

Does Psychometric Testing in Microfinance Actually Work?—The Case of Sogesol

Rocheny Sifrain

Independent Researcher, Port-au-Prince, Haiti

Email: rsifrain@gmail.com

How to cite this paper: Sifrain, R. (2020). Does Psychometric Testing in Microfinance Actually Work?—The Case of Sogesol. *Journal of Financial Risk Management*, 9, 278-313.
<https://doi.org/10.4236/jfrm.2020.93016>

Received: August 10, 2020

Accepted: September 26, 2020

Published: September 29, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Psychometric testing is claimed to be a powerful innovation in credit scoring. Pioneered by the Entrepreneurial Financial Lab (EFL), this technique would enhance credit decisions by screening out high-risk applicants. This paper aims to evaluate the predictive power of the EFL's psychometric credit scoring model in microfinance through evidence from Sogesol, a Haitian microfinance institution. This evaluation has been conducted at two different levels: 1) A sample of clients has been selected from Sogesol's database to carry out a back test of the EFL tool, using performance metrics such as the Kolmogorov-Smirnov (K-S) statistic, the area under the ROC curve (AUC) in comparison with the existing socio-demographic model in use at Sogesol; 2) We conduct an analysis of causality between the quality of the portfolio and the credit decisions made based on the EFL tool and/or the traditional credit scoring model through the estimation of a linear regression model. The results show that the psychometric credit scoring model would present low predictive power in terms of K-S and AUC. However, the EFL tool would outperform the socio-demographic credit scoring model in use at Sogesol. The study further indicates that there would not be any statistically significant relationship between the risk level and the decision of granting a loan or not. The paper concludes that psychometric testing in its original format would not be efficient in the context of Sogesol's microcredit operations. Thus, the paper develops a new credit scoring model along traditional socio-economic and behavioral lines, using logistic regression. This new model presents a better discriminatory power than the EFL tool, regarding K-S and AUC. In addition, it is well-calibrated, considering the results of Hosmer-Lemeshow (HL) test and the Brier score. If properly maintained and integrated into the client selection process, this new model could significantly improve credit risk management practices at Sogesol.

Keywords

Credit Risk, Credit Scoring, Psychometric Testing, Microfinance

1. Introduction

Risk management constitutes one of the core functions of banks and other types of financial institutions, because risk is inherent in all of their activities. Among the different types of risks they are facing, credit risk may be considered the most important source and the biggest exposure. That is why credit risk management plays more and more a critical role, in that financial institutions have to constantly calibrate the tradeoff between risk and return. After the global financial crisis of 2008-2009 that started in United States with subprime housing loans, a particular focus has been put on credit risk management. It is commonly admitted that one of the causes of the financial crisis was a lack of rigorous credit risk assessment. To address this issue, the Basel Committee and local regulatory authorities made it mandatory for banks and other financial institutions to be equipped with tools that will provide better visibility of credit risk. Credit scoring is considered to be one of those tools. Statistical credit scoring models based on socio-demographic variables are developed to estimate the probability of default of borrowers. This traditional scoring model is largely used by microfinance institutions since their clients are considered very vulnerable, taking into account their lack of collateral that prevents them from accessing conventional bank credit.

The big challenge for those lenders is to find a tool that can really help assess the risk, while increasing the financial inclusion which is one of the Sustainable Development Goals (United Nations, 2015). Hence the importance of the integration of alternative data in credit risks assessment. This can be understood as the motivation of developing models with the incorporation of psychometric variables. Psychometric testing seeks for ways to assess borrower's willingness to repay when she/he has no credit history with the lender, no credit history in a bureau, which is common in microfinance.

The psychometric scoring has been implemented by financial institutions in Africa, Asia, Latin America, in the Caribbean. In psychometric scoring, applicants for loans have to answer a list of questions measuring their intelligence, business skills, personality, ethics, and character as a way to evaluate their willingness to repay their loans. The Inter-American Development Bank (IDB) argued in one of its articles that the implementation of psychometric testing by banks and other financial institutions can reduce defaults by a 20 to 45 percent and a 15 to 30 percent increase in profits, with operational costs of the lending process at less than 40 percent of the cost of traditional evaluation and due diligence (Inter-American Development Bank, 2013).

This argument of the IDB constitutes one of the main motivations behind this paper, in that we are interested to know whether that assumption is confirmed in practice. Hence the *title*: “Does psychometric testing in microfinance actually work?—The case of Sogesol”.

1.1. Purpose and Objective

The objective of this paper is to evaluate the predictive power of the psychome-

tric scoring model implemented at Sogesol (Société Générale de Solidarité) in 2012, the largest microfinance institution in Haiti in terms of outstanding portfolio (more than US\$ 40 million for almost 35,000 loans, as of September 2018). As of April 2016, Sogesol had tested 5517 applicants. The psychometric tool was developed and implemented in a microfinance institution where a socio-demographic scoring model had been in place since 2006. That is why the intention of Sogesol was to develop a hybrid model combining the psychometric factors with the socio-demographic variables in order to enhance credit decisions. The paper determines the effectiveness of the psychometric tool by way of analyzing statistical metrics as well as the loan repayment performance.

In order to complete the validation process, the research also reviews the calibration of the psychometric model. The model may be good in terms of discriminatory power, but not sufficiently calibrated. Assessing the calibration requires that clients must be grouped by class of scores. The rule is that a model with a correct calibration would indicate similar default rates for clients belonging to the same class of scores.

Finally, based on the results of the psychometric model review, our purpose is also to look for opportunities to re-estimate/recalibrate the socio-demographic model, so as to enhance financial inclusion at Sogesol with improved visibility of credit risk in the micro-entrepreneurial market.

1.2. Problem Statement

In developing countries where the socio-economic conditions are difficult, many low-income individuals seek to create a livelihood by running their own entrepreneurial activity. These activities are of different sizes: micro, small and medium. Hence the concept of Micro, Small and Medium Enterprise (MSME) emerged. They constitute real sources of revenues for their owners. In general, micro-enterprises are not formally structured. They do not have financial statements or accounting systems that provide information on their financial performance. The lack of documented financial performance creates a problem of information asymmetry for lenders. Besides, they are very vulnerable to any social or economic shocks. The owners of micro-activities typically also lack collateral to offer if they want to obtain loans for their business. Consequently, traditional financial institutions are not interested in financing their business, being considered too risky. To meet their funding requirements, entrepreneurs are obliged to borrow from informal moneylenders at exorbitant interest rates that may deplete their working capital and drive them deeper into debt.

This gap in the credit markets has created an opportunity for microfinance as an alternative financing technology. Microfinance is specifically designed to meet the needs of MSMEs by providing tailored working capital funding solutions. One should point out that microfinance is also considered an important economic development tool, created to help eradicate poverty worldwide. Specialized microfinance organizations then emerged that focused on serving un-

banked entrepreneurs. However, the problem of information asymmetry remains. From the 2000s, microfinance institutions have started using the traditional credit scoring to address this problem and enhance the risk assessments made by loan officers. More recently, an alternative credit scoring based on psychometric features has been introduced as a credit risk tool for new borrowers lacking any form of verifiable credit history.

In 2006, the Entrepreneurial Financial Lab (EFL) started developing a credit scoring model with psychometric factors at Harvard University. The objective is to address the information asymmetry when considering the creditworthiness of the micro-entrepreneurs, mainly in the context of absence of credit bureaus and where applicants have no credit history. The psychometric tool should enable financial inclusion by selectively giving access to loans to previously unbankable clients without significantly increasing the portfolio credit risk. From the 2010s, microfinance institutions have started deploying psychometric scoring in their credit decisions. Sogesol was the first microfinance organization in the Caribbean region to implement a psychometric model.

This alternative tool is presented as a method that screens out applicants with high-risk. In other words, psychometric testing promises to be effective in the prediction of borrowers' repayment. This paper tests the two following hypotheses, using the case of Sogesol's psychometric credit scoring model:

Hypothesis 1: The psychometric scoring model offers low predictive power that cannot discriminate between good borrowers and bad borrowers. If this hypothesis is true, the model may fail to predict the probability of default of borrowers.

Hypothesis 2: There is no significant impact of the psychometric model on Sogesol's portfolio quality. If this hypothesis is true, the psychometric tool may not mitigate credit risk in Sogesol's portfolio.

In order to verify these hypotheses, the paper uses the data of Sogesol on the psychometric scoring and the borrower's repayment performance. A sample of clients is selected, using different definitions of good and bad clients, including the one used in the development of the customized psychometric model. The evaluation of the psychometric tool is realized using statistical methods. Furthermore, since the psychometric tool was deployed in a socio-demographic scoring environment, a comparative analysis of the two different scoring models is carried out, in order to obtain relevant insights. In the end, the paper proposes the re-estimation and the recalibration of the socio-demographic model, using a logistic regression model. R software and Excel are used to conduct the analysis. More details are provided in the data and research methodology section.

2. Analytical Framework of Psychometric Testing

2.1. Definition of Credit Risk

In respect of offering credit, there is a common element to take into consideration: the need to study the creditworthiness of borrower or counterparty or the

need for credit risk analysis. It means checking whether the prospective borrower is worthy to receive credit (Joseph, 2013). A borrower who is not creditworthy has a high propensity to default on credit.

Credit risk is defined as the potential that a contractual party will fail to meet its obligations in accordance with the agreed terms. Credit risk is also called default risk, performance risk or counterparty risk (Brown and Moles, 2014).

2.2. Definition of Credit Scoring

Credit scoring is referred to as the use of statistical models to determine the likelihood that a prospective borrower will default on a loan. Credit scoring models are then largely used to assess business, consumer loans, and so on (Abdou and Pointon, 2011). In addition, credit scoring is defined as the set of decision models and their underlying techniques that help lenders grant loans. These techniques decide who will get a loan, how much prospective borrowers should obtain, and what operational strategies will enhance the profitability of the borrowers to lenders (Abdou and Pointon, 2011).

To develop a credit scoring models, many variables are used. These variables are socio-demographic, financial. Credit bureau data are added to enhance the decision making of extending, mainly for applicants found in the grey area of internal scores formula, taking into account the cut-off score defined. These data are called traditional data. With the evolution of technologies and statistics, other types of data are used to predict borrower's repayment behavior. These types of data are known as alternative data. Psychometric attributes are one type of alternative data.

2.3. Psychometric Testing in Credit Risk Assessment

Psychometric assessment is widely used to measure personality traits, knowledge, skills and attitudes. The possibility to screen many people at low cost is seen as one of the benefits of the psychometric assessment. That is why employers use it to select the best-talented employees for their business. It is also demonstrated that the personality dimensions are correlated to the entrepreneurship status.

The success of psychometric assessment in predicting job performance has encouraged the transfer of that method to other areas, such as small and medium enterprise credit and microfinance, where screening applicants is very costly and time consuming. Not only the psychometric tool can help reduce the costs, but also it offers a solution to the asymmetric information while assessing the credit risk of the applicants. The psychometrician made the assumption that there is a personality trait or a set of traits characterizing low-risk versus high-risk loan applicants. The purpose is then to identify those traits and build a measure that has suitable psychometric properties and predictive value. The questions identified by the psychometrician must systematically be tested on real-world loan applicants, and their predictive validity confirmed by best prac-

tices of credit scoring (Arráiz et al., 2015).

2.4. EFL and Psychometric Testing

The use of psychometrics in screening credit applicants and in predicting their repayment behavior was originated by the Entrepreneurial Finance Lab (EFL), by experimenting with psychometric credit scores at Harvard University in 2006. Initially, the objective was to address information asymmetry at the Harvard Center for International Development Research. Afterward, EFL extended its business all over the world, partnering with leading financial institutions, and winning global awards such as the G-20 SME Finance Challenge recognizing EFL as one of the most innovative solutions for SME finance in the world and the African Business Award for Innovation.

EFL offers psychometric credit scoring assessments, taking into account the local culture. The evaluation can be completed on a tablet or a laptop without accessing to the internet. The purpose is to assess qualitative measures such as personal initiative, situational judgment, creativity, and business acumen. The psychometric test is designed to integrate into a financial institution's existing underwriting tools and methods. EFL began by quantifying the characteristics of borrowers who had defaulted on a past loan versus those who had not, and of borrowers who owned small businesses with high versus low profits. EFL grouped these characteristics into three categories: personality, intelligence, and integrity (Arráiz et al., 2018). EFL originally worked with a personality assessment based on the five personality dimensions, also known as the "Big Five" model (Costa Jr. and MacCrae, 1992), an intelligence assessment based on digit span recall tests (a component of the Wechsler adult intelligence scale), the raven's progressive matrices tests (Spearman, 1946), and integrity assessment adapted from Bernardin and Cooke (1993).

The assumption formulated by the EFL researchers was that these assessments would enable them to identify the two core determinants of an entrepreneur's intrinsic risk: the ability to repay a loan and the willingness to do so. Entrepreneurial traits, measured through personality and intelligence tests, define entrepreneur's ability to generate cash flows in the future, cash flows that can, in turn, serve to repay any loan contracted. Honesty and integrity traits, measured through the integrity test, explain the entrepreneur's willingness to pay, independently of the ability to do so. EFL identified questions that could potentially predict credit risk and tested a first prototype of their psychometric tool. Afterward, EFL developed a commercial application based on the responses to their tool and subsequent default behavior. The commercial application is based on the same quantitative methods applied to generate traditional credit scores, comprised of questions developed internally and licensed by third parties relating to individual attitudes, beliefs, integrity, and performance, in addition to traditional questions and the collection of metadata (indicating the interaction of the applicant with the tool).

The EFL application produces a 3-digit score that classifies the relative credit risk of the person who took the test. Financial institutions can apply this score in different ways: for approvals, or modifying the price, size or other margins of loan.

3. Microfinance, Sogesol and Psychometric Testing

3.1. Haitian Microfinance Overview

3.1.1. Definition of Microfinance and the Microfinance Institutions

The Haitian Central Bank (Banque de la République d'Haïti or BRH) defines the microfinance as the sector that extends small-scale credit to low-income people, allowing them to create and manage their microenterprises. Its goal is to expand financial access to low-income people or to those previously excluded from the formal financial system. It allows them to permanently access quality and affordable financial services in order to finance income-generating activities, to save, to accumulate assets, stabilize their consumer spending and to protect themselves against risks (BRH, 2018).

According to the Central Bank, there are three groups of microfinance institutions (MFI) in Haiti:

- Mutualist microfinance institutions or cooperatives: They constitute a group of people, part of a non-profit organization and founded upon the principles of cooperation, of solidarity and mutual support primarily with the objective to collect savings of its members and/or credit granting. They extend loans to members and to individuals as well. They are regulated by the 2002 law on savings and credit cooperative.
- Solidarity credit unions: A solidarity credit union is defined as a group of people with strong ties to each other (Socio professional origin, place of residence, facility, friendship, etc.) that decide to create a fund fed by their contributions, in order to reach a clearly defined goal: the granting of credit to the members of the group on a rotating basis. Unlike the community banks, the solidarity credit unions are independent from the start: operation rules are established by the group itself without the interference of any MFI, even though this one may be an alternative source of funds to supplement the inadequacy of internal resources and provide a technical assistance as well.
- Non-cooperative microfinance institutions: Those MFIs differ from those organized in a mutualist fashion or the cooperatives. They grant credit from the borrowed funds of the banking system or an international financial organization or donations from NGO. Those MFIs can be NGOs, associations, bank subsidiaries and anonymous societies. ACME (Association pour la Coopération avec la Microentreprise), MCC (Microcrédit Capital), MCN (Microcrédit National) and Sogesol (Société Générale de Solidarité) are examples of MFIs that are not cooperatives.

3.1.2. Industry Environment

Haiti is the poorest country in the Western hemisphere, with Gross Domestic Product (GDP) per capita of US\$ 870 in 2018. More than 6 million Haitians (over fifty percent) live below the poverty line with less than US\$ 2.41 per day and more than 2.5 million fall below the extreme poverty line i.e. US\$1.23 per day (World Bank, 2019).

A slight growth of GDP was observed in one year, passing from 1.2% in 2017 to 1.5% in 2018. Such a result was associated with an expanding budget deficit of 4.35% of GDP in 2018 against 1.9% in 2017. This deficit is progressively being financed by the Central Bank. Consequently, the national currency (the Gourde) continues to depreciate, driving double-digit inflation (about 15%) and further penalizing the poorest households. Those macroeconomic conditions accompanied by weak tax revenues, have reduced the room for Government to increase in improving the budgetary allocation for social issues.

Haiti has struggled with several periods of instability caused by demonstrations, strikes and civil unrest at the national level. Since July 2018, the situation has progressively deteriorated with protests on a regular basis in the streets of the Haitian Capital, Port au Prince. Moreover, Haiti is greatly exposed to natural disasters (floods, hurricanes, earthquakes). According to the World Bank, more than 96% of the population are vulnerable to these natural disasters. After the devastating earthquake of 2010, hurricane Matthew in 2016 affected the country and caused losses and property damage estimated at 32% of GDP.

3.1.3. Sector Performance

According to the census results of the microfinance industry in Haiti, conducted in 2018 on 67 institutions, a total of 274 service points covering ten departments of Haiti are distributed as follows: 57 in Port au Prince, the Haitian capital, 117 in urban areas and 100 in rural areas. 35% of those service points belong to savings and credit unions, 34% to limited companies, 28% to bank subsidiaries and 3% to NGOs and associations/foundations (USAID, 2018).

In 2017, the gross portfolio of the microfinance sector was estimated at more than US\$ 207 million against more than US\$ 152 million in 2016. In 2017, the credit unions occupied about 42% of the whole gross portfolio, bank subsidiaries 34%, limited companies 22%, and others 0.59%. About 40% of the gross portfolio is allocated to women.

Moreover, the number of borrowers was estimated at 281,263 in 2017 compared to 242,140 in 2016, representing a year-over-year growth of 6%. Women borrowers represent about 41% of the total number of loans. In terms of credit methodology, the loans distribution is as follows: individual loan (63%), joint-liability group (20%), community bank (15%), NGOs and Others (1.36%), and Solidarity credit unions (0.36%).

In spite of the efforts of the microfinance sector in terms of outreach, the portfolio quality remains a big challenge. On average, the portfolio quality of the sector has declined in 2017, with a PAR30 rate of 13.44% against 12.22% in 2016.

3.2. Sogesol Overview

Sogesol (Société Générale de Solidarité) was created as a service company for Sogebank, Haiti's one of the largest commercial banks. Its mission is to promote Haitian entrepreneurship by adapting traditional banking services to the needs of micro and small businesses. Sogesol has known some changes in its shareholding. In the beginning, Sogesol's shareholders were: Sogebank (35%); Accion International (19.5%); ProFund (20.5%) and Individuals (25%). From 2018, Sogesol's shareholders are: Sogebank (51%) and Individuals (49%). Sogesol disbursed its first loan in November 2000, providing initially individual working capital in urban areas.

Since its foundation, Sogesol has significantly grown in spite of tough economic, social and political context of Haiti. Nowadays, Sogesol offers a full range of credit products: working capital finance, agricultural loans, consumer credit and housing microfinance. The majority of its clients are micro and small business owners and agricultural producers. Sogesol has 17 branches, of which seven are in metropolitan zones and 10 in rural areas. In addition, Sogesol has 5 other points of services to better serve the customers. Sogesol is then a national network, of which headquarter is located in Port-au-Prince, the Haitian capital.

3.2.1. Sogesol's Outreach Performance

From inception in 2000 to September 2018, Sogesol has served 202,546 customers and disbursed a total amount of US\$ 345,550,631 for 532,754 loans. At the end of the fiscal year 2018, Sogesol's outstanding portfolio was US\$ 44,158,354 against US\$ 35,153,304 in 2017, representing a growth rate of 26% on a year-on-year basis. Conversely, the number of borrowers has known a drop of 7.1%, passing from 36,209 in 2017 to 33,653 in 2018. This performance may be explained by Sogesol's strategy to increase its portfolio outstanding by improving the average amount disbursed.

Moreover, it is important to analyze Sogesol's performance with regard to the competition. To do so, we consider the 3 biggest non-cooperative MFIs of the Haitian microfinance sector. The two following graphs present respectively the evolution of outstanding portfolio and number of borrowers of the competition from 2012 to 2018.

As observed in **Figure 1**: in 2018, Sogesol (US\$ 44,158,354) has the largest outstanding portfolio among the four biggest non-cooperative MFIs, followed by MCN (US\$ 43,400,552). The same rank was observed in 2017, with Sogesol (US\$ 35,153,304) and MCN (US\$ 33,398,684). However, from 2012 to 2016, MCN occupied the first place, followed by Sogesol from 2014 to 2016 and by ACME in the first two years under study. MCC (US\$12,657,972) is the smallest MFI in terms of outstanding portfolio.

Regarding the number of borrowers, at the end of the fiscal year 2018, ACME (35,760), followed by MCN (34,640). Historically, we can observe that ACME and Sogesol have had the greatest number of borrowers, even though MCN dominated the market in 2016 and 2017. The drop observed in Sogesol's number

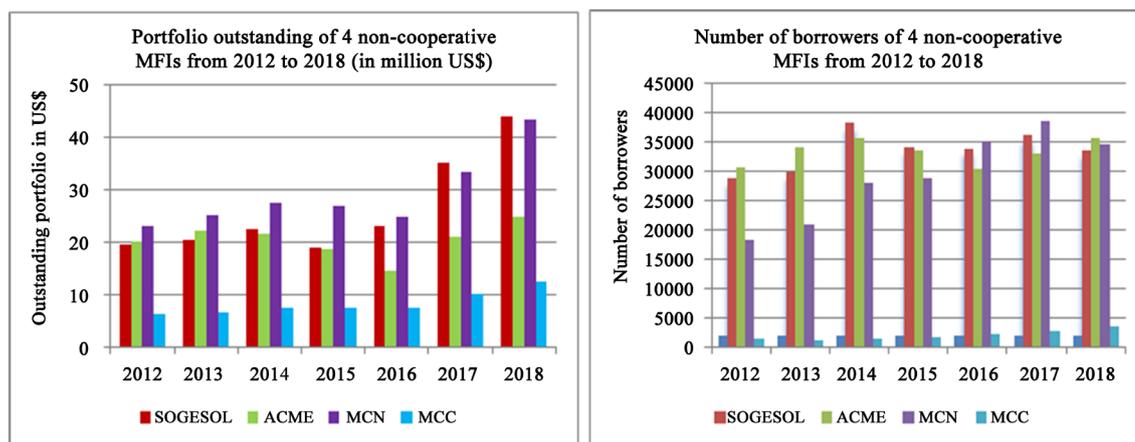


Figure 1. Portfolio outstanding and number of borrowers of 4 non-cooperative MFIs.

of borrowers might be explained by a shift in its commercial strategy, putting more emphasis on the volume of the portfolio than the number of costumers. It is important to mention that MCC (2025) is specialized in Small and Medium Enterprise (SME) loans, which explains its weak number of borrowers, since it is easier to find a microenterprise than an SME. It is also indicated to underline that except ACME that is an association, all of the 3 MFIs are banks affiliates dedicated to provide microfinance services.

3.2.2. Sogesol's Portfolio Quality

At the end of the fiscal year 2018, Sogesol's portfolio quality has deteriorated compared to 2017. The PAR 30 has passed from 6.6% to 7.98% in one year. The result of the competition is not different from the one of Sogesol. The graph below displays the evolution of the portfolio at risk more than 30 days from 2012 to 2018.

As shown in Figure 2, all of the 4 MFIs have known deterioration in 2018. ACME presented the worst performance (8.84% in 2018 against 6.5% in 2017), followed by Sogesol (7.98% in 2018 against 6.6% in 2017). MCC displayed the

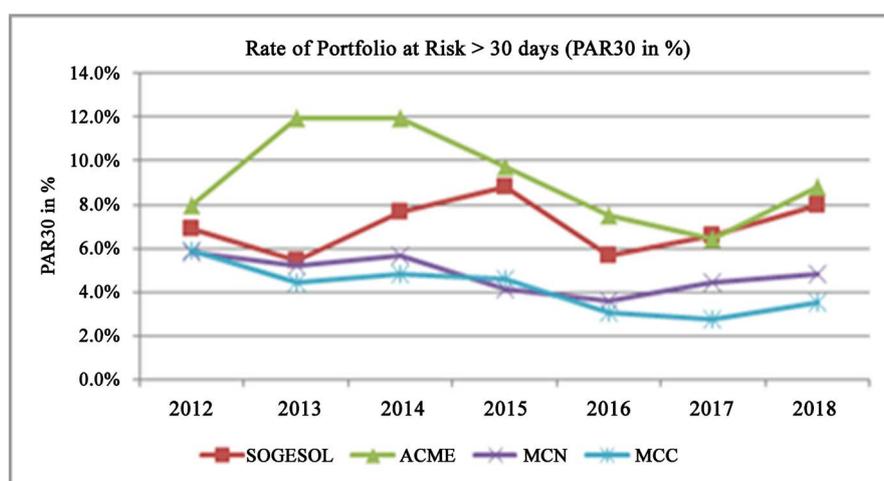


Figure 2. Rate of portfolio at risk > 30 days (PAR 30 in %).

lowest rate of PAR 30 days (3.49% in 2018). The overall performance was impacted by the deterioration of the socio-political environment of Haiti in 2018, mainly after the eruption of major and violent protests in July 2018, as the government announced a price increase of 38 percent to 51 percent for gasoline, diesel and kerosene. Those events truly affected the conditions of the Haitian population. The best rate of PAR30 registered by MCC on the period under study is related to the fact that its portfolio is exclusively comprised of SMEs which are less vulnerable than the microenterprises which are part of the portfolio of the 3 other MFIs.

3.3. Sogesol and Credit Scoring

The use of credit scoring began at Sogesol in 2006, with the technical support of Accion International. The goal of Sogesol was to implement its first credit scoring model while launching a new type of working capital product dedicated to the most vulnerable borrowers. This product is known as a nano loan, meaning for a smaller amount than even a microloan (\leq US\$ 500). Sogesol became then the first Haitian financial institution to incorporate scoring in its loan process.

With the introduction of the sociodemographic scoring model in its operations, Sogesol mainly aimed to improve customer service by accelerating the loan approval process, maximize efficiency of collection activities, and improve portfolio quality.

Based on the performance of its first sociodemographic credit scoring, Sogesol decided to extend the use of credit scoring to microenterprise working capital loans. Sogesol has then developed four credit scoring models: two models to assess new borrowers (1 for nano loans and 1 for microenterprise working capital loans) and two models for repeat borrowers (1 for nano loans and 1 for microenterprise working capital loans). In addition, Sogesol has also initiated in its credit process, the psychometric testing.

3.4. Sogesol and Psychometric Testing

In 2012, Sogesol and EFL teamed up to incorporate psychometric credit scoring into the credit process. Sogesol signed an agreement with EFL for the development of the psychometric credit scoring model. Sogesol was interested to test whether the EFL tool could help it enhance its credit approval process. Since Sogesol had several years of experience in implementing traditional credit scoring models, the objective was to develop a hybrid credit scoring model (socio-demographic/psychometric), in order to increase the predictive power of the existing credit scoring model. MSME's owners who applied for a working capital of US\$ 1000 and more have been screened by the EFL as part of the application process. On average, the EFL application took 61 minutes to complete.

Implementation Phases

The implementation of the psychometric credit scoring model at Sogesol consists of the following phases:

- Data collection to test the EFL global model;
- Adaptation of this model to the local context;
- Development of a semi-calibrated and a calibrated model;
- Testing the psychometric model alone.

The data collection phase lasted about two years and half, from August 2012 to February 2015. In April 2015, EFL developed a semi-calibrated model based on the set of data previously collected. A fully customized model was finally developed by EFL in October 2015. While implementing the psychometric score, clients kept being assessed under the conventional credit scoring model. While waiting for the development of the final calibrated model, Sogesol decided to test the robustness of the EFL tool using the semi-calibrated psychometric model, without informing its loan officers to avoid bias that may arise during the usual credit process.

4. Performance Evaluation of the Psychometric Model

4.1. Data and Research Methodology

4.1.1. Data Sources

The data points used in the evaluation of the psychometric credit scoring model have been collected from three different sources of Sogesol's database, which is the information system used for decision-making, coordinating, control, analysis and visualization of credit information. For the back testing of the calibrated psychometric model, we used a sample of 3671 working capital clients whose loan was disbursed between June 2013 and April 2016. This sample represents the number of tests, which were disbursed and reported properly, out of a population of 5717 tests administered during the period under study. These data include the EFL scores generated by the calibrated model, the Day Past Due (DPD) on Months on Book (MoB), the date disbursement, amount disbursed, loan outstanding after MoB, loan term, and other some socio-demographic variables (Age, gender, marital status, education level), used in the psychometric credit scoring development.

Additionally, a sample of 2014 observations has been selected from data of the traditional scores between February 2014 and December 2017. These data have been used to make the back testing of the traditional credit scoring model, for a performance comparison with the psychometric model of EFL. The dataset is comprised of variables such as home ownership, number of dependents, time in the same location, gender, marital status, education level, loan term, disbursement date, and so on.

Finally, we collected data from the pilot conducted with the semi calibrated EFL scores for recruiting new clients between May 2015 and June 2016. This category of data contains 250 observations. The data include the scores of the traditional model implemented by Sogesol, the EFL scores, the ratings (A, B, C and F) associated with each score, the DPD after MoB, amount disbursed, dis-

bursement date, loan term, loan outstanding after MoB, and some socio-demographic characteristics, such as gender, age, marital status, education level.

4.1.2. Research Methodology

1) Back testing of the calibrated psychometric model

In order to meet the objective and test our hypotheses, we used the data of Sogesol on the psychometric scores and borrowers' repayment performance. We adopted different definitions of good and bad clients, including the one used in the development of the customized psychometric model. We proceeded to an exploratory data analysis, considering the bad and good definition used in the development of the calibrated model, by visualizing the bad rate by some socio-demographic variables used in the psychometric model. Since the EFL model was developed and implemented in a proprietary black-box approach, the paper did not produce any analysis on the psychometric variables. A vintage analysis, using several MoBs, has been produced.

This first level of analysis is followed by the back testing of the EFL calibrated model. This part contains two components: 1) the assessment of the EFL model discriminatory power and 2) the assessment of the calibration of the EFL model. The first one includes key metrics such as:

- a) Kolmogorov-Smirnov (K.S.);
- b) Receiver Operating Characteristics Curve (ROC)/Area Under the Curve (AUC)

With regards to the calibration, the following statistics are used:

- a) Hosmer-Lemeshow (HL);
- b) Brier Score.

Since the psychometric model was deployed in a socio-demographic scoring environment, we produced a comparative analysis of the two different scoring models, for each good and bad definition adopted by the paper.

2) Assessment of the pilot of the EFL model

In order to test the ability of the psychometric scoring model to mitigate risk, Sogesol has conducted a pilot from May 2015 to June 2016, where the priority in granting new loans was given to the psychometric score. For this purpose, a threshold score was defined. Based on that, commercial strategies were defined. Since the EFL score had not been applied yet alone to show its predictive capacity, a loan was rejected if and only if it received the rating F in both traditional model and psychometric model. Sogesol wanted to make sure not to lose potential clients by basing its decision-making exclusively on the EFL. The methodology applied to carry out the analysis is explained in the section 4.4.2.

Excel and R Language are the two tools used to conduct the evaluation.

4.2. Definitions of Good and Bad Clients

4.2.1. Definition of the Psychometric Model Development

A client is defined as bad if his/her Days Past Due was over 30 after 6 months on Book (Bad30MoB6). Otherwise, the client is qualified good.

4.2.2. Alternative Definitions of Good and Bad Clients

In order to reinforce our analysis, besides the main definition indicated above, we used 5 other definitions. They are as follows:

- 1) A bad client is the one who had more than 60 days late after 6 months on book (DPD60MoB6). Otherwise, s/he is defined as good.
- 2) A bad client is the one who had more than 90 days late after 6 months on book (DPD90MoB6). Otherwise, s/he is defined as good.
- 3) A bad client is the one who had more than 30 days late after 9 months on book (DPD30MoB9). Otherwise, s/he is defined as good.
- 4) A bad client is the one who had more than 60 days late after 9 months on book (DPD60MoB9). Otherwise, s/he is defined as good.
- 5) A bad client is the one who had more than 90 days late after 9 months on book (DPD90MoB9). Otherwise, s/he is defined as good.

All of these definitions are used to evaluate both psychometric model and traditional model.

4.3. Back Testing of the Psychometric Model

Assessing the validity of a predictive model is a very critical task. There are two methods that are generally used to measure the performance of a predictive model: discrimination tests and calibration methods. This section aims to evaluate the psychometric model, using those two methods. Prior to that, the score distribution is analyzed.

4.3.1. Discriminatory Power of the Psychometric Model

Discrimination assesses a model's ability to correctly classify clients. In other words, it measures the capability of the model to separate good clients from bad ones. There are several tests to achieve the assessment. But the paper is interested in the following ones:

1) Kolmogorov-Smirnov (K-S) test

The Kolmogorov-Smirnov test (K-S test) is a non-parametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K-S test), or to compare two samples (two-sample K-S test)

(https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test). The K-S statistic measures a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The null hypothesis distribution of the K-S statistic is that the sample is indeed drawn from the reference distribution in the one-sample case, or that the samples are drawn from the same distribution in the two-sample case.

The empirical distribution function F_n for n i.i.d (independent and identically distributed) ordered observations X_i is given as follows:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{[-\infty, x]}(X_i)$$

where $I_{[-\infty, x]}(X_i)$ is the indicator function, equals to 1, if $X \leq x_i$.

The K-S statistic for a given cumulative distribution function $F(X)$ is:

$$D_n = \sup_x |F_n(x) - F(x)|$$

where \sup_x is the supremum of the set of distances.

In the field of credit risk management, the K-S test quantifies maximum vertical separation (deviations) between two cumulative distributions (good and bad clients) in credit scoring modelling. In other words, this statistic measures the degree of discrimination between good and bad clients. The result of the test can be between 1 and 100, where the higher the K-S, the better the discrimination.

Using the Bad30MoB6 as definition, the K-S of the psychometric model is 19.67%.

To confirm the performance of both psychometric and traditional models, we used several alternative definitions to calculate the K-S. The results obtained for each definition is presented in the table below.

As illustrated in **Table 1**, the psychometric model would present a better predictive capacity than the traditional model. The max K-S of 40% is reached by the psychometric model with the Bad90MoB6, while the traditional model displays a 22% K-S. The lowest K-S is observed for both psychometric and traditional models in the Bad30MoB9, which is respectively 16% and 11%. Whatever the DPD (30, 60, 90) used, with the MoB6, the two models perform better than with the MoB9.

Table 1. The K-S statistic using different definitions of bad clients.

Definition	Psychometric model	Traditional model
BAD30MOB6	20%	14%
BAD60MOB6	32%	22%
BAD90MOB6	40%	22%
BAD30MOB9	16%	11%
BAD60MOB9	19%	12%
BAD90MOB9	21%	13%

Source: Author's own calculations.

2) Receiver Operating Characteristic (ROC)

A Receiver Operating Characteristic curve (ROC curve) is a graphical plot that shows the diagnostic ability of a binary classifier system as its discrimination threshold is varied (wikipedia.org). The ROC curve is built by plotting the true positive rate (TPR) or sensitivity against the false positive rate (FPR) or (1-specificity) at various threshold settings.

The ROC is one of the methods used in credit risk management to determine the discriminatory ability of a credit scoring model. Let us consider C as a cut-off providing a simple decision rule to divide clients into potential good and bad. As indicated in **Table 2**, four situations can happen.

Table 2. Decision results.

	Observed	
	Bad clients	Good clients
Scores Above C	True positive prediction (Sensitivity)	False positive prediction (1-Specificity)
Below C	False negative prediction (1-Sensitivity)	True negative prediction (Specificity)

Source: Adapted from Wu, 2008.

If a client with a score above C is identified as bad client or a good client has a score below C, the prediction of the model is then correct. Otherwise, the credit scoring model makes wrong prediction. The proportion of correctly predicted bad clients is named Sensitivity and the proportion of correctly predicted good clients is named Specificity.

It is very important to remember that the false positive prediction (1-Specificity) is known as type I error, defined as the error of rejecting a null hypothesis that should have been accepted. The false negative (1-Sensitivity) is known as type II error, i.e. the error of accepting a null hypothesis that should have been rejected.

From a risk perspective, Sensitivity denotes those cases which are both actually bad clients and predicted to be bad clients as a proportion of total bad cases. Specificity indicates cases which are both actually good clients and predicted to be good clients as a proportion of total good cases (Abdou et al., 2016).

The ROC curve is a representation of a set of coordinates specified by the True Positive Rate (TPR) and False Positive Rate at different values of C (cut-off). For the perfect model, the graph for the ROC curve passes through the upper left corner, where the share of the false positive outcomes is equal to zero. The closer is the curve to the upper left corner, the higher is the predictive power of the model. The diagonal line (line of no discrimination or random guess) indicates the bad model (Garanin et al., 2014).

The area under the ROC curve (AUROC) offers a measure of a model's discriminatory power. The AUC is between 0.5 and 1. For a random model (useless model), the AUC is 0.5, for a perfect model, the AUC is 1. A model with greater power presents a larger AUC.

In general, AUC values are interpreted as follows (Abdou et al., 2016):

- a) $0.5 \leq \text{AUC} < 0.6$ = fail;
- b) $0.6 \leq \text{AUC} < 0.7$ = poor;
- c) $0.7 \leq \text{AUC} < 0.8$ = fair;
- d) $0.8 \leq \text{AUC} < 0.9$ = good;
- e) $0.9 \leq \text{AUC} \leq 1.0$ = excellent.

In the case of Sogesol, with a Bad30MoB6 definition, the ROC for respectively the psychometric model and the traditional model is as follows.

As observed in Figure 3, the psychometric model displays a greater discriminatory power. Its ROC curve is more distant to the diagonal line than the one of the traditional model. At some levels, the ROC curve of the traditional model is even confounded with the diagonal line which is referred to as a model without

discriminatory power.

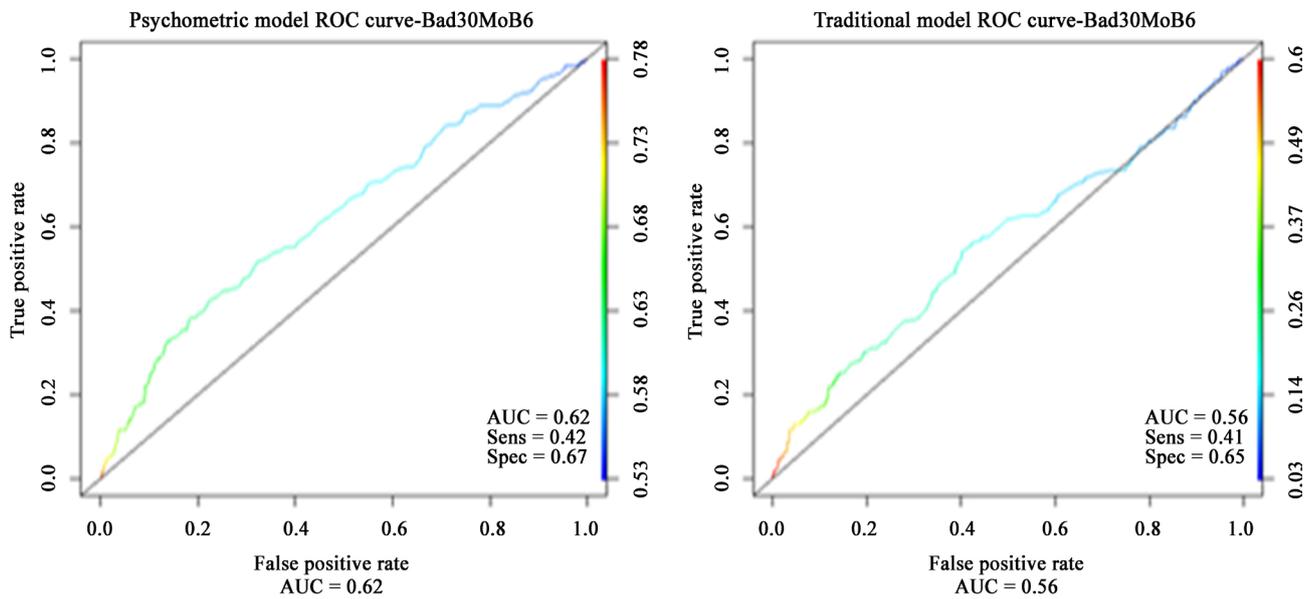


Figure 3. Psychometric and traditional models ROC curve.

That observation is corroborated by the AUC values, which are respectively 0.62 for the psychometric and 0.56 for the traditional model. However, both models present a relatively poor performance, since the AUC is less than 0.70. And we can even speak about the failure of the traditional model. The following table shows the AUC of both models for different definitions of bad clients:

Looking at Table 3: we can assert that the psychometric model provides a better discriminatory capability than the traditional model. For all of the definitions, the EFL model displays a greater AUC. As for the K-S, the maximum AUC is attained by the Bad90MoB6.

Table 3. The AUC statistic using different definitions of bad clients.

Definition	Psychometric model	Traditional model
BAD30MOB6	0.62	0.56
BAD60MOB6	0.68	0.60
BAD90MOB6	0.70	0.60
BAD30MOB9	0.61	0.53
BAD60MOB9	0.63	0.52
BAD90MOB9	0.64	0.52

Source: Author's own calculations.

On the other hand, even though the difference is not significant when comparing the sensitivity and the specificity of the two models, we can observe that the psychometric model is better with a sensitivity (0.42) and specificity (0.67) larger than the sensitivity (0.41) and the specificity (0.65) of the traditional mod-

el. In fact, for a given level of specificity, the model with the higher sensitivity is preferred. As well, for a given level of sensitivity, the model with the higher level of specificity is preferred (Abdou et al., 2016).

4.3.2. Calibration of the Psychometric Model

As illustrated previously, the discrimination tests inform about the ability of the EFL model and the traditional model to separate good and bad clients. However, they are not able to confirm whether or not the credit scoring model is calibrated. Calibration is defined as the ability of the model to make unbiased estimates of the probabilities of default. The calibration of a credit scoring model compares the realized default frequency with the estimates of the conditional probability of default, given the score and analyzes the difference between the observed default frequency and the estimated probability of default. Calibration tests can be then used to measure the reliability of predicted probabilities (Fenlon et al., 2018). The Hosmer-Lemeshow test and the Brier test are retained in this research.

1) Hosmer-Lemeshow (HL) test

The Hosmer-Lemeshow test is a statistical test for goodness of fit for logistic regression models (Wikipedia.org). It is commonly used in risk prediction models. The test evaluates whether or not the observed rates match expected event rates in subgroups of the model population. The Hosmer-Leshow test precisely identifies subgroups as the deciles of the risk fitted values. Models for which expected and observed event rates in subgroups are similar are well calibrated. A small p-value (usually less than 0.05) suggests the rejection of the null hypothesis stipulating that the expected and observed event rates are similar. The prediction is poor (lack of fit), indicating problems with the model.

A HL test was conducted with Sogesol's data for both psychometric and traditional models. The results are presented below.

Table 4 and **Table 5** display the Hosmer-Lemeshow test for the psychometric model and traditional model, respectively. As indicated in both tables, the observed (Good, bad clients) are not similar to the expected (Good, bad clients), even though we can observe for the traditional model, the observed and the expected for both good and bad clients are relatively closer. Additionally, using R software, we obtained a Chi-square statistic of 4720 for Psychometric model and 324.65 for the traditional model, with a p-value less than 0.05 (p-value < $2.2e-16$ for both models). The p-value is below $\alpha = 0.05$, so the null hypothesis that the observed (Good, bad clients) and the expected (Good, bad clients) are the same across all score ranges is rejected. The result suggests that both models are not a good fit. So, the probabilities obtained from the models would be biased. In other words, the probabilities might not inform about the actual client's risk profile. They are not consistent. It is very important to mention that the HL test was also conducted for the alternative bad loan definitions. The results are not different from the ones observed for the definition used in the psychometric model development, Bad30MoB6.

Table 4. Hosmer-Lemeshow test groups. Psychometric (Bad30MoB6).

Scores range	Observed		Expected	
	Good	Bad	Good	Bad
10	372	11	173.86	209.14
9	363	8	162.82	208.18
8	356	13	158.13	210.88
7	334	16	146.82	203.18
6	365	14	155.9	223.1
5	348	15	145.93	217.07
4	359	16	146.95	228.05
3	348	13	136.82	224.18
2	331	28	129.48	229.52
1	323	38	113.94	247.06

Notes: For reason of confidentiality, the range of the EFL scores are not displayed, they are replaced by number from 10 to 1, where 10 represents the class with the higher scores and 1 the one with the lowest scores.

Table 5. Hosmer-Lemeshow test groups. Traditional model (Bad30MoB6).

Scores range	Observed		Expected	
	Good	Bad	Good	Bad
10	223	14	220.59	16.41
9	202	13	193.07	21.93
8	238	11	217.55	31.46
7	192	10	173.51	28.49
6	286	11	247.32	49.68
5	140	9	121.53	27.47
4	227	19	194.35	51.65
3	192	10	155.97	46.03
2	206	19	158.48	66.52
1	202	23	116.95	108.05

Notes: For reason of confidentiality, the range of the EFL scores are not displayed, they are replaced by number from 10 to 1, where 10 represents the class with the higher scores and 1 the one with the lowest scores.

2) Brier test

The Brier score is referred to as an overall goodness-of-fit check for a model predicting binary or categorical response values (Brier, 1950). Alongside other metrics, it is ordinarily used to measure credit scoring model performance. The original definition of Brier is as follows (Kraus, 2014):

$$BS = \frac{1}{n} \sum_{j=1}^r \sum_{i=1}^n (p_{ij} - Y_{ij})^2$$

where p_{ij} denotes the forecast probabilities, Y_{ij} takes the value 0 or 1 according to whether the event occurred in class j or not and r defines the possible classes ($r = 2$ for default and non-default). So, the Brier score is defined as the squared difference of the predicted probabilities p_{ij} and the observed default rates within each category. The Brier score takes on a value between zero and one. The lower the Brier score of a model, the better is the predictive performance. The Brier score incorporates elements of both discrimination and calibration, since it compares numerical outcomes (in the case of a binary result, 0 and 1) to predicted probabilities, without the grouping used by the other calibration techniques (Fenlon et al., 2018).

The Brier score was calculated for the sample under study with the different definitions of good and bad clients, in order to measure the accuracy of the psychometric model implemented at Sogesol. The table below presents the results of the test:

Surprisingly, as shown in **Table 6**: the best Brier scores are obtained for the traditional model for every definition of good and bad clients. The maximum Brier score attained is 0.36 (Psychometric model) and 0.12 (Traditional model), while the minimum reached is 0.33 (Psychometric model) and 0.06 (Traditional model). As observed, the difference between the Brier scores of both models is relatively large and the Brier scores of the traditional model are closer to 0. We should point out that the best Brier scores (0.06 and 0.07) of the traditional model are obtained with the definitions (Bad60MoB6 and Bad90MoB6) for which we also obtained the higher AUC values (0.6 and 0.6).

Table 6. Brier score comparison: Psychometric and traditional models.

Definition	Model Type	
	Psychometric model	Traditional model
BAD30MOB6	0.35	0.09
BAD60MOB6	0.36	0.07
BAD90MOB6	0.36	0.06
BAD30MOB9	0.33	0.12
BAD60MOB9	0.33	0.12
BAD90MOB9	0.34	0.11

Source: Author's own calculations.

However, the result of the Brier score alone is not sufficient to indicate that the traditional model is calibrated, since the HL test suggests the contrary. Moreover, it would be controversial when taking into account the outputs of the KS and the AUC that would suggest the poor performance of the traditional model while the EFL tool performs better.

4.4. Evaluation of the Psychometric Model Pilot

4.4.1. Implementation of the Psychometric Model as a Pilot

In order to test the predictive power of the psychometric model, Sogesol made

the decision to evaluate all of the new working capital loans through the semi-calibrated model, since at that time the calibrated model had not been developed yet by EFL. That pilot was conducted in all of the branches from May 2015 to June 2016. A total of 250 clients were evaluated. The objective is to see if the psychometric score could increase the discriminatory power in the credit decision, since Sogesol wanted to have a hybrid model with the traditional model and the psychometric tool. To do so, a cut-off was defined by the implementing institution and was shared with EFL. A matrix containing classes (A, B, C and F) was therefore established. A potential client was rejected a loan if and only if he/she obtained an F score in both psychometric and traditional models. The following table provides the results of the credit decisions:

As observed in **Table 7**: 64.4% of loans would have been accepted by both models. 16.8% would have been rejected by EFL score and accepted by the traditional model. The 10.8% that would have been rejected by the traditional model would have been accepted by EFL score. Finally, only 8.0% of the loans were actually rejected, since the score was F for both models.

Table 7. Credit decisions based on the EFL Score and the Traditional Score.

		Traditional score decision	
		Accepted	Rejected
EFL decision	Accepted	0.644	0.108
	Rejected	0.168	0.08

Source: Author's own calculations.

4.4.2. Psychometric Model and Risk Reduction

The following table displays the DPD rate for clients accepted under both models and clients accepted by the traditional model and rejected by the EFL model according to different definitions of DPD after months on books. This allows us to measure the contribution of the psychometric model in helping the traditional model screening out high risk potential borrowers. Therefore, we expected to obtain a better DPD rate for clients accepted under both models.

As shown in **Table 8**, taking into consideration all of the DPD definitions, loans approved under both models would display a better DPD rate. A look at the difference column quickly informs about that performance, since only for the DPD30MOB6, the difference is positive, indicating a lower DPD rate for clients accepted by the traditional model and rejected by EFL score. This result would suggest that the psychometric model could help improve credit decisions. However, it is critical to rule out that the result is not only due to chance. To do so, we analyzed the correlation between different DPD definitions and a binary variable. The statistical approach adopted is the estimation of a linear regression model taking the following form:

$$y_i = a + bx_i + \epsilon_i, i \in S$$

where y_i is either a binary variable or a continuous variable. For instance, total

Table 8. Days past due (DPD) rate according to credit decisions.

Days past due (DPD)	Accepted under both Traditional and EFL (DPD rate)	Accepted by Traditional and rejected by EFL (DPD rate)	Difference
More than 5 days past due after 3 months on Book	3.11%	4.76%	-1.66%
More than 5 days past due after 3 months on Book	3.11%	4.76%	-1.66%
More than 30 days past due after 6 months on Book	5.59%	4.76%	0.83%
More than 60 days past due after 6 months on Book	2.48%	4.76%	-2.28%
More than 30 days past due after 9 months on Book	11.80%	11.90%	-0.10%
More than 60 days past due after 9 months on Book	9.94%	11.90%	-1.97%
More than 90 days past due after 12 months on Book	4.97%	9.52%	-4.55%
More than 30 days past due after 7 months on Book	11.18%	16.67%	-5.49%
More than 30 days past due after 8 months on Book	12.42%	19.05%	-6.63%

Source: Author's own calculations.

number of days past due after 6 months on book for each client i (continuous variable) or a variable equal to 1 if the client i had more than 90 days past due after 9 months on book, and 0 if the client i was less or equal to 90 days past due after 9 months on book (binary variable); x_i is a variable equal to 1 if the client i was rejected by the EFL score and accepted by Sogesol's traditional score and 0 if the client i was accepted by both psychometric model and socio-demographic model; ϵ_i is the regression error term; and S is defined as the sample of clients under study.

The outputs of the linear regression of each DPD definition with the binary variable for the adverse psychometric test are presented in the table below.

As indicated in **Table 9**, there would not be any relationship between the risk

Table 9. Linear regression results.

DPD	Dummy variable			
	estimate	Std. Error	t value	Pr (> t)
DPD_MOB12	7.928	7.428	1.067	0.28711
DPD_MOB9	-0.2236	5.5647	-0.040	0.968
DPD_MOB8	3.530	5.214	0.677	0.499
DPD_MOB6	1.984	2.626	0.756	0.45065
DPD_MOB4	-0.02484	1.61773	-0.015	0.98776
DPD_MOB3	1.220	1.181	1.033	0.30292
DPD90MOB12	0.04555	0.04094	1.113	0.26721
DPD90MOB9	0.02588	0.05186	0.499	0.618304
DPD30MOB9	0.001035	0.056221	0.018	0.985
DPD60MOB9	0.01967	0.05301	0.371	0.711
DPD60MOB6	0.02277	0.02945	0.773	0.4402
DPD30MOB6	-0.008282	0.039415	-0.210	0.83379
DPD_MOB2	0.7484	0.3701	2.022	0.04448*

level and the decision of granting a loan or not. The adverse psychometric test variable is not statistically significant in any linear regression with the different DPD variables, except with DPDMOB2 (Total days past due after 2 months on book), it is statistically significant at the level of 5%. Contrary to the results observed previously, this would suggest that the psychometric tool would not have significantly mitigated risk in Sogesol's portfolio, since the DPD cannot be expected to reliably respond to observing the adverse psychometric test result in the credit decisions.

4.5. Discussion of Results

Based on the different definitions of good and bad clients and the scores, the paper evaluated the performance of the psychometric model implemented at Sogesol, in comparison with the existing traditional model. The findings indicate that the model developed by EFL would display a better discriminatory ability, with higher values for the discrimination tests (KS, AUC). However, for the six definitions used, the maximum obtained for each metric is not strong enough to assert that the psychometric model is of a good quality. That would confirm our hypothesis that the psychometric score shows a low predictive power. In addition, the results of the calibration tests suggest that the default probabilities obtained from the psychometric model are not accurate, whereas results are better for the traditional model.

On the other hand, the evaluation of the pilot indicated that the psychometric model would not have contributed to screening out high-risk clients, since there is no statistical significance when analyzing the relationship between various behavioral response variables and the binary variable: 1 = clients accepted by the traditional model and rejected by EFL model; 0 = clients accepted under both models. So, as our second hypothesis stipulates it, the EFL score may fail to recruit clients of low-risk profile. These findings suggest that the development of a hybrid model of psychometric and traditional scores would not provide any added value in the credit decision.

However, it is very important to note the limitations of the findings summarized above. The results may be influenced by the definitions of good and bad clients chosen. For instance, regardless the number of months on books, the higher the DPD chosen, the better the metrics. Maybe with a definition of a DPD higher than 90, the metrics could be improved. Regarding the calibration, we note that the tests, for instance the HL, are often criticized as being suboptimal in the evaluation of the probability prediction.

Furthermore, one of the limitations of the pilot is the size of the sample (250 loans). Drawing a conclusion with such a sample of clients may not be reliable. Nonetheless, the model can give an indication of the future potential value of using psychometrics in microfinance. Also, the production of loans in the category of Sogesol's clients studied here is rather small, with less than 40 new loans on average disbursed monthly. We should also note that the pilot was not con-

ducted with the final model delivered to Sogesol by EFL, since Sogesol did not want to wait for the availability of the latest model to experiment with the psychometric model alone in its credit decisions.

5. Sogesol's New Scorecard Development

5.1. The Data

The data used to develop the model for new working capital clients is collected from Sogesol's database. The sample contains 5,776 loans disbursed from October 2012 to December 2017, more than five years of historical data on repayment behavior of new clients that are sufficient to develop a credit scoring model. The dataset is comprised of 23 variables (numerical or categorical). They are described in the following table.

The variables described in **Table 10** are part of Sogesol's loan application. However, some of them have not been used yet in the existing credit scoring models of the institution under study. Those variables are: Phone number, Sogesol Awareness, Credit experience, Bank account, Bank account, Sogebank savings account and Client reference. Their choice in the sample is due to the fact that they could likely be predictive of borrowers' repayment behavior.

Table 10. Variables definition.

Var. #	Variable name	Description	Variable type
1	Branch (BV, CB, CF, CH, CY, DM, GV, JM, JR, MB, OUA, PG, PP, PV, RD, SM)	Branch where the loans were granted.	Categorical
2	Gender (female, male)	Sex of the client.	Categorical
3	Marital status (Divorced, living with partner, married, single)	Set of indicators that stand for the marital status of the client.	Categorical
4	Education Level (None, primary, secondary, technical, higher)	Indicates the education level reached by the client.	Categorical
5	Economic sector (Trade, service, production)	The economic sector in which the client is occupied.	Categorical
6	Activity type (Sundries, food supplies, restaurant, new clothes, used clothes, soft drinks, building materials, shoes, other)	Indicates the economic activity in which the client evolves.	Categorical
7	Phone number (One, two and more)	Indicates the quantity of phone numbers the client owns.	Categorical
8	Local enterprise type (Rent, ownership with debt, paid ownership, with others, other)	Set of indicator variables stand for client's current situation regarding the ownership of the business local.	Categorical
9	Local residence type (Rent, ownership with debt, paid ownership, with others, other)	Set of indicator variables stand for client's current situation regarding the ownership of the house where s/he lives.	Categorical
10	Sogesol Awareness (Advertising, information session, recommendation, segment visit, other)	Set of indicator variables stand for the way the client is informed about Sogesol.	Categorical
11	Credit experience (Yes, no)	Indicates if the client has credit experience, besides the loan s/he requests.	Binary
12	Bank account (Yes, no)	Indicates if the client owns a bank account or not.	Binary
13	Sogebank savings account (Yes, no)	Indicates if the client has a savings account at Sogebank (Sogesol's parent company)	Binary

Continued

14	Client reference (Friend, neighbor, parent, other)	Set of indicator variables that stand for the origin of testimony given about the client.	Categorical
15	Origin of other income (None, local, rent, other enterprise, remittance, salary, other)	Set of indicator variables that stand for the origin of client's other incomes.	Categorical
16	Monthly Income (≤ 500 ; [500, 1000]; [1000, 2000]; 2000<)	Set of indicator variables that stand for the range of monthly incomes generated by client's business in US\$.	Categorical
17	Monthly sales (≤ 2500 ; [2500, 4000]; [4000, 7500]; 7500<)	Set of indicator variables that stand for the range of monthly sales realized by client's business in US \$.	Categorical
18	Age (≤ 30 ; [30, 40]; [40, 50]; 50<)	Age of the client measured in years.	Categorical
19	Number of dependents (≤ 1 ; [1, 2]; [2, 3]; 3<)	Number of persons the client is financially responsible for.	Categorical
20	Enterprise since (≤ 36 ; [36, 72]; [72, 120]; 120<)	Set of indicator variables that stand for the range of the time of existence of client's enterprise in month. It indicates the experience of the client as entrepreneur.	Categorical
21	Present residence since (≤ 36 ; [36, 66]; [66, 120]; 120<)	Set of indicator variables that stand for the range of the time that the client lives at the current residence in months.	Categorical
22	Present enterprise since (≤ 36 ; [36, 66]; [66, 120]; 120<)	Set of indicator variables that stand for the range of the time that the client owns his/her current business in months.	Categorical
23	Over 90 Days past due after 9 Months On Book (DPD90MOB9)	Indicator variable that marks 1 when the client has more than 90 days after 9 months on books, and 0 if otherwise.	Binary

Lastly, we should mention that the numerical variables such as age, enterprise since, monthly sales, etc. were discretized to obtain the different bins, using percentiles or quartiles as for obtaining the interval cut-offs.

5.2. Definition of Good/Bad Client

With regards to the development of the new model for Sogesol, this research adopts the following definition of good/bad client:

- A bad client is the one that has more than 90 days past due after 9 months on books (DPDMOB9 > 90).
- A good client is the one whose days past due are less or equal to 90 after 9 months on books (DPDMOB9 \leq 90)

The DPD90MOB9 turns out to be considered an optimal definition, since among definitions tested, it would provide a more performing model. In addition, the choice of DPD90 could be explained by the fact that Sogesol considers loans as nonproductive (in default) after 90 days past due.

5.3. Scorecard Development Methodology

On the subject of developing a scorecard model to predict the repayment behavior of loan borrowers, several mathematical techniques are available. Among others, we can mention: logistic regression, decision tree, random forest, neural networks, and so on (Siddiqi, 2006).

This research adopts one of the most common, successful and transparent

technique, i.e. logistic regression.

5.3.1. Logistic Regression

Like most of other predictive modeling techniques, logistic regression uses a set of predictor variables to estimate the probability of a particular outcome (Siddiqi, 2006). The equation for the logit transformation of a probability of an event is displayed as follows:

$$\text{Logit}(p_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

where:

p = posterior probability of “event”, given inputs;

$x_1 \dots x_k$ = input variables;

β_0 = intercept of the regression line;

$\beta_1 \dots \beta_k$ = regression coefficients.

In the case of a scorecard, the event is set to “bad” and the non-event to “good”.

Logit transformation is defined as the natural logarithm of the odds, i.e. $\ln(p(\text{bad})/p(\text{good}))$. It serves to linearize posterior probability and limits outcomes of estimated probabilities in the model to between 0 and 1. To estimate the regression coefficients β_1 to β_k , maximum likelihood is used. These parameters determine the rate of change of logit for one change in the input variable (adjusted for other inputs). In other words, they represent the slopes of the regression line between the target variable (good/bad) and their respective input variables x_1 and x_k . It is important to underline that the parameters depend on the unit of the input. In order to facilitate the analysis, they have to be standardized.

Regression requires a target or response variable and a set of independent inputs. Different forms of inputs exist. However, the method commonly used is to consider the raw input data for numeric variables and create binary variables for categorical data. To counterbalance the effects of input variable units, standardized estimates are used in the analysis.

In order to obtain the best possible model using all options, regression can be run. It is usually referred to as “all possible” regression techniques. The three following techniques are generally used in logistic regression (Siddiqi, 2006):

- Forward selection: This technique chooses first one characteristic model based on the individual predictive power of each characteristic, then adds further characteristics to this model to create the best two, three, four, and so on characteristic models incrementally, until no remaining characteristics have p -values of less than some significant level (for instance 0.5), or univariate Chi-Square above a determined level. The forward technique is an effective one. However, it can present some weakness if characteristics are too much numerous or there is high correlation.
- Backward elimination: This method is the opposite of forward selection. It

begins with all the characteristics in the model and sequentially eliminates characteristics that are considered the least significant, given the other characteristics in the model, until all the remaining characteristics have a p -value below a significant level or based on some other measure of multivariate significance. With this method, all of the characteristics of lower significance have a higher chance to be part of the model.

- **Stepwise:** This method combines the two previous ones. It implies adding and removing characteristics dynamically from the scorecard in each step, until the best combination is attained. Minimum p -values (or Chi-Square) required can be set to be added to the model, or to be kept in the model.

In this research, the stepwise logistic regression method is used to build the model for assessing new applicants for a working capital loan at Sogesol. To do so, R language is used as a statistical tool to run the regression.

5.3.2. Training/Validation Samples

In the scorecard development process, the total sample should be divided into training sample and validation sample. The training sample serves to develop the scorecard and the validation sample is held aside to test the scorecard obtained.

Based on the definition of good/bad client above, the whole sample of Sogesol is divided as indicated in the table below.

As shown in **Table 11**, 75% of the whole sample is used to develop the scorecard and 25% for validation. The distribution of bads is almost the same for the different samples (training = 11.1%; validation = 10.5% and total = 11.0%).

Table 11. Training/validation samples and distribution of good/bad clients.

Sample	Goods	% Goods	Bads	% Bads	Total	%
Training	3849	88.9%	483	11.1%	4332	75.0%
Validation	1293	89.5%	151	10.5%	1444	25.0%
Total	5142	89.0%	634	11.0%	5776	100.0%

Source: Author's own calculation.

5.3.3. Scorecard Model Results

The credit scoring model developed to screen new applicants is based on a binary logistic regression, where the target variable is the good/bad repayment classification as defined previously. The following table displays the results of the model:

As observed in **Table 12**, 10 out of 23 variables of the dataset are retained in the scorecard. The scorecard has a constant of -1.39845 and 39 elementary characteristics. The coefficients of the variables are positive or negative. The outcome corresponds to the probability of bads placed between 0 and 1.

5.3.4. Score Calculation

Once we have the estimates of each variable, we can calculate the scores. The

Table 12. Sogesol's working capital selection scorecard.

Variable	Characteristic	Coefficient = β	#Variable (v)
Intercept		-1.39845	
Branch	CF	-0.21292	v1
	CY	-1.27435	V2
	GV	-0.66413	V3
	JR	-0.42539	V4
	OUA	-1.485	V5
	PP	-0.64245	V6
	RD	-0.06284	V7
	CB	-0.55401	V8
	CH	-1.13636	V9
	DM	0.1043	V10
	JM	-1.16115	V11
	MB	-1.50126	V12
	PG	-0.45969	V13
	PV	-0.89793	V14
	SM	-0.42613	V15
Marital status	Living with Partner	-0.47696	V16
	Single	-0.20719	V17
	Married	-0.6439	V18
Phone number	Two numbers and more	-0.77233	V19
Local residence type	Ownership with debt	-1.57365	V20
	Paid ownership	0.08787	V21
	Rent	0.43309	V22
	With others	0.33447	V23
	Information session	-0.92283	V24
Awareness Sogesol	Recommendation	-0.30732	V25
	Segment visit	0.02489	V26
	Other	-2.10276	V27
Credit experience	Yes	-0.53663	V28
Savings account Sogebank	Yes	-0.54762	V29
Origin of other incomes	Other enterprise	1.31072	V30
	Remittance	0.70014	V31
	None	1.70619	V32
	Salary	1.46415	V33
Monthly Sales	[2500, 4000]	0.16314	V34

Continued

	[4000, 7500]	-0.1332	V35
	7500<	-0.48805	V36
Age	[30, 40]	0.26965	V37
	[40, 50]	0.19598	V38
	50<	-0.13675	V39

Source: Author’s own calculation.

model equation is given as follows:

$$model = constant + (V1\beta_1) + (V2\beta_2) + (V3\beta_3) + \dots + (V39\beta_{39})$$

where v = the variables from the scorecard and β = the value of the coefficient of each variable.

The probability of bads is defined by the following equation:

$$p = \frac{1}{e^{-model}}$$

The probability of good is then equal to 1 minus the probability of bad ($p_{good} = 1 - p$).

Using p_{good} the score is calculated as follows:

$$Score = p_{good} * 1000$$

So, the higher the score, the less risky is the new applicant.

5.4. Model Performance Assessment

Discriminatory Power of the Model

We assess the discriminatory power of the above model for new working capital clients at Sogesol with the same metrics used previously to evaluate the psychometric model developed by EFL for Sogesol.

1) Model performance-training sample

The KS, AU values indicate that the model is robust. The outcomes are shown in table below:

As shown in **Table 13**, a KS of 32.59% indicates that the model has the ability to discriminate between good clients and bad clients. In other words, based on that result, the model can separate goods from bads. In addition, with regards to the respective benchmarks established for AUC, we would attest that the model developed performs well. Finally, the model outperforms the psychometric model, since, over all, the metrics of this traditional scorecard built with some new variables included are greater than those obtained in the psychometric assessment.

Table 13. Metrics values-training sample.

Metric	Value
KS	32.59%
AUC	0.72

Source: Author’s own calculation.

The ROC curve below supports the robustness of the model:

As observed in **Figure 4**: the ROC curve displays enough lift above the 45° line of a useless model, indicating the predictive power of the model.

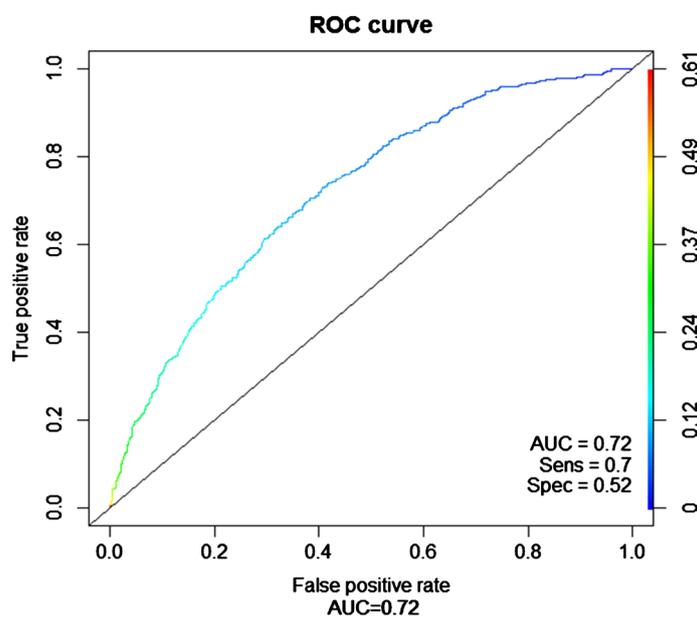


Figure 4. ROC Curve-training sample.

2) Model performance in the validation sample

No matter how good the model may appear within the training sample, if it is not the case in the test sample, the model will not be considered reliable in predicting new applicants' repayment behavior. As shown in **Table 14**, the results obtained for the validation sample are almost the same as those observed for the training sample.

Table 14. Metrics values-validation sample.

Metric	Value
KS	32.38%
AUC	0.7

Source: Author's own calculation.

These indicators are also within the good range of benchmark respectively. Compared to the results obtained for both psychometric model and traditional model from the previous back testing, the new model is shown better.

The predictive power of the model is illustrated in the ROC curve below:

As shown in **Figure 5**: like in the graph of the training sample, the ROC curve of the validation sample is far enough from the line of a useless model to confirm the robustness of the new model.

It is important to highlight that KS and AUC values for the validation sample were not obtained from running another stepwise forward logistic regression on the validation sample, but by applying the model previously estimated to the

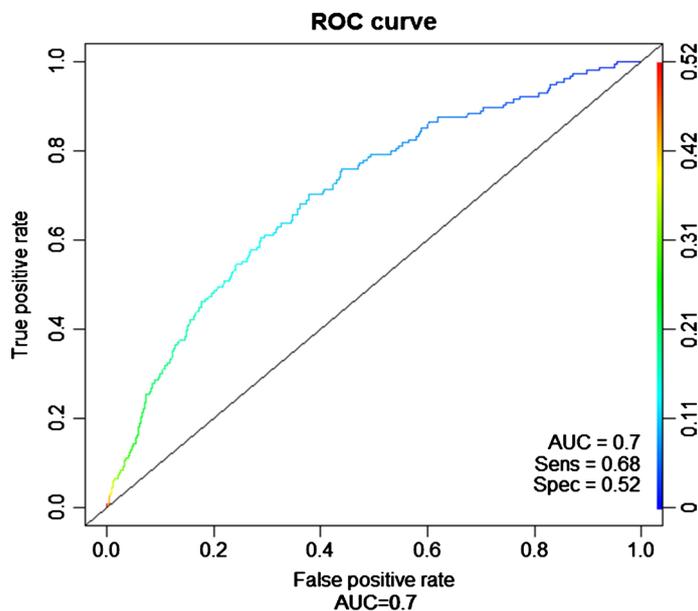


Figure 5. ROC Curve-validation sample.

reserved validation sample and then stepping through various cut-points to create the ROC curve.

5.5. Calibration of the New Model

The previous statistical tests inform about the ability of the model to separate goods from bads. Nonetheless, they are not sufficient to conclude about the ability of the model to really predict borrowers’ repayment behavior. That is why it is critical to run statistical tests to know if the model is well-calibrated. For that, we used the HL test and the Brier test.

The HL test gives a *p*-value = 0.99, bigger than 0.05, indicating no evidence of poor fit. The model is therefore correctly specified. The following table of observed vs expected has been obtained:

As shown in Table 15, in each bin the observed goods are close enough to the expected goods. The same result is observed for the bads, suggesting the accuracy of the probabilities obtained from the model.

Table 15. Hosmer-Lemeshow test groups-New model.

Scores range	Observed		Expected	
	Good	Bad	Good	Bad
971≤	426	8	424.94	9.06
957 - 970	423	13	419.99	16.01
944 - 956	409	21	408.45	21.55
929 - 943	403	30	405.47	27.53
913 - 928	397	37	399.69	34.31
893 - 912	392	41	391.02	41.98

Continued

869 - 892	380	53	381.46	51.54
833 - 868	368	64	367.81	64.19
772 - 832	347	87	349.49	84.51
≤771	304	129	300.68	132.32

Source: Author's own calculation.

That observation is reinforced by the outcome of the Brier test. In fact, a Brier score of 0.09 is found, suggesting that the model is well-calibrated. The Brier score gives a result between 0 and 1, the lower the score, the better model.

5.6. Models Comparison

It is very useful to compare the predictive power of the new traditional credit scoring proposed by this paper and the EFL tool evaluated in the previous chapter. The following table illustrates the comparison:

As indicated in **Table 16**: the new traditional outperforms the socio-demographic model currently used at Sogesol to screen out high risk new loan applicants. Compared to the EFL tool, the result is not different, since the new model presents both higher KS and AUC, indicating that a better predictive power than the psychometric model.

Table 16. Traditional and Psychometric models comparison.

Metric	New traditional model	Current traditional model	Psychometric tool
KS	32.38%	13.94%	19.70%
AUC	0.7	0.52	0.62

Source: Author's own calculation.

5.7. Discussion of Results

The outcomes of the back testing of the psychometric model and the existing traditional scorecard used at Sogesol for new borrowers have suggested the construction of a new scorecard as an alternative. To do so, we have selected a sample of historical data from Sogesol's information system. The dataset has been extended to more than five years, containing enough data to build the predictive model.

Several alternative definitions of good/bad clients have been tested. The retained one (Bad90MOB9) has been found to be optimal, since it displays the best statistical performance both in terms of discriminatory power and calibration. Besides the robustness of the new model, one important thing discovered is the predictive power of the new variables (credit experience, awareness Sogesol, etc.), never used before at Sogesol. This analysis teaches us that all data is credit data, which suggests that Sogesol should continue to optimize the use of the information captured in its credit evaluation process.

Moreover, the new model could help Sogesol in screening good borrowers that would increase its portfolio while enhancing portfolio quality. It is more reasonable to use such a scorecard than a psychometric model that remains a black box for Sogesol, where the institution could never be sure about the predictive power of the variables used. We believe that alternative data, such as psychometric characteristics, can potentially enhance the predictive performance, when combined with traditional data. However, Sogesol would need better control over the selection and design of the psychometric elements, such that those alternative data could be tested for predictive performance and customized to the Haitian country context.

Finally, an eventual combination of the EFL model and the existing sociodemographic scorecard would not have added value to Sogesol's efforts regarding risk mitigation, since results obtained in the back testing are relatively poor in terms of discriminatory power. The implementation of the new scorecard alone would produce better results, while helping Sogesol to enhance its contribution to financial inclusion in Haiti.

6. Conclusion

The paper has evaluated the predictive power of psychometric testing in micro-finance, based on evidence from Sogesol. The evaluation contains two parts: one consists of a back testing, using statistical tools (ROC curve, AUC, K-S) with different definitions of good and bad clients, to measure the predictive power of the psychometric model used by Sogesol to assess the creditworthiness of MSME's owners applying for a working capital loan of US\$ 1000 and more. For this purpose, a sample of 3671 clients has been selected from Sogesol's database. Using R software and Excel, we analyzed the performance of psychometric testing in comparison with the existing traditional working capital scoring. On the other hand, the paper analyzed a pilot conducted with a sample of 250 borrowers, where Sogesol used psychometric testing alone to grant loans. This empirical analysis of the research helped to measure the contribution of the psychometric credit scoring model in risk reduction at Sogesol.

The research finds that the psychometric model as implemented at Sogesol would present a low predictive power, confirming the initial hypothesis arguing that the psychometric scoring model cannot discriminate between good and bad borrowers. In fact, using different definitions of bad clients (Bad30MoB6, Bad60MoB6, Bad30MoB9, Bad60MoB9, Bad90MoB9), all of the metrics calculated display values under the normal values that would show low predictive power of the psychometric tool developed by EFL. The AUCs obtained vary from 0.62 to 0.70 (less than 0.70 = poor performance). Regarding the K-S, the values vary from 16% to 40%. The maximum AUC and K-S were registered for the Bad90MoB6. In addition, the psychometric tool would not be well-calibrated, with HL test presenting p -value < 0.05 . The probabilities of default obtained would then be biased. However, it is worth mentioning that the psychometric

model outperforms the traditional credit scoring model. Compared to results of the psychometric model, the values of the metrics (AUC, K-S) found for the socio-demographic model are smaller, varying from 11% to 22% for the K-S and from 0.52 to 0.60 for the AUC.

Moreover, the analysis of the pilot results indicates that the EFL tool would not contribute to mitigate risk in Sogesol's portfolio, confirming the hypothesis of the paper stipulating that the psychometric model has no significant impact on Sogesol's portfolio quality. This second hypothesis has been tested within a linear regression model, where the dependent variable is defined as Days Past Due and the independent variable is a binary variable set equal to 1, if the client was rejected by EFL score and accepted by the traditional score, and set to 0 if the client was accepted by both models. Overall, the binary variable is not statistically significant in any linear regression with the different DPD variables. Therefore, the DPD would not be affected by the credit decisions, while the objective of Sogesol was to strengthen the predictive power of the socio-demographic model by experimenting with the psychometric tool, by developing a hybrid model.

In this case, we found that the psychometric testing as applied by Sogesol at the time presents low discriminatory power. Based on this outcome, the paper proposed to re-estimate the existing socio-demographic credit scoring model, through a stepwise logistic regression using R software and Excel, in order to enhance its predictive power, since the EFL tool would not have added value. A sample of 5776 loans, disbursed between October 2012 and December 2017, was selected from Sogesol's database, while including also variables found in the loan application (credit experience, Sogesol awareness, quantity of phone members, etc.) that have not been used before in credit scoring at Sogesol. Using the Bad90MoB9 definition, the new model outperforms the EFL tool, with a K-S of 32.38% and an AUC of 0.70 compared to 19.70% and 0.62 for the psychometric tool when considering the validation sample. The results of the development sample are better showing a K-S of 32.59% and an AUC of 0.72. Contrary to the psychometric model, the new traditional model is shown to be reasonably calibrated, the HL test giving a p -value = 0.99 > 0.05. This result indicates that the model is a good fit. Instead of using a psychometric tool which remains a black box and under the control of EFL, Sogesol could improve its portfolio quality by implementing this new socio-demographic credit scoring model.

Unlike the assessments conducted by EFL or related consultants, this paper provides an independent evaluation which helps gauge the robustness of the psychometric model as developed and implemented at Sogesol. Nevertheless, we should highlight that the research presents some limitations that could impact the outcomes of the empirical analysis. The sample size of borrowers used by the paper to evaluate the predictive power of the psychometric tool may be considered too small (3671 clients). With a greater number of clients who took the test, results might have been improved. As well, the data collected (250 cases) to

assess the pilot of the EFL tool may be considered to be insufficient to enable the research to be conclusive, even though it could constitute a data point to further research. So, a critical mass of clients is important. Besides, the research was conducted on borrowers of only one microfinance institution, Sogesol. A larger number of institutions with a larger sample size might have affected results.

However, psychometric tests may be shown difficult to standardize, based on the cultural context. Questions related to certain personality traits, attitudes and behaviors, may be understood or interpreted differently depending on the cultural background. For instance, the way the Haitian micro entrepreneur understands honesty may be different from the Dominican micro entrepreneur's point of view. That being said, if the psychometric test developed by EFL was not well customized to the Haitian context and particularly to Sogesol's borrowers, the results might be biased and fail to discriminate between good borrowers and bad ones. It would be helpful to conduct a statistical analysis on the psychometric variables used in the tool implemented at Sogesol, in order to gauge their real predictive power. This could constitute the purpose of future studies.

Acknowledgements

This research was carried out with the valuable support of my Ph.D. thesis advisor, Dr. Joachim Bald. I would like then to express my sincere gratitude to him.

This paper would not have been possible without Sogesol that accepted to provide me with data necessary to conduct my research. Sogesol's Managing Director, Ms. Daphné Louissaint, has been very delighted to share with me documents and resources related to Sogesol's experiment with psychometric testing.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Abdou, H. A. et al. (2016). Predicting Credit Worthiness in Retail Banking with Limited Scoring Data. *Knowledge-Based Systems*, 103, 89-103.
<https://www.journals.elsevier.com/knowledge-based-systems>
<https://doi.org/10.1016/j.knosys.2016.03.023>
- Abdou, H., & Pointon, J. (2011). *Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature, Intelligent Systems in Accounting, Finance & Management*. Manchester: University of Salford. <https://doi.org/10.1002/isaf.325>
- Arráiz, I. et al. (2018). *Are Psychometric Tools a Viable Screening Method for Small and Medium Enterprise Lending? Evidence from Peru, Development through the Private Sector Series, TN No. 5, IDB/Invest*. <https://doi.org/10.1596/1813-9450-8276>
- Arráiz, I., Bruhn, R., & Stucchi, R. (2015). *Psychometrics as a Tool to Improve Screening and Access to Credit*. IDB Working Paper Series No. IDB-WP-625.
<https://doi.org/10.18235/0000199>
- Bernardin, H. J., & Cooke, K. D. (1993). Validity of an Honesty Test in Predicting Theft among Convenience Store Employees. *Academy of Management Journal*, 36,

- 1097-1108. <https://doi.org/10.2307/256647>
- BRH (2018). *Le secteur de la microfinance en Haïti, document d'information, MAE/BRH, DI-004.*
- Brier, W. G. (1950). Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78, 1-3.
[https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2)
- Brown, K., & Moles, P. (2014). *Credit Risk Management, Edinburgh Business School Herriot-Watt University, United Kingdom.*
- Costa Jr., P. T., & McCrae, R. R. (1992). Four Ways Five Factors Are Basic. *Personality and Individual Differences*, 13, 653-665. [https://doi.org/10.1016/0191-8869\(92\)90236-I](https://doi.org/10.1016/0191-8869(92)90236-I)
- Fenlon, C. et al. (2018). A Discussion of Calibration Techniques for Evaluating Binary and Categorical Predictive Models. *Preventive Veterinary Medicine*, 149, 107-114.
<https://doi.org/10.1016/j.prevetmed.2017.11.018>
- Garanin, D. A. et al. (2014). The Evaluation of Credit Scoring Models, Parameters Using Roc Curve Analysis. *World Applied Sciences Journal*, 30, 938-942.
- Inter-American Development Bank (2013). *Providing Credit to Latin America's "Missing Middle"*.
<https://www.iadb.org/en/news/webstories/2013-04-30/pychometric-testing-to-assess-s-me-creditworthiness%2C10437.html>
- Joseph, C. (2013). *Advanced Credit Risk Analysis and Management*. Croydon: Wiley.
<https://doi.org/10.1002/9781118604878>
- Kraus, A. (2014). *Recent Methods from Statistics and Machine Learning for Credit Scoring*. Munich: Faculty of Mathematics, Informatics and Statistics.
- Siddiqi, N. (2006). *Credit Risk Scorecards, Developing and Implementing Intelligent Credit Scoring*. Hoboken, NJ: John Wiley & Sons, Inc.
- Spearman, C. E. (1946). Theory of the General Factor. *British Journal of Psychology*, 36, 117-131. <https://doi.org/10.1111/j.2044-8295.1946.tb01114.x>
- United Nations (2015). Transforming Our World: The 2030 Agenda for Sustainable Development. In *the General Assembly*.
<https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>
- USAID (2018). *Census of the Microfinance Industry in Haiti*. Finance Inclusive Haiti.
- World Bank (2019). *Haiti World Bank Data*. <https://data.worldbank.org/country/HT>
- Wu, X. Z. (2008). *Credit Scoring Models Validation*. Amsterdam: Faculty of Science Korteweg-de Vries Institute for Mathematics, University of Amsterdam.