of applicants with a 20 percent bad rate and leave a pool of accepted applicants with an 8.5 percent bad rate. Choosing a higher cutoff that rejects the bottom third of EFL scorers would leave a default rate of 19.2 percent among "rejects" and 4.9 percent among "accepts." Rejecting the bottom two thirds would split the sample into rejects with a 13.9 percent default rate and accepts with only a 1.9 percent default rate. The optimal cutoff depends on the cost to keep clients who enter into arrears versus the cost of rejecting good applicants; the important conclusion is the strong discriminatory power of the psychometric-based credit score, even when it is developed out of time and out of sample—conditions that impose an even stricter test. Although these results show only the additional power of the model in defining risk among approved clients, the model's key value is not necessarily the ability to lower risk on approved clients, but to help identify low-risk applicants who would otherwise be rejected by traditional criteria because they lack credit history and collateral. Furthermore, unlike many other tools used in current approval processes, this model uses information that can be provided by any applicant.

Figure 15. Default and Rejection Rates for Different Cut-Off Scores



Source: Authors' calculations based on pilot implementation data.
Note: curves are smoothed.

7. Conclusions and Further Directions

The results of this pilot study show that, unsurprisingly, citizens of different countries clearly have personality patterns and psychological characteristics that differ, based on divergent cultures, languages, and values. Yet despite this diversity, dimensions that are related to entrepreneurial performance and credit risk are common across countries. A psychometrics-based scorecard built on data from countries as different as South Africa and Nigeria still can provide significant discriminatory power within Argentina. Moreover, because these tests were applied on existing clients, the scorecard has discriminating power over and above traditional screening tools that use some of the subjective assessments and financial ratios described above.

Therefore, psychometric-based scorecards would allow BCBA to lower its default rate on existing SME clients and allow a more complete risk analysis. Even more exciting, both from a development perspective and from a private business opportunity viewpoint, is the potential for such scorecards to help open up new segments of SMEs that are currently more difficult to judge with existing criteria. Along with other policies to increase the amount of information available to verify performance and assess default risk for SMEs in Argentina, psychometric-based screening tools could increase financing for the 40 percent of Argentine SMEs for which the major constraint to business growth is access to finance. This finding represents both a potential boost to Argentina's economic output, as well as a ripe profit opportunity for banks such as BCBA.

8. References

BCRA (Banco Central de la República Argentina). 2010a. Financial Stability Report: Second Half 2010. Argentina: Economic Research Deputy General Management, Central Bank of Argentina.

———. 2010b. Prudential Regulations. Argentina: Central Bank of Argentina.

Baron, R. 2006. "Opportunity Recognition as Pattern Recognition: How Entrepreneurs 'Connect the Dots' to Identify New Business Opportunities." *Academy of Management Perspectives* 20: 104–19.

Barrick, M., and M. Mount. 1991. "The Big Five Personality Dimensions and Job Performance: A Meta–Analysis." *Personnel Psychology* 44: 1–26.

Baum, R., M. Frese, and R. Baron. 2006. *The Psychology of Entrepreneurship*. London, UK: Psychology Press.

Bernardin, H., and D. Cooke. 1993. "Validity of an Honesty Test in Predicting Theft among Convenience Store Employees." *The Academy of Management Journal* 36(5): 1097–1108.

Bleger, L. 2011. "Financial Regulations in Argentina, Brazil, Chile, Colombia, Mexico and Peru." Mimeographed document.

Castro, L. 2010. "Instituciones y negocios: desafíos institucionales para la Competitividad Argentina. Mejora del clima de negocios." Washington, DC: Inter-American Development Bank.

Centro de Estudios Monetarios Latinoamericanos, Banco Mundial, and First Initiative. 2008. *Sistemas de reporte de préstamos bancarios y créditos en Argentina.* Iniciativa de Sistemas de Reporte de Préstamos Bancarios y Créditos del Hemisferio Occidental. México: Centro de Estudios Monetarios Latinoamericanos (1st ed. in Spanish, 2011).

Chell, E. 2008. *The Entrepreneurial Personality: a Social Construction*. London, UK: Routledge.

Ciavarella, M., et al. 2004. "The Big Five and Venture Survival: Is There a Linkage?" *Journal of Business Venturing* 19: 465–83.

Costa, P. and R. McCrae. 1994. The stability of personality: observations and evaluations. American Psychological Society, 3(6), 173–75.

EIU (Economist Intelligence Unit) 2010. *Country Finance Report: Argentina*. 2010. *Country Finance Report: Argentina*. The Economist Intelligence Unit Limited.

Girault, M. 2007. "Modelos de credit scoring: ¿Qué, cómo, cuándo, y para qué?" Buenos Aires, Argentina: Gerencia de Investigación y Planificación Normativa, Subgerencia General de Normas, BCRA.

Hogan, J., P. Barrett, and R. Hogan. 2007. "Personality Measurement, Faking, and Employment Selection." *Journal of Applied Psychology* 92(5): 1270–1285.

Hunter, J. 1986. "Cognitive Ability, Cognitive Aptitudes, Job Knowledge, and Job Performance." *Journal of Vocational Behavior* 29(3), 340–362.

Lazear, E. 2005. "Entrepreneurship." *Journal of Labor Economics* 23(4), 649–680.

McClelland, D. 1961. *The Achieving Society*. Princeton, NJ: Van Nostrand.

Musacchio, A. 2011. "Banco Ciudad (B): Transformation at Work." *Harvard Business School Supplement* 712-030.

Neisser, U., Boodoo, G.,. Bouchard, T., Boykin, A., Brody, N., Ceci, S., Halpern, D., Loehlin, J., Perloff, R., Sternberg. R., Urbina, S. 1996. "Intelligence: Knowns and Unknowns." *American Psychologist* 51(2): 77–101.

OPyME (Fundación Observatorio PyME). 2009. "Inversión y Financiamiento, datos comparativos." Informes especiales. Buenos Aires, Argentina: OPYME.

Patterson, A., et al. 2012. A 30-country multilevel analysis of managerial skillsets important for success. Paper to be presented at the 27th annual meeting of the Society for Industrial and Organizational Psychology, San Diego, California.

Ree, M., J. Earles, and M. Teachout. 1994. "Predicting Job Performance: Not Much More than g." *Journal of Applied Psychology* 79(4): 518–24.

Schmidt, F. and J. Hunter. 1998. "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings." *Psychological Bulletin* 124(2): 262–74.

Seibert, S. and H. Zhao. 2006. "The Big Five Personality Dimensions and Entrepreneurial Status: A Meta–Analytical Review." *Journal of Applied Psychology* 91(2): 259–71.

Snow, R. and E. Yalow. 1982. *Handbook of Human Intelligence*. London, UK: Cambridge University Press.

Annex 1. Recruitment Message.

Figure A.1. Message Utilized by BCBA's Credit Officials[6]

Evaluación de impacto de técnicas psicométricas

CIPPEC políticas públicas public policies

Implementación

Mensaje

Buenos días/Buenas tardes,

Me comunico con Ud. para informarle que el Banco Ciudad está colaborando con una investigación de **una nueva herramienta que permitirá mejorar los productos crediticios ofrecidos a las PyMEs**. Se trata de un novedoso instrumento desarrollado por la Universidad de **Harvard** que ayudará al Banco a evaluar más adecuadamente las **necesidades de nuestros clientes**.

Para ello, recurrimos a la colaboración voluntaria de la clientela, y nos gustaría que usted participe. Solo deberá responder una **encuesta autoadministrada** que se ejecuta a través de una computadora portátil. No llevará mucho tiempo, y podremos efectuarla en el **banco o en su propia empresa**.

Por otro lado, garantizamos que **los resultados de la encuesta no influirán de manera alguna en su relación con el banco** . En agradecimiento por el tiempo prestado, participará del **sorteo de 3 laptops** que se hará entre los participantes de este estudio.

El test será llevado a cabo por **CIPPEC**, un prestigioso centro independiente de estudios sobre políticas públicas. De ser de su interés participar, CIPPEC se comunicará con Ud. a la brevedad para coordinar fecha, hora y lugar que a usted le convenga para realizar el test.

Muchas gracias por su tiempo y predisposición.

16

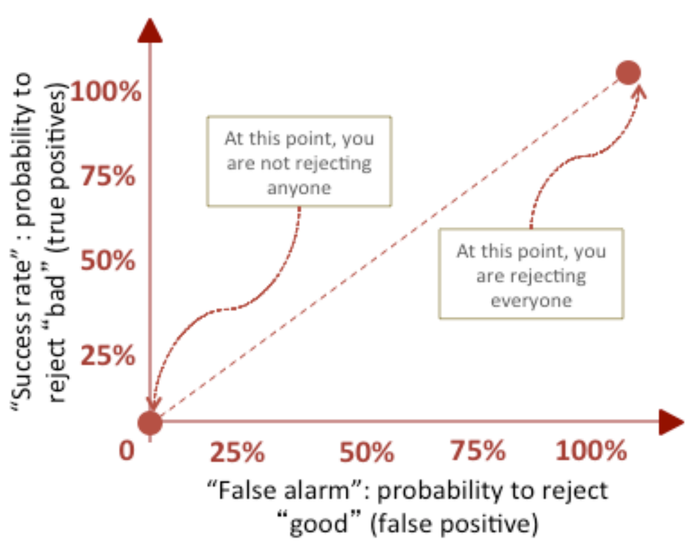
[6] Translation: Good morning/Good evening, I am calling you to inform you that the BCBA is collaborating in a project aimed at evaluating a new tool that will contribute to improve our financial products for SMEs. It is an innovative tool developed by Harvard University that will help the BCBA to evaluate more adequately the needs of our SME clients. Therefore, we would like to kindly ask for your cooperation. You just need to complete a brief auto-administered survey that will be provided in a laptop. It will not take much of your time and can be completed either in your office or in one of our branches. We guarantee that your answers will not affect your current credit standing with the Bank. In appreciation for your participation, you will enter into a lottery for three laptops. The test is being carried out by CIPPEC, a prestigious independent public policy think tank. If you are interested in participating, CIPPEC will contact you to schedule an appointment.

Annex 2. Graphing and Interpreting an AUC and a Gini Coefficient

A common set of metrics is used to assess credit scores. It is important to remember that any credit score does not trigger a decision to accept or reject an applicant; assessment is a continuous relative measure of risk, and lenders can make an accept/reject decision based on any score cutoff point they select. Metrics to assess the predictive power of credit scores therefore evaluate the score's ability to sort applicants by credit risk rather than by imposing any single cutoff. If a score is closely related to default, then participants with a low score should be much more likely to default than those with a high score. Credit scores with little value for directing lending do not separate high-risk applicants from low-risk applicants, and both categories tend to be evenly distributed across the score's spectrum.

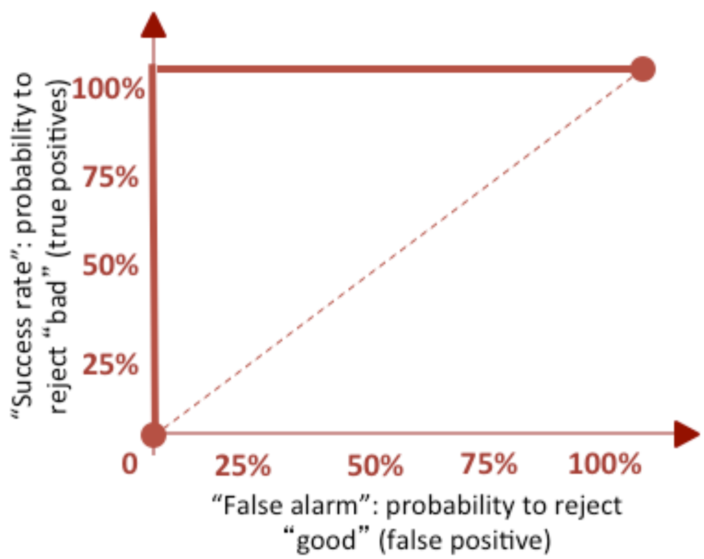
This ability of a model to sort applicants based on level of default risk is typically illustrated by a receiver-operating characteristic curve, or ROC curve (see Figure A.2). This curve plots on the x-axis the percentage of "goods" (non-defaulters) below any particular score level, while the y-axis shows the percentage of "bads" (defaulters) below that score. Any credit scoring model represents a curve on this graph, with each point on the curve showing the impact of a potential cutoff score.

Figure A.2. Building an ROC curve



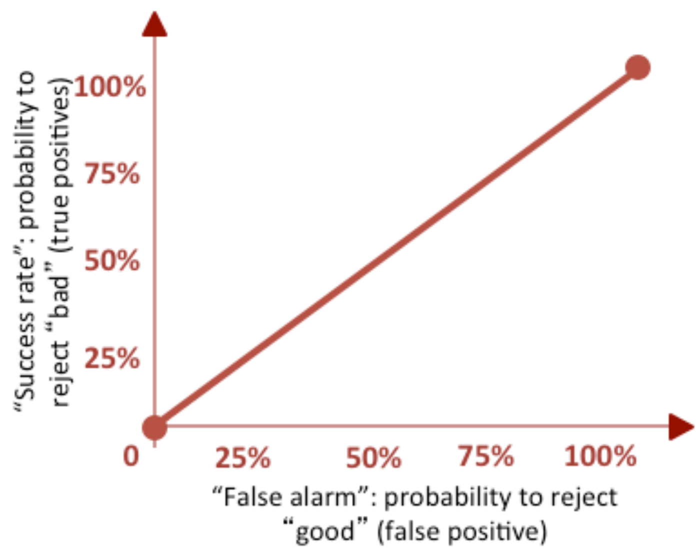
As illustrated in Figure A.3, a perfectly predictive credit model would assign the lowest score to all defaulting clients, and therefore when rejecting applicants with the lowest score, only the defaulters would be rejected, producing a move up the y-axis while the x-axis remains at 0 percent. Only after raising the rejection score cutoff to the point that 100 percent of the "bads" were rejected (top left corner), would raising the rejection score cutoff start rejecting the "goods," moving from the top left corner horizontally along the y-axis until it reaches the maximum score so that all applicants would be rejected (top right). So the ROC curve for a perfectly predictive credit score would appear at the top left side of the figure.

Figure A.3. ROC Curve of a Perfect Model



Conversely, the ROC curve for a credit score containing no predictive power would fail to distinguish "bads" from "goods" at any level, but would be like flipping a coin. This means good and bad credit risks would be evenly distributed across all scores. So in Figure A.4, beginning from the lowest score and increasing the rejection score cutoff would lead to a rejection of both "goods" and "bads" in equal proportion. In other words, the ROC curve would be a straight diagonal line.

Figure A.4. ROC Curve of a Random Model



Most scoring models fall between these two extremes. A better credit scoring model would resemble the 90 degree angle than the diagonal line above; when the line bows up and to the left, the model gives lower scores to a greater proportion of "bads" and higher scores to a greater proportion of "goods."

In summary, the credit scoring industry typically describes a model's power using the Area Under the ROC curve (AUROC or often shortened to AUC). The higher an AUROC, the better the model. The perfectly predictive model above (Figure A.3) has an AUROC of 1, while the useless model (Figure A.4) has an AUROC of 0.5. A common transformation of the AUROC is a Gini coefficient, which is simply (2*AUROC)-1. While comparing AUROC or Gini coefficients across different samples is not recommended because they are sensitive to absolute level of default, these metrics are useful for comparing the performance of different models, using the same data.