DEMS
Department of Economics, Management and Statistics

DEMS WORKING PAPER SERIES

Activating women cognitive abilities: Impact of a financial literacy pilot program in India

Lucia Dalla Pellegrina, Giorgio Di Maio,
Paolo Landoni and Beatrice Rama

No. 412 – May 2019

<

Department of Economics, Management and Statistics
University of Milano – Bicocca
Piazza Ateneo Nuovo 1 – 2016 Milan, Italy
http://dems.unimib.it/

Activating women cognitive abilities:
Impact of a financial literacy pilot program in India

Lucia DALLA PELLEGRINA [*]
Università degli Studi di Milano-Bicocca
Department of Economics, Management and Statistics (DEMS), Milano, Italy
lucia.dallapellegrina@unimib.it

Giorgio DI MAIO
Università degli Studi di Milano-Bicocca
Department of Economics, Management and Statistics (DEMS), Milano, Italy
giorgio.dimaio@unimib.it

Paolo LANDONI
Politecnico di Torino
Department of Management and Production Engineering (DIGEP), Torino, Italy
paolo.landoni@polito.it

Beatrice RAMA
Politecnico di Milano
Department of Management, Economics and Industrial Engineering (DIG), Milano, Italy
beatrice.rama@mail.polimi.it

[*] Corresponding author, lucia.dallapellegrina@unimib.it.
The authors would like to thank Dr. Sujit Kumar Brahmochary and Alim Sardar, Arvind Ravishankar, Mattia Bosisio, Francesca Conti, Ludovica Tonella, and Elena Diceglie for their contribution to the research project, and the participants to the CERSEM's 8th International Research Workshop in Microfinance for their comments on a previous version of the paper. The usual disclaimers apply.

Activating women cognitive abilities:
Impact of a financial literacy pilot program in India

Abstract

This study is based on a randomized control trial (RCT) aimed at understanding the effectiveness of a financial literacy pilot program conducted in 2014-2016 at the Institute for Indian Mother and Child, a non-profit microfinance institution based near Kolkata. Significant impacts are revealed in terms of improvements in saving accumulation and punctuality in the repayment of the loan instalments for borrowers belonging to the treated group, compared to the group of borrowers who did not participated to the program. In particular, positive contribution emerges from the evolution of both the cognitive skills and the level of financial knowledge developed by the beneficiaries during the training program. Estimates provide evidence that enhancing cognitive abilities turn out to be strongly beneficial in fostering the accumulation savings, whereas financial principles also had an (although weaker) impact on stimulating a more timely reimbursement of the instalments. We conclude that the financial literacy pilot program has significantly activated women cognitive abilities, giving them the opportunity to apply them both in the course and in their businesses.

Keywords

Microfinance; financial literacy; randomized control trial, repayment performance; savings.

JEL classification

I21, I24, I28, J13, J24.

Introduction

In a world where more than 2 billion people don't have any sort of bank account (Consultative Group to Assist the Poor, 2006), including 75% of those earning less than 2 US$ per day, policymakers are noticing that financial exclusion may represent a risk to political stability and prevents economic advancement. As a result, perdurable access to a sufficiently broad range of financial services (financial inclusion) is seen as a means to help people to take charge of their lives, with benefits that could go beyond the individuals' needs (income inequality drop and the increase of economic growth).

During the Nineties, microfinance programs experienced a significant economic upturn, which led to their rapid expansion and to the improvement of the lending methodologies and of the products offered. New NGOs and banks were founded, not only thanks to a greater efficiency and profitability achieved in the overall sector, but also because of improved customer retention over time.

Nonetheless, and perhaps due to excessive customers' dependence on credit, the sector's reputation has recently suffered a remarkable drop, as evidenced by the crisis that hit the Indian state of Andhra Pradesh in 2010. Increasing attention has thus been put on the limitations and lack of efficacy of some microcredit initiatives (Lascelles, Mendelson, & Rozas, 2014).

Microfinance, indeed, is blamed to be characterized by many unresolved issues. Still, both practitioners and academics struggle to understand how to increase its outreach, while contemporaneously guaranteeing microfinance institutions' (MFIs) self-sustainability (Köhn, 2013; Quayes, 2015). Furthermore, it is not yet clear what the effective use of loans by the borrowers is, and there is no full agreement about the impact of microcredit on women empowerment, job creation, children education, social harmony and health.

In this very fragile context, financial literacy has been found to play a fundamental role against the mixed impacts of the very many attempts to seek for improvements in the field. While, at the beginning, the interest in financial education was concentrated on high-income countries, now the attention has been extended to the poorer part of the world. This can be documented by the number of countries that have started financial literacy programs (Holzmann, 2010).

A common belief across the practitioners is that financial literacy is a key success factor for guiding individuals in making a correct use of money, helping them to understand how it should be allocated to the business and other investments aimed at increasing households' wellbeing. Such good practices would stimulate a virtuous cycle, as customers' improved repayment behaviour (Roslan & Karim, 2009) would eventually benefit the MFIs, allowing an easier achievement of sustainability and better standards of outreach.

The general purpose of this study is to evaluate the impact of a financial literacy program conducted between September, 2014, and August, 2016, at the Institute for Indian Mother and Child (IIMC), a non-governmental development organization headquartered near Kolkata, India, which provides microfinance services in an extremely poor context.[1]

The outcomes of the program are evaluated through a randomized control trial (RCT) to assess whether the training actually brought tangible improvements in the saving capability of the beneficiaries and in the punctuality of the repayments. A number of microcredit beneficiaries were randomly assigned to either treatment or control groups; only the treatment group received the training during compulsory weekly afternoon sessions.

A baseline survey was conducted before the intervention; a second evaluation survey took place at the end of the course, and a follow up evaluation survey after another 20 weeks. In addition, the financial cash-flow data of both the treatment and the control group were analyzed over a period of 99 weeks, before, during, and after the experiment.

Simple difference (SD) estimates suggest that the training program had a positive and significant impact on the beneficiaries in the form of improved saving accumulation and punctuality in the repayment of instalments.

Our study differs from previous RTCs aimed at directly evaluating the effects of educational programs on borrowers' repayment performance. In particular, we apply an Item Response methodology that allows an extremely accurate quantification of the specific educational skills, or abilities, developed by customers during the training, which encouraged them towards greater financial discipline.

In a recent paper, in fact, Heckman et al. (2013) observes that "A growing literature establishes that high quality early childhood interventions targeted toward disadvantaged children have substantial impacts on later life outcomes. Little is known about the mechanisms producing these impacts". In line with this approach, our methodology allows analyzing the mechanisms, or channels, through which the pilot program boosted participants' outcomes.

The regression analysis shows that experimentally induced changes in cognitive and financial skills actually explain a sizable portion of the treatment effects. On the one hand, cognitive abilities, including an enhanced attitude towards problem solving, have a strong significant role in fostering the accumulation savings. On the other hand, increasing the knowledge of financial principles has positive (although downsized) effects on both the incentives to save and to observe a more timely

[1] In 2016, the IIMC's microcredit program involved around 22,000 women living in 249 villages and gathered in 1,159 groups in seven branches (Institute for Indian Mother and Child, 2017).

reimbursement of the instalments. The improvement of the cognitive abilities seems to be the most important by-product of the financial literacy program. We interpret this effect as an activation of the women cognitive abilities. Treated women were likely to already possess these abilities, and the course seems to have given them the opportunity to activate these skills, applying them to their businesses. Consequently, by making the latter more productive and profitable women succeeded in fostering a timely repayment of loans and increasing the accumulation of their savings at IIMC.

There are important differences between this work and other studies that previously have managed the problem of explaining the mechanisms that make a program successful with respect to some kind of outcome (see mainly Heckman et al., 2013). First of all, we do not study the long-term effects on the individuals' social behavior. Rather, we focus more closely on those short and medium-term conducts of the borrowers that may have positive consequences in terms of increased sustainability of the MFIs.

Second, we have panel-type observations with a high-frequent temporal dimension. Hence, even if (as in Heckman et al., 2013) the subjects who participated in the experiment are limited in number, we observe them repeatedly over time, resulting in a considerable number of observations. This allowed us to exploit the individual time-invariant heterogeneities to eliminate the effects of possible initial unbalances between the treated and the control set due to the short individual dimension.

Third, we designed the project and the questionnaires for the specific purpose of this research. Therefore, we were not required to interpret any forms of latent skills. In fact, the questionnaire administered to IIMC's clients has been structured with separate sections, each containing questions suitable to measure each specific type of skill, previously defined by the authors. The outcome of this different approach is that while in Heckman et al. (2013) the measurable skills make a significant but limited contribution to the treatment effect compared to non-measurable skills and fixed effects, in this study measurable skills explain most of the treatment effect.

The article is organized as follows. In section 1 we provide literature backgrounds. In section 2 we outline the institutional context, the details of the intervention and the experimental design. The database is presented in section 3. Section 4 presents the estimation methods used to construct the cognitive and financial ability scores and to estimate the impact of their evolution on the outcome variables. The experiment-induced changes in individuals' skills and the analysis of the treatment effects is conducted in section 5. Finally, Section 6 concludes.

1. Literature background

Currently, MFIs are strongly recommended to offer Business Development Services (BDS) and technical assistance to loan beneficiaries in order to increase the odds of their projects' success and reduce default rates (Chowdhury & Mukhopadhaya, 2012; Ledgerwood, 1998). In fact, there is consolidated evidence suggesting that borrowing alone does not empower women (Alam, 2012; Holvoet, 2005). It is not by chance that the most successful microcredit programs are not confined to credit provision, but offer a group atmosphere and a safety net in which women can share information and set up networks. Combining financial and business training with microcredit is therefore considered crucial to this end, in a sort of "microfinance-plus" approach to poverty reduction (Bhatt & Tang, 2001; Bulte, Lensink, & Vu, 2016; Lensink, Mersland, Vu, & Zamore, 2018; Mahmud, 2003; Morduch, 2000). However, financial literacy programs have thus far been of extremely limited use due to MFIs' commitment to cost reduction.

Notwithstanding, the literature provides evidence that borrowers who are granted access to basic financial literacy can select profitable projects and/or generate more cash out of a specific activity (Godquin, 2004). It has also been observed that sustained success in microfinance can depend on participants' literacy and numeracy skills (Coppock, Desta, Tezera, & Gebru, 2011) and that better educated entrepreneurs have a greater ability to understand and analyze complex information, resulting in more effective business decisions (Baklouti, 2013; Bhatt & Tang, 2002).

Some studies have found little or no evidence of changes in borrowers' key-outcomes, such as business revenues, profits, or employment (Bruhn & Zia, 2013; Sayinzoga, Bulte, & Lensink, 2016), especially when considering female clients (Berge, Bjorvatn, & Tungodden, 2015; Giné & Mansuri, 2014). Nevertheless, the positive impact on business knowledge improvements and increased client retention rates for the MFIs providing formal training in business skills is widely assessed (Drexler, Fischer, & Schoar, 2014; Karlan & Valdivia, 2011; McKenzie & Woodruff, 2014). Loan repayment rates also appear to be positively related to training (Roslan & Karim, 2009).

In particular, enhancing women's financial management skills has become a priority for several microcredit providers (Afrin, Islam, & Ahmed, 2008). Since they are generally poorer and less educated, it is unquestionable that women (especially those living in rural areas) are more in need of financial literacy (Armendáriz de Aghion & Morduch, 2005; Mayoux, 2001) and are expected to better respond to it (e.g. Field, Jayachandran, & Pande, 2010). This may also help women in becoming more independent and increase their self-esteem and confidence. For example, Edgcomb (2002), Cook, Belliveau and Von Seggern (2001), and Dumas (2001) analyze MFIs offering integrated business development training, finding that the latter significantly improves microenterprise performance and generates microentrepreneurs' empowerment (Brau & Woller, 2004).

Microfinance clients perform a wide variety of business activities. For this reason, the financial training provided by MFIs usually focuses on improving general business skills rather than technical or sector-specific knowledge, as documented by McKenzie and Woodruff (2014). Some, however, claim that most training courses are indeed excessively general, which means that they frequently do not substantially improve the trainees' chances of success in conducting their own business (Bhatt & Tang, 2002). In some cases, borrowers even consider the training sessions to be a waste of time (Eversole, 2003).

On the contrary, an in line with our approach, Drexler, Fischer and Schoar (2014) find that a simple, rule-of-thumb based approach to training does better than a more formal training program, and improvements are significantly larger for micro-entrepreneurs who have lower skills or poor initial financial practices. Teaching basic skills, such as maintaining business records, and instilling rudimentary principles of computational strategy, appears as a necessary step for the development of "managerial capital", i.e., the organizational abilities required to manage an effective operational scale-up (Bruhn, Karlan, & Schoar, 2010; McKenzie & Woodruff, 2014).

In the next section we detail the type of intervention and describe the institutional context in which the experiment took place.

2. Intervention

2.1. Institutional context

The Institute for Indian Mother and Child (IIMC) is a non-government organization headquartered in the rural South-West districts of Kolkata, India. IIMC offers microfinance services, microcredit and micro savings collection, and social programs such as medical-care, a health promotion program, an education and child sponsorship program.

IIMC's microcredit program is dedicated to married (or widowed) women. To enter in the program, women must give proof of their permanent residence in their village and they also need their husbands' permission.

IIMC applies a group lending methodology with individual liability. Groups' formation process starts with IIMC credit managers visiting villages where potential groups can be created, mostly in rural areas. After the first contact, IIMC monitors the potential customers for a period of three months and provides basic training to teach how to start a business and how to behave in a group. This training is aimed at instilling the IIMC's philosophy, and is a good chance to understand the women's motivation and to establish forms of interrelation based on mutual trust. Each client is responsible for

the repayment of her loan only because there is no joint responsibility. However, group members, who usually live in the same village and carry out similar business activities, are invited to support each other and to exchange information and business experiences.

After the completion of the baseline training, IIMC's field officers select women and create groups of maximum 25 members. After three months of saving commitment, each woman can ask for a first loan of at most 3,000 rupees, provided that she has an amount of at least 1/10 of the loan size in her savings account. After the first loan is disbursed, savings are not mandatory anymore, although IIMC carefully monitors its customers' saving balance and encourages continuous saving accumulation. The annual interest rate paid by IIMC on the clients' saving balance is 4%.

The clients have to repay the loan within one year. They are expected to start the repayment after one month, with a total number of weekly installments equal to 44. Interests, equal to 10% of the loan amount, and are paid with the capital in the weekly installments.

If a client does not repay a loan in one year, IIMC gives her 3 months extra time, and only after that time the borrower will be considered a defaulter. A strong point of the IIMC microcredit system is the attitude towards its participants: if women are in trouble with the repayment of the loan, they are granted more time and support to go beyond the difficulty. If the problems persist, the consequence will be the impossibility to get a new loan, although the client does not receive any fines. If women prove to be reliable in repaying the first loans they can apply for greater loans. After 6-7 years they can reach the maximum loan amount that is set to 15,000 rupees.

Each credit group attends weekly meetings during which the clients repay the loan installments and deposit their savings to IIMC. Clients' financial transactions are registered on a personal register, called "Passbook for savings and loans" which is signed by the credit officer next to every transaction, and on a group register called "Group collection book". These documents are the official financial statements of the clients' transactions with IIMC that we used as a data collection support for the empirical analysis.

2.2. Experimental design

The experiment has been structured in four main building blocks: i) the conduction of a baseline survey and a first evaluation survey (S1 and Q1, week 27) followed by a pilot course of weekly lessons (weeks 28-39); ii) a second evaluation survey at the end of the course (Q2, week 40); iii) a follow-up third evaluation survey (Q3, week 60) aimed at verifying the persistence of the main outcomes; and iv) the collection of financial cash-flow data (weeks 1-99).

In order to better understand the advancement of the project, it is possible to refer to the timeline in Figure 1.

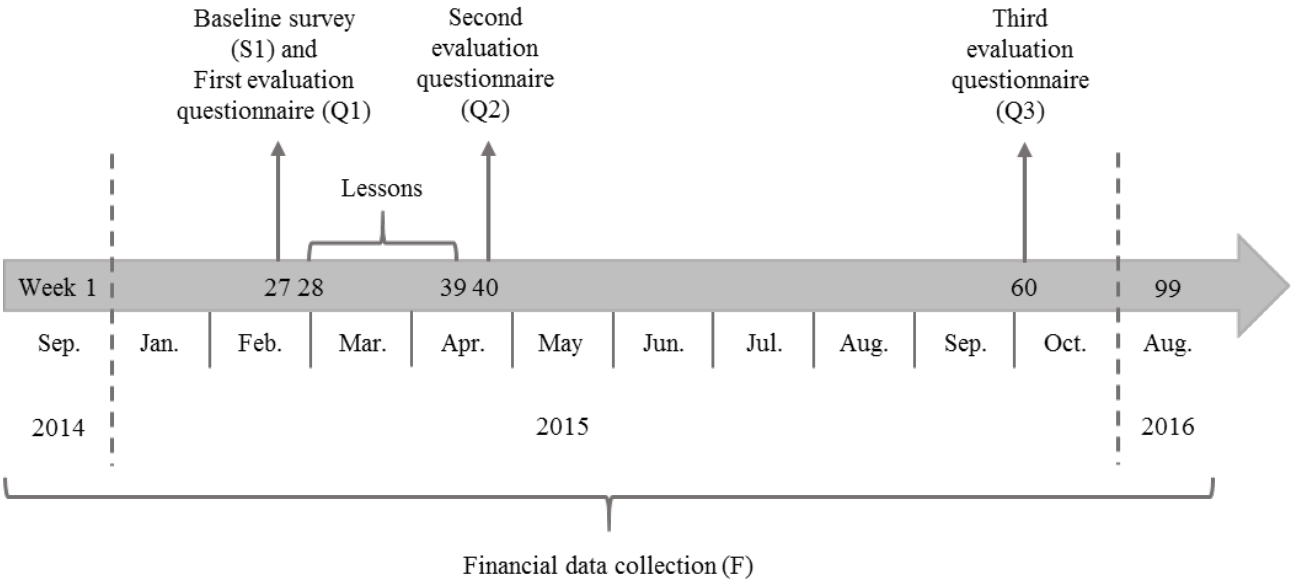

Figure 1 - Project Timeline

The RCT has been conducted in three branches of IIMC (Hogolkuria, Chakberia and Kalyanpur), where a group of 128 randomly selected women received weekly financial management training sessions during ad-hoc meetings. Another randomly selected group of 85 women did not receive any training, although they were required to complete both the baseline and the evaluation questionnaires.[2] Women belonging to the treated group were divided in three sub-groups in order to allow an adequate level of understanding in relatively small classes.

The financial literacy lessons covered topics such as basic accounting and profit calculation, how to separate money between business and household uses, understanding loan purposes and the relevance of savings accumulation, business problem solving, and business opportunities. The lessons were developed taking into account the peculiarities of the local context. In particular, applying a

[2] Our sample of 213 individuals is larger than the sample used by Heckman et al. (Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010b, 2010a) and Heckman et al. (Heckman et al., 2013) to evaluate the effects of the Perry Program (123 participants). Although this has led some researchers to question the validity and relevance of their results (Hanushek & Lindseth, 2009; Herrnstein & Murray, 1994), Heckman et al. (2010a) used an exact inference method that is valid in small samples, still obtaining statistically significant effects of the program, even after accounting for several tests for compromised randomization. In our case there are repeated observations for each individual in the sample; therefore the overall sample dimension is considerably higher (4,223 observations).

rule-of-thumb approach (Drexler, Fischer, and Schoar, 2014), trainers made use of posters, drawings, skits and folkloric tales to allow for full participation of the illiterate women[3].

As standard practice, the baseline survey was administered to collect personal data of the women involved in the experiment before the training took place. The first evaluation questionnaire conducted along with the baseline survey also had the scope of evaluating whether the initial levels of the cognitive skills of each participant to the experiment were balanced. Instead, we assumed that the initial knowledge regarding the financial management principles taught during the course was absent, therefore equal across the two groups.[4]

Questionnaires were initially translated in Bengali thanks to the support of the IIMC staff and volunteers. All surveys and questionnaires were distributed in paper form to the women reunited during ad hoc meetings. In cases of illiteracy, an IIMC's support worker was available to offer assistance, with the commitment not to enter into the merits of the answer nor to help women in any way, except for the translation. Questionnaires aimed at evaluating the cognitive abilities and financial knowledge of the participants were made of multiple choice questions. All the answers have been turned into binary responses, where the value of 1 corresponds to a correct answer and 0 to wrong/missing answers. Each of the meetings, as well as the lessons, lasted approximately two hours. Lessons were held in the afternoon in order not to divert too much the attention of the women from their own business activities. We estimated the initial proficiencies of the examinees and how both cognitive and financial skills have been evolving pursuant to the training.

Finally, we completed the experiment with the collection of the financial statements related to the loans taken by all the women participating to the experiment, repayment delays and overdue instalments, as well as the weekly savings. In particular, financial statements have been copied from paper group registers and digitalized afterwards. These data refer to weekly observations from September, 2014 to August, 2016, for a total of 99 weeks.[5]

[3] The lessons were given by an Indian student flanked by a Western student, both trained by the authors of this paper, and IIMC staff.

[4] Given the extremely low level of financial competence of the participants, we opted for not administering a first financial questionnaire as the rate of non-response or guessed responses would have been too high. As a proof of this, the staff who held the course stated that none of the women in the treated group had basic knowledge of profit calculation and money management principles. Being the two groups randomly selected, the assumption that the previous financial knowledge of treated and non-treated women were the same appears quite reasonable, in particular in light of the fact that the two groups were balanced in terms of initial cognitive abilities (see Section 3).

[5] In September 2016, IIMC adopted a new policy on savings collection that imposed a cap on the individual savings at INR 3,000. This new (and unexpected) policy, introduced to comply with a new government regulation, forced those

In principle, the empirical analysis is expected to reveal a significant divergence in the reimbursement rates between the two groups of women, in terms of punctuality in the repayments and saving capability. This outcome, however, could be challenged by the opportunity cost of attending the classes, since compulsory attendance may have implied subtracting time to the treated group' s own business in the training phase. For these reasons, it has been extremely important to monitor the pattern of the impact variables, both during the training and in the follow-up period.

3. Data

As aforementioned, the main data sources of the study are the financial statements of the beneficiaries (F), a baseline survey (S1), and three evaluation questionnaires (Q1, Q2, Q3). The latter were divided into two sections: the first included questions aimed at assessing the cognitive abilities of the women and their attitude toward problem-solving (Q1L, Q2L, Q3L), while the second section was intended to evaluate their knowledge of the financial practices which were taught during the course, such as profit calculation and separation of personal money from business revenues (Q2F, Q3F).

Relevant information from the baseline survey included women' name, group's name and number, age, educational achievement, village and microcredit branch,[6] marriage status,[7] number of children and total family members, working individuals within the household, current microcredit and Women Peace Council (WPC)[8] program membership, number of years spent in each program,[9] and the

clients with savings above the limit to withdraw the exceeding part and limited the accumulation of new savings by the clients below the maximum amount. For this reason, we limited our follow-up analysis to the financial data of August 2016.

[6] Besides individual (random) effects, in the regression analysis we opted for including fixed effects at the branch level, rather than at the village level. Unlike the information about the village in which the women live, the microcredit branch (or the potential branch for non-borrowers) is always known by the survey administrators, and is the one in which the women went to fill out the questionnaires. Instead, in the rural area where the experiment took place villages are often too small and not identifiable, as the village names indicated by the women cannot be traced back to any exact geographical location shown on maps.

[7] We did not use this information in the regression analysis, as almost all women were married.

[8] The WPC program aims to provide social, cultural and intellectual support to poor village mothers. A WPC is composed of borrowers from the microcredit program who meet weekly with compulsory attendance in sites located in several different villages to discuss issues regarding the community where they live in and to engage in activities targeted to solve them.

[9] Some WPC affiliates were not part of the program at the time of the survey but were members in the past. In the case of microcredit, they could be members of the program at the time of the survey but without having any outstanding loans.

number of other (not from IIMC) loans outstanding, including informal ones. As previously mentioned, F includes information on loan disbursement, repayments and (weekly) savings,[10] collected by the credit managers during the microcredit group meetings and subsequently digitalized by the authors.

Table 1 provides a description of the outcome variables considered for the evaluation of the impact of the intervention, along with the baseline characteristics of the respondents that will be used as covariates in the regression analysis.

Table 1 – Descriptive statistics – Baseline survey (S1)

Variable	Obs.	Mean	Std. Dev.	Min	Max
age (at the time of the S1 survey)	213	37.214	7.658	20	62
nchild (number of children)	213	1.934	1.075	0	7
fam_members (total family members, incl. respondent)	213	4.188	1.669	0	11
year_school (respondent's school attendance)	213	3.930	3.792	0	14
nfam_work (nr. household members with regular working activity)	213	1.216	0.895	0	5
mc (IIMC loan outstanding at the time of the S1 survey)	213	0.479	0.501	0	1
wpc (WPC member at the time of the S1 survey)	213	0.408	0.493	0	1
year_mc (years microcredit program membership at the time of the S1 survey)	213	5.185	4.246	0	12
year_wpc (years WPC program membership at the time of the S1 survey)	213	2.361	3.874	0	8
nloans (number of outstanding loans, incl. informal at the time of the S1 survey)	213	1.629	1.153	0	4
savings (weekly savings, avg. week 1-99)	213	97.2	552.3	0	2,025
delays (% delayed repayments, week 1-99)	102	0.078	0.270	0	1

On average, women are 37 years old, ranging from 20 to 62, and achieved almost 4 years of education. They live in households of about 4 members, of which two are children. In most of the

For some other interviewed, the borrowing activity may have covered only part of the observation period, as they may have had a loan at the moment of the survey but the last instalment was paid before the end of the observation period. For all these reasons the correlation between current participation and years spent in a program is lower than one may expect (see Table A1 in Appendix). Consequently, we decided to keep both program membership dummies and years of membership as covariates.

[10] The variable "savings" used in the empirical analysis is net of withdrawals. There are, indeed, very few withdrawals in the database.

families only one member has a regular working activity, reasonably the husband. Half of the interviewed have an outstanding IIMC micro-loan at and have been members of the microcredit program for more than 5 years. WPC membership is less frequent as it involves 40.8 per cent of the interviewed (2.4 years of participation).

Each woman has from one to two outstanding loans, besides the one obtained by IIMC. Considering the overall period of observation (pre and post-treatment), *savings* amount to 97 rupees per week, while repayment delays occur 8 per cent of the times. Savings are observed weekly, for the entire time period and for all women. Notice, instead, that the number of observations is lower for the variable *delays*, as only 102 over 213 women have an outstanding IIMC loan at the time of the survey.

Table 2 reports a breakdown of the baseline characteristics of the women belonging to either the treated or control groups at the time of the baseline (S1) survey, and the corresponding t-test for mean equality. The t-statistics confirm the random split of women between the treated and control group, as there is no (below 5 per cent) significant difference in means for any of the baseline characteristics. Randomness also results from the comparison of the average savings and delays at the date of the S1 survey (referred to the 27 weeks prior to S1).

Table 2 – t-Tests – Baseline survey (S1)

Variable	**Group**	**Obs.**	**Mean**	**Std. Err.**	**Std. Dev.**	**[95% Conf. Interval]**	
age	0	85	37.40611	0.927729	8.553236	35.56122	39.25100
	1	128	37.08618	0.621629	7.032933	35.85609	38.31627
						t = 0.2979	
nchild	0	85	1.8	0.105586	0.973457	1.590030	2.009970
	1	128	2.023438	0.100114	1.132657	1.825331	2.221544
						t = -1.4895	
fam_members	0	85	4.2	0.157715	1.454058	3.886367	4.513633
	1	128	4.179688	0.159367	1.803032	3.864329	4.495046
						t = 0.0868	
year_school	0	85	3.8	0.433473	3.996427	2.937991	4.662009
	1	128	4.015625	0.323761	3.662933	3.374961	4.656289
						t = -0.4056	
nfam_work	0	85	1.294118	0.097336	0.897397	1.100554	1.487682
	1	128	1.164063	0.079039	0.894228	1.007658	1.320467
						t = 1.0380	
loan_ip	0	85	0.4823529	0.0545205	0.502654	0.373933	0.5907729
	1	128	0.4765625	0.0443191	0.5014129	0.3888631	0.5642619
						t = 0.0825	
wpc	0	85	0.435294	0.054096	0.498738	.3277188	.5428694
	1	128	0.390625	0.043293	0.489808	.3049554	.4762946
						t = 0.6471	

Variable	Group	Obs.	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
year_mc	0	85	5.570588	0.475091	4.380118	4.625819	6.515358
	1	128	4.929688	0.367072	4.152949	4.203318	5.656057
						t = 1.0791	
year_wpc	0	85	2.529412	0.432110	3.983861	1.670113	3.388711
	1	128	2.25	0.336801	3.810470	1.583531	2.916469
						t = 0.5146	
nloans	0	85	1.576471	0.123416	1.137839	1.331044	1.821897
	1	128	1.664063	0.103028	1.165632	1.460188	1.867937
						t = -0.5422	
savings	0	85	81.76471	38.07139	351.0010	6.055572	157.4738
	1	128	107.5781	57.79617	653.8890	-6.790067	221.9463
						t = -0.3333	
delays	0	41	0.073170	0.041176	0.263652	-.010048	.1563895
	1	61	0.081967	0.035414	0.276591	.0111289	.1528055
						t = -0.1604	

All data refer to week 27 (S1). See Table 1 for a description of the variables' content.

4. Methodology

In this section, we estimate the effects of the treatment on the evolution of two latent individual traits pertaining to participants in the pilot project: cognitive skills and financial knowledge. We use the Item Response Technique (IRT) to create scores that summarize the skills acquired (see, among others, Baker 2001, Baker & Kim 2017, DeMars 2010, Fox 2010, Hambleton & Swaminathan 1985 and 1991, Lord 2008, Partchev 2004, Reise 2014, Reise & Revicki 2015, van der Linden 2016). Then we estimate the relationship between the previously estimated skills and the outcomes, i.e. changes in customers' repayment performance and ability to save.

The next section presents our model for the outcome estimation. Skill measurement strategy is detailed afterwards.

4.1. Outcome equation

We define y_{it} as the outcome measure for individual i at week t, and T_i as a binary variable taking the value of 1 if the individual receives the treatment and 0 otherwise. In particular, y_{it} refers to the repayment and saving performance of the borrowers.

Given the randomized design of experiment, single-difference (SD) estimation is performed. This choice is also supported by the outcome of the t-tests performed on both the pre-treatment observations of the dependent variable and on all the covariates (Table 2), which show that the differences in mean between the two groups are not statistically significant.

We estimate the following outcome equation:

$$y_{it} = \beta_0 + \beta_1 T_i + \beta_2 \theta_{it}^{ql} + \beta_3 \theta_{it}^{qf} + \beta_4 X_{it} + \beta_5 sb_{i0} + u_i + \varepsilon_{it} \quad (1)$$

where θ_{it}^{ql} and θ_{it}^{qf} are, respectively, a cognitive ability score and a score reflecting the knowledge of financial principles.[11] [12]

Model (1) is estimated using data collected from week 28 (beginning of the training program) onwards, including the balance of the saving account observed at the time of the baseline survey (sb_{i0}).[13] All variables are in natural logs.

Estimates are performed with and without (time invariant) covariates (X_{it}), also collected in the baseline survey. Notice that, given the presence of the (time variant) proficiency scores among the regressors, T_i only captures the residual effect of the treatment, net of the intermediated contribution occurring through an improvement of the cognitive and financial abilities.

Furthermore, given the relevant temporal characterization of our dataset, we account for the possible presence of women' specificities affecting the final outcome y_{it}. We initially assume that these effects are random (u_i) and not related to the treatment. In an alternative version of the model, we relax this hypothesis allowing for possible correlation between women's unmeasurable specificities and the treatment. Defining λ_i as the woman's fixed-effect, the alternatively estimated equation takes the following form:

$$y_{it} = \beta_0 + \beta_1 \theta_{it}^{ql} + \beta_2 \theta_{it}^{qf} + \lambda_i + \varepsilon_{it} \quad (2)$$

[11] We also perform regressions including an aggregate measure of ability (which will be simply denoted by θ_{it}), computed without distinguishing between questions aimed at extracting cognitive skills and questions aimed at extracting financial knowledge. However, a separate analysis of the two measures provides insightful information regarding whether cognitive and financial skills operate separately on women's repayment behavior or, instead, they tend to reinforce each other.

[12] About 10 per cent of ability scores are missing in the dataset, as some women were not present at the date of the evaluation questionnaire. In particular the missing values correspond to a specific branch because a flood occurred and women decided to go to save the harvest. Because of the random nature of this event, we perform the econometric analysis by relying on a missing at random (MAR) assumption. Multiple imputation techniques (see, for example, Little & Rubin, 2002) have been used to replace non-available observations.

[13] Another standard practice is to include the baseline value of the dependent variable. In our case, one could use the average number of delays and weekly savings in the pre-treatment period. We opted for using a less volatile measure that accounted for the possible gaps in the initial repayment capability, i.e. the stock of savings.

Notice that, due to collinearity, the treatment variable and other time invariant measures do not enter equation (2) as they cannot coexist with the lambdas. In this case, the treatment effects are accounted for by the presence of the lambdas. All regressions are performed using linear techniques,[14] and standard errors are always clustered at the women's level.

4.2. Measurement of cognitive skills and financial knowledge

We evaluate women's ability, considered as a latent trait (random variable) whose realization cannot be observed directly, but can be measured indirectly through the women's manifest performance in answering the quantitative questionnaires (Q1L, Q2L, Q3L, Q2F, Q3F), conditional on the characteristics of each proposed question. We apply the IRT methodology to estimate the respondents' abilities in two different dimensions: cognitive aptitudes and financial skills.

In general, IRT derives the probability of answering correctly to each question (item) as a function of latent traits. As opposite to broad criteria adopted to produce respondents' evaluation scores (which often consist of a mere sum of the number of correct answers) IRT techniques consider possible item characteristics that adjust both for the difficulty of each question and for its capacity to discriminate between respondents' talent on the recovered dimension(s).

In our context, items consist of multiple-choice questions where the responses can be either correct or incorrect. The output of the analysis is a numerical value representing the ability of each and every person that completes the test. Ability scores, denoted by θ, are constrained in the interval $[-2,+2]$.[15] The probability of answering a particular item k correctly is denoted by $P_k(\theta)$, while the probability of a wrongful response is equal to $1\text{-}P_k(\theta)$.

To estimate each individual's abilities, we adopt a two-parameter logistic model (2PL) (e.g. Lord, 1952; Nunnally & Bernstein, 1994; van der Linden, 2016; Wells & Faulkner-Bond, 2016)[16] defining the probability of a correct response as an exponential function as follows:

[14] When the outcomes are binary, as in the case of repayment delays, we estimate the effects using a Linear Probability Model (LPM). Textbook concerns about functional form misspecification for the LPM does not apply to impact estimation since treatment status is a binary variable (http://www.hhs.gov/ash/oah/oah-initiatives/assets/lpm-tabrief.pdf). See also Angrist and Pischke (2008).

[15] It is standard practice to constrain scores in a zero-centred range, so both positive and negative theta values are allowed.

[16] As opposite to the one-parameter response model (1PL), or Rasch model, the two-parameter model assumes that items do not discriminate equally between the respondents, and therefore that an increase in the respondents' ability is not associated with the same increase in the probability of success. Three-parameter models (3PL) also admit the possibility of guessing (DeMars, 2010; Fox, 2010). However, it was revealed by the women themselves that they left the questionnaire blank when they did not know the answer. For this reason we qualified blank answers as wrong and avoid models with guessing due to their higher computational expensiveness.

$$P_k(Y_i = 1|\theta_i, \beta_k, \alpha_k) = \frac{exp(\beta_{k(}\theta_i - \alpha_k))}{1+exp(\beta_{k(}\theta_i - \alpha_k))} \quad (3)$$

where *i* is the individual (*=1,2,…,N*) and θ_i his/her ability level[17]; *k* is the item (*=1,2,…,n*); α_k is the item difficulty parameter, and β_k is a discrimination parameter, which measures how effectively the item can distinguish between skilled examinees and less-proficient ones.

Following Partchev (Partchev, 2004) we simultaneously estimate the item parameters and the examinees' ability, in such a way that the ability scores are item-independent (see below). Item calibration and proficiency estimation are retrieved through an iterative process, where convergence is reached after a number of cycles.

We use Monte Carlo Markov Chain (MCMC) techniques to estimate the model.[18] We compute the respondents' ability scores and item parameters for each wave of the test, separating between financial and cognitive sections. Each model is run for 1,200,000 iterations, discarding the first 2,000 as burn-in. The thinning interval that we use in the simulations is 10. Gibbs sampling is adopted.
The fitness of the data to the model is evaluated recurring to three criteria: the calculation of the R^2, the test information function (TIF), and separation and reliability indexes (G and R, respectively). The first criterion, R^2, is used to reveal the relationship between the proficiency levels and the results of the tests. The individual results in each test are computed using both raw scores, i.e. just the count of right answers given (Wright & Stone, 1999b), and true scores, i.e. the sum of the probabilities of a correct response to each item, $TS_i = \sum_{k=1}^{n} P_k(\theta_i)$ (Baker & Kim, 2017). R^2, reported in Table 3, suggest that the model fits well the data, as no misalignments occur between the raw scores, the true scores and the levels of proficiency.

[17] In order to simplify the notation, in this section we do not distinguish between cognitive and financial skills, and also omit the time index. In this case one can interpret θ_i as the individual *i*'s ability score measured in a given round of the evaluation questionnaires.

[18] We use the MCMC pack for R to implement the estimation (Martin, Quinn, & Park, 2011, 2018). Bayesian methods applied to IRT guarantee less variable estimates than Maximum Likelihood criteria, especially in small samples and short tests (Natesan, Nandakumar, Minka, & Rubright, 2016).

Table 3 – IRT estimates

	Q1L (25 items)	Q2L (13 items)	Q2F (20 items)	Q3L (15 items)	Q3F (10 items)
Theta					
Mean	-0.031	-0.031	0.021	-0.025	-0.014
Median	-0.087	0.002	0.192	0.035	-0.066
Std. Dev.	0.877	0.865	0.874	0.785	0.809
Max	1.882	1.756	1.597	1.724	1.777
Min	-2.202	-2.000	-2.000	-2.000	-2.000
R^2					
R^2 Raw scores vs theta	0.9606	0.9420	0.9184	0.8772	0.9703
R^2 True scores vs theta	0.9947	0.9953	0.9918	0.9967	0.9842
Unidimensionality					
Critical Values	7	9	1	5	1
Total Values	2,662	1,092	1,729	1,705	1,085
%Critical Values	0,26%	0,82%	0,06%	0,29%	0,09%
Separation and Reliability indexes					
G2- theta	576.22	379.17	555.58	275.73	323.56
R - theta	99.83%	99.74%	99.82%	99.64%	99.69%
G2- alpha	466.53	150.42	146.62	141.53	110.17
R - alpha	99.79%	99.34%	99.32%	99.30%	99.10%
Number of items with negative slope removed					
	3	1	1	4	3

The second criterion used to assess data fitness is the test information function (TIF) that, assuming local independence, is the sum of the item information functions (IIF) of the questions of each test (Reise & Revicki, 2015).[19] The IIF describes how much ability information, or discrimination, an item provides at each level of the latent variable (Reise, 2014).[20] We computed the TIF about the three waves of the test. The corresponding test information curves (see Appendix) tend to assume a definite bell shape, rather tall and narrow, indicating that the items of the three waves are characterized by a relevant discriminating capability.

In relation to the information criterion, we consider the three standard hypotheses of applicability of IRT techniques: *monotonicity*, *local independence* and *unidimensionality*. Typically, an item

[19] Because of the local independence of the questions the item information functions demonstrate the features of addictiveness (Zięba, 2013).

[20] The analytical expression of the IIF of item k is $I_k(\theta) = \beta_k^2 * P_k(\theta) * (1 - P_k(\theta))$ and the TIF is given by $I(\theta) = \sum_{k=1}^{n} I_k(\theta)$ (Moghadamzadeh, Salehi, & Khodaie, 2011) The test estimates the ability θ with a precision inversely related to the information provided at that point, i.e. the standard error of the estimation is $SE(\theta) = 1/\sqrt{I(\theta)}$ (Hambleton et al., 1991).

characteristic curve (ICC)[21] has an increasing logistic trend, which means that the probability of giving a correct answer to a specific item increases in a monotonous way with the raise of the ability level of the individual. To validate the *monotonicity* assumption (Fox, 2010), it is standard practice to remove the answers to those items with negative slope of the ICC. In our specific case, a relatively low number of items and the associated answers were removed over all questionnaire waves (see Table 3).

Furthermore, when the test measures just one ability or trait (i.e., the scale is *unidimensional*), the parameters related to the item difficulty must be independent of who is answering the item itself, and individual abilities must be independent of what items the examinees answer (*local independence*). We exploit the *unidimensionality* condition of the scale,[22] in order to have *local independence* of the items.[23]

We use the package *psych* in R (Revelle, 2018) to implement an *analysis of eigenvalues* in order to check the above assumptions. In particular, we compute the difference between the observed answers (binary values) and the answers predicted by the model (probability of correct response). Noticeably, the standardized residuals are z-scores. If the model well-fits the data, all the residuals are equal to zero, or show negligible deviation from zero. An adequate compliance with this hypothesis is confirmed by the relatively low number of critical values of the standardized residuals, as reported in Table 3.

Then, a Principal Component Analysis (PCA) is applied to the matrix of the standard residuals, with the aim to check for the absence of other factors explaining the correlations between them. In fact, once the contribution of the factor that is expected to drive the answers (ability) is extracted, no other elements should remain. In this sense, it is required to check if the first eigenvalue is equal to 1, as well as the ratio between the first and the other eigenvalues. In our case, the assumptions of unidimensionality and local independence can be considered truthful (see scree plots in Appendix).

---

[21] The relationship between the probability of a correct response to an item k and the ability θ is described by the item characteristic curve (ICC), which has the expression $P_k(\theta) = P_k(Y = 1|\theta) = 1/(1 + exp(-\beta_k(\theta_i - \alpha_k)))$. α_k and β_k are also known as the location and slope of the ICC, respectively. The ICC is near zero at the lowest levels of ability and it increases until the probability of correct response reaches 1, in correspondence of the highest levels of ability. Each item has its own ICC.

[22] This means either separately estimating cognitive abilities and financial skills, or compute a single measure of ability without distinguishing among the two types of skills.

[23] These considerations can be expressed by the following equation, where Y is the dichotomous random variable (y=1 denotes a correct answer, while y=0 denotes an incorrect answer): $P(Y_1, Y_2, .., Y_n|\theta) = P(Y_1|\theta) * P(Y_2|\theta) \dots P(Y_n|\theta) = \prod_{i=1}^{n} P(Y_i|\theta)$.

Finally, as a third criterion to assess data fitness, we compute both the separation index G, which defines the dispersion of the sample of the items or examinees on the continuum scale, and the reliability index R, which is a measure of the trustworthiness of the model (Wright & Stone, 1999a).

We first calculated the sample standard deviation (SSD) of the respondents' ability ($SSD^2 = \sum_1^N \theta_i^2 /N - \sum_1^N (\theta_i /N)^2$) and of the items' difficulty ($SSD^2 = \sum_1^n \alpha_k^2 /n - \sum_1^n (\alpha_k /n)^2$) and the standard deviation of the error (SDE) of the respondents' ability $\left(SDE^2 = \sum_i^N SE_{\theta i}^2 /N\right)$ and of the items ($SDE^2 = \sum_i^n SE_{\alpha k}^2 /n$). Then, we derived the value of the corrected variance of the estimates, i.e. the expected standard deviation of the target sample, $SDT^2 = SSD^2 - SDE^2$. Finally, we computed the separation index *G=STD/SDE* and the reliability index $= G^2/(G^2 + 1)$. Obviously, good values of reliability approximate 100%. The two indexes (see Table 3) suggest that the model was suitable to represent the data collected in the research study. The computed G indexes, in fact, indicate that all waves of the test are enough sensitive to distinguish between high and low performers and that the sample is sufficiently wider to prove the item hierarchy. Also, the values obtained for the R indexes are close to 100%.

4.3. Evolution of cognitive skills and financial knowledge as a consequence of the treatment

As mentioned earlier, pre-treatment cognitive skills and financial knowledge have been tested before the beginning of the training program. As anticipated, we limited this step of the analysis to the evaluation of cognitive abilities, as the attendees had almost no basic financial knowledge prior to the treatment.

T-test have been performed to check whether the average scores in the treated and control groups were equal. The result in Table 4 show that for the cognitive ability scores there are no difference in means at the date of the evaluation survey conducted before the beginning of the training program (Q1).

Table 4 – t-test on mean equality: IRT estimates on baseline cognitive abilities (Q1L)

Variable	Group	Obs.	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
qlL	0	85	0,038059	0,197221	0,763834	-0,38494	0,4610562
(round 1)	1	128	-0,07278	0,102128	0,835956	-0,27668	0,1311271
						t=0.4710	

Total observations 2,107 (number of questions answered*number of respondents).

As consequence of the fact that the two groups were randomly selected, the test does not show discrepancies between the score means of the two groups. This should be sufficient to ensure that a possible divergence on the average scores of the two groups is attributable to the effect of the course.

To corroborate these considerations, and before undertaking the analysis on repayment and saving performances outlined in section 5.2, we perform a set of regressions aimed at testing the effects of the treatment on skill measures.

We proceed also in this case with an SD analysis where the effects of the treatment on the IRT scores are evaluated on both θ_{it}^{ql} and θ_{it}^{qf}.

$$\theta_{it}^{q\cdot} = \gamma_0 + \gamma_1 T_i + \gamma_2 X_{it} + \beta_3 y_{i0} + \upsilon_{it} \quad (4)$$

Results of this stage of the analysis are reported in in the next section before commenting the output of the main analysis, i.e. the effect of the evolution of knowledge and skills on savings and repayment delays.

Finally, to the purposes of the regression analysis, all scores are scaled to fit a range of positive values (by adding +2 to all the estimated θ_i) in order to avoid problems in the sign of the interaction terms that will be included in some augmented versions of equations (1) and (2).

5. Results

In this section we present the regression outcomes, starting from the analysis of the treatment effects on cognitive skills and financial knowledge. Then we will investigate how changes in the scores reflecting the evolution of these skills impact on the final outcomes of interest.

5.1. Treatment effects on cognitive skills and financial knowledge

As expected, there are important positive effects of the treatment on both θ_{it}^{ql} and θ_{it}^{qf}. The regression output corresponding to equation (4) is summarized in Table 5, where columns (a)-(c) report the treatment effects on cognitive abilities, θ_{it}^{ql}, while columns (d)-(f) show the estimates of treatment on financial knowledge, θ_{it}^{qf}. In the last columns, (g)-(i), the dependent variable is the interaction of the two previous scores. Furthermore, in columns (a), (d) and (g) we report the baseline results, whereas the estimates in columns (b), (e) and (h) include the covariates. Estimates in columns (c), (f) and (i) account for women's fixed effects.

Table 5 – Treatment effects on cognitive abilities and financial competences

	θ_{it}^{ql} (a)	θ_{it}^{ql} (b)	θ_{it}^{ql} (c)	θ_{it}^{qf} (d)	θ_{it}^{qf} (e)	θ_{it}^{qf} (f)	$\theta_{it}^{ql}\theta_{it}^{qf}$ (g)	$\theta_{it}^{ql}\theta_{it}^{qf}$ (h)	$\theta_{it}^{ql}\theta_{it}^{qf}$ (i)
treatment (T_i)	0.112**	0.119***	0.112**	0.033	0.023**	0.032	0.131	0.104***	0.130*
	(0.057)	(0.010)	(0.047)	(0.060)	(0.010)	(0.051)	(0.085)	(0.015)	(0.073)
_cons	1.088***	1.634***	1.087***	1.043***	0.855***	1.045***	1.125***	1.287***	1.128***
	(0.107)	(0.079)	(0.099)	(0.131)	(0.074)	(0.106)	(0.193)	(0.107)	(0.152)
Random (id)intercepts	yes	yes	no	yes	yes	no	yes	yes	no
id FE	no	no	yes	no	no	yes	no	no	yes
covariates	no	yes	no	no	yes	no	no	yes	no
N	4,223	4,223	4,223	4,223	4,223	4,223	4,223	4,223	4,223

(*) significant at 10% / (**) significant at 5% / (***) significant at 1%

The regression outcome in Table 5 provides evidence of a positive treatment effect on both types of skills, as highlighted by the sign associated with the parameter associated to the treatment variable, T_i. Furthermore, the greater capacity of the training program in developing cognitive-type skills is also evident, whereas the treatment improved financial skills to a lesser extent. Indeed, the magnitude of the parameter associated with T_i is greater in columns (a)-(c) compared to the corresponding parameters in columns (d)-(f). In the baseline estimates, significance is also higher for the parameter in column (a), whereas in column (d) the parameter is even not significant.

There could be several reasons for this result. A possible explanation is that the women belonging to the treated group had the opportunity to apply cognitive skills and the attitude towards problem-solving learnt during the course in the very short run, as these notions are reasonably more immediate and easier to grasp, especially through role plays and tales, which were part of the teaching tools. Financial skills, on the contrary, require time to be understood, studied, and applied before obtaining some feedback. Treated women were busy with the course attendance, and may have had less need and less time available to practice even basic financial principles, such as the calculation of profit.

Furthermore, in the rural area where the experiment was carried out, the economy is strongly based on agriculture and for some women the production cycle, the sale of products and the consequent calculation of profit is delayed several weeks after the end of the treatment. So, as long as women did not need to apply the financial principles learned during the course they may not have had incentives to make these principles their own knowledge.

5.2. The effects of changes in cognitive skills and financial knowledge on saving and repayment behavior

The estimated effect of changes in the scores measuring the skills accumulated by the attendees during the experiment are reported in Table 6 (estimates with random effects, without covariates), Table 7 (estimates with random effects, including covariates) and Table 8 (estimates with fixed effects, without covariates).

Table 6 – Treatment effects on saving accumulation and repayment delays - Random (id) intercepts Simple differences (SD) model without covariates

	savings (a)	**savings** (b)	**savings** (c)	**savings** (d)	**delays** (e)	**delays** (f)	**delays** (g)	**delays** (h)
treat	0.169	0.217	0.168	0.193	-0.067	-0.072	-0.071	-0.069
	(0.160)	(0.159)	(0.160)	(0.158)	(0.045)	(0.046)	(0.046)	(0.045)
θ_{it}^{ql}	0.610***		0.631***		-0.033		-0.008	
	(0.173)		(0.217)		(0.033)		(0.055)	
θ_{it}^{qf}		0.397*	0. 354			-0.043*	-0.038	
		(0.208)	(0.212)			(0.026)	(0.048)	
$\theta_{it}^{ql}\theta_{it}^{qf}$				0.322**				-0.025
				(0.128)				(0.021)
savingbalance _00	0.143	0.145	0.142	0.142	-0.002	-0.002	-0.002	-0.002
	(0.096)	(0.094)	(0.096)	(0.096)	(0.027)	(0.027)	(0.027)	(0.027)
_cons	1.114	1.290*	1.134	1.379*	0.262	0.279	0.281	0.258
	(0.712)	(0.729)	(0.722)	(0.719)	(0.201)	(0.200)	(0.200)	(0.200)
N	4,223	4,223	4,223	4,223	3,267	3,267	3,267	3,267

(*) significant at 10% / (**) significant at 5% / (***) significant at 1%

Table 7 – Treatment effects on saving accumulation and repayment delays - Random (id) intercepts Simple differences (SD) model with covariates

	savings (a)	**savings** (b)	**savings** (c)	**savings** (d)	**delays** (e)	**delays** (f)	**delays** (g)	**delays** (h)
treat	0.169	0.230*	0.167	0.202	-0.043	-0.049	-0.047	-0.046
	(0.122)	(0.128)	(0.123)	(0.123)	(0.030)	(0.031)	(0.031)	(0.030)
θ_{it}^{ql}	0.583***		0.610***		-0.037		-0.011	
	(0.173)		(0.217)		(0.033)		(0.053)	
θ_{it}^{qf}		0.368*	0.250			-0.046*	-0.039	
		(0.196)	(0.200)			(0.025)	(0.046)	
$\theta_{it}^{ql}\theta_{it}^{qf}$				0.309**				-0.027
				(0.124)				(0.021)
year_school	0.109	0.110*	0.110	0.105	-0.026**	-0.025*	-0.025*	-0.025*
	(0.067)	(0.066)	(0.067)	(0.067)	(0.013)	(0.013)	(0.013)	(0.013)
loan_yes	-0.200***	-0.209***	-0.201***	-0.205***				
	(0.068)	(0.068)	(0.068)	(0.067)				
nloans	0.277	0.164	0.280	0.218	-0.155***	-0.148***	-0.150***	-0.152***
	(0.245)	(0.245)	(0.247)	(0.242)	(0.057)	(0.057)	(0.057)	(0.056)
branch_1	-0.574***	-0.512***	-0.574***	-0.577***	-0.074	-0.074	-0.073	-0.071
	(0.149)	(0.149)	(0.150)	(0.153)	(0.046)	(0.046)	(0.046)	(0.046)
branch_3	0.604***	0.621***	0.600***	0.607***	-0.324***	-0.328***	-0.328***	-0.325***
	(0.138)	(0.137)	(0.139)	(0.136)	(0.051)	(0.051)	(0.051)	(0.051)
savingbalance _00	0.233***	0.248***	0.232***	0.237***	-0.016	-0.017	-0.017	-0.016
	(0.080)	(0.078)	(0.079)	(0.080)	(0.022)	(0.022)	(0.022)	(0.022)
_cons	1.079	1.727	1.071	1.670	0.358	0.325	0.339	0.323
	(1.211)	(1.238)	(1.211)	(1.220)	(0.325)	(0.323)	(0.329)	(0.322)
N	4,223	4,223	4,223	4,223	3,267	3,267	3,267	3,267

(*) significant at 10% / (**) significant at 5% / (***) significant at 1%. Other (non significant) covariates included: age, nchild fam_members, nfam_work, wpc, year_mc, year_wpc. Refer to table 1 for variable description.

Table 8 – Treatment effects on saving accumulation and repayment delays – Fixed-effects (id) Simple differences (SD) model without covariates

	savings (a)	**savings** (b)	**savings** (c)	**savings** (d)	**delays** (e)	**delays** (f)	**delays** (g)	**delays** (h)
θ_{it}^{ql}	0.635***		0.651***		-0.037		-0.007	
	(0.175)		(0.225)		(0.035)		(0.059)	
θ_{it}^{qf}		0.428**	0.250			-0.049*	-0.045	
		(0.206)	(0.220)			(0.027)	(0.052)	
$\theta_{it}^{ql}\theta_{it}^{qf}$				0.338**				-0.027
				(0.131)				(0.022)
_cons	2.189***	2.410***	2.199***	2.471***	0.218***	0.231***	0.233***	0.210***
	(0.185)	(0.216)	(0.203)	(0.151)	(0.036)	(0.028)	(0.032)	(0.025)
N	4,223	4,223	4,223	4,223	3,267	3,267	3,267	3,267

(*) significant at 10% / (**) significant at 5% / (***) significant at 1%.

Throughout all Tables 6 to 8 there is evidence of a strong positive effect of the cognitive skills on the ability to accumulate savings (columns (a) and (c) of each table).

In some regressions (e.g. column (b), Table 6) it turns out that improvements in the scores reflecting financial knowledge play a positive role on the ability to save, although the effect is much more downsized compared to the effect exerted by cognitive skills. Instead, knowledge of financial principles seems to have a greater influence in promoting the reduction of repayment delays (e.g. columns (f) and (g), Tables 6 and 7).

It is interesting to note how the contribution of the course, beyond the intermediate role played by the accumulation of skills measured by the two IRT scores, is rather modest. In fact, the parameter associated with the variable *treat*, which collects all those factors other than the increase in the cognitive skills and the financial knowledge that the course has helped to modify, is rarely significant. And where significant, it is limited to an increase of savings (column (b) Table 7).

There is also some interesting evidence regarding the contribution of the other factors influencing delay rates and the ability to save. These factors can be found by looking at Table 7, which reports the output of the random-effects model with covariates.

First, having an outstanding loan at IIMC reduces the accumulation of savings.[24] On the contrary, women who are part of the microcredit program but have no current loans are nevertheless encouraged to save money.

[24] Note that the variable indicating whether the woman has a loan in progress cannot be included among the covariates in the regression having the delays as a dependent variable since only women who have outstanding loans can record delays.

Second, the level of educational attainment is negatively correlated with delays, which is an expected result as education could also lead to more diligent behaviour.

Third, another expected outcome is related to the positive sign of the parameter associated with the initial level (at week 28) of the saving balance. Also this aspect indicates, presumably, that a greater initial level of savings shows a greater saving capacity, i.e. a sort of wealth effect.

Fourth, there is strong evidence of a reduction of delays associated with the presence of other outstanding loans, in addition to that provided by IIMC. One possible interpretation is the possible use of other sources, probably of informal nature, to repay the loan at IIMC.

Fifth, branch fixed effects are statistically significant. In this case, the customers of branch 3 seem to be more virtuous compared to the residual branch (branch 2), in fact they exhibit a systematic ability to accumulate more savings and record less delays in repaying their loans. The opposite occurs for the women belonging to branch 1. This should find possible explanations both in the different extent of the support offered by branch managers and in the various activities carried out by customers belonging to each branch. For example, a high density of Muslim merchants in one of the branches may partly explain such discrepancies.

Finally, to exclude the possibility that the financial and cognitive ability scores are endogenous in the equations describing customers' repayment and saving behaviour, we computed the correlations between the residuals and the scores. All correlations are not significant and lower than 5 per cent.

6. Conclusions

This study is based on a randomized control trial aimed at understanding the effectiveness of a financial literacy pilot program conducted in 2014-2016 at the Institute for Indian Mother and Child, a non-profit microfinance institution based in the rural South-West districts of Kolkata, India.

The data collection process consisted in the submission of a baseline questionnaire and three evaluation surveys. We also monitored the loan repayment performance of and savings of customers involved before and throughout all the period of the experiment (99 weeks). The first evaluation survey was administered at the beginning of the course, the second one at its end, and the third one six months after the end of the lessons.

We used these surveys to measure the development of the women cognitive abilities and financial skills by means of the IRT technique. We found that borrowers who attended the financial literacy course showed greater cognitive abilities than the women in the control group, both at the end of the program and in the follow-up evaluation survey.

The great impact of the training program on the cognitive skills of the women is very interesting. In fact, the financial literacy program was intended, as it is common in the literature, to provide the

women with basic financial knowledge and money management skills. We consider the improvement of the borrowers' cognitive ability as a relevant by-product of the training and we interpret this effect as an activation of the women cognitive abilities. We deem that the treated women already possessed these abilities, and that the financial literacy program gave them the opportunity to activate them. Applying these skills to their businesses and making the latter more productive and profitable the women achieved a more timely repayment of loans and an increased accumulation of savings.

The analysis of the treatment effects confirms this hypothesis. Indeed, we found significant impacts in terms of improvements in saving accumulation and punctuality in the repayment of the loan instalments for customers belonging to the treated group, compared to the group of women who did not participated to the program.

In particular, estimates provide evidence that enhancing cognitive abilities turns out beneficial in fostering the accumulation savings, whereas financial principles had an impact also on a more timely reimbursement of the instalments. However, the latter effects are considerably downsized compared to the stimulus exerted by cognitive skills on saving behaviour.

Summarizing, the policy message that derives from this study is extremely important from the point of view of the microfinance institutions, especially those operating in the poorest areas of the world. In fact, besides confirming the effectiveness of financial education programs that are not overly specific, the results of our analysis show that the most significant effects on repaying performance are fostered by the acquisition of cognitive skills that help customers to improve their general problem-solving skills. Therefore, we suggest that MFIs focus more on enhancing such kind of competences rather than pushing their clients to acquire more specific and sophisticated financial management skills.

In conclusion, although in our case the program has been subsidized, the cost of providing training to small groups of women is noticeable. As a consequence, the possibility of extending the program to the overall institution deserves careful attention. Bundling training modules to microfinance programs, and/or getting borrowers involved both in teaching and learning, could help obtaining more tangible improvements and save on organizational expenses.

7. References

Afrin, S., Islam, N., & Ahmed, S. U. (2008). A Multivariate Model of Micro Credit and Rural Women Entrepreneurship Development in Bangladesh. *International Journal of Business and Management, 3*(8), 169–185.

Alam, S. (2012). The Effect of Gender-Based Returns to Borrowing on Intra-Household Resource Allocation in Rural Bangladesh. *World Development*, *40*(6), 1164–1180.

Angrist, J. D., & Pischke, J. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Armendáriz de Aghion, B., & Morduch, J. (2005). *The Economics of Microfinance*. Cambridge, MA: The MIT Press.

Baker, F. B. (2001). *The Basics of Item Response Theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.

Baker, F. B., & Kim, S.-H. (2017). *The Basics of Item Response Theory Using R*. Cham: Springer International Publishing.

Baklouti, I. (2013). Determinants of Microcredit Repayment: The Case of Tunisian Microfinance Bank. *African Development Review*, *25*(3), 370–382.

Berge, L. I. O., Bjorvatn, K., & Tungodden, B. (2015). Human and Financial Capital for Microenterprise Development: Evidence from a Field and Lab Experiment. *Management Science*, *61*(4), 707–722. https://doi.org/10.1287/mnsc.2014.1933

Bhatt, N., & Tang, S.-Y. (2001). Making Microcredit Work in the United States: Social, Financial, and Administrative Dimensions. *Economic Development Quarterly*, *15*(3), 229–241.

Bhatt, N., & Tang, S.-Y. (2002). Determinants of Repayment in Microcredit: Evidence from Programs in the United States. *International Journal of Urban and Regional Research*, *26*(2), 360–376.

Brau, J. C., & Woller, G. M. (2004). Microfinance: A Comprehensive Review of the Existing Literature. *Journal of Entrepreneurial Finance and Business Ventures*, *9*(1), 1–26.

Bruhn, M., Karlan, D., & Schoar, A. (2010). What Capital is Missing in Developing Countries? *American Economic Review*, *100*(2), 629–633.

Bruhn, M., & Zia, B. (2013). Stimulating managerial capital in emerging markets: the impact of business training for young entrepreneurs. *Journal of Development Effectiveness*, *5*(2), 232–266.

Bulte, E., Lensink, R., & Vu, N. (2016). Gender training and female empowerment: Experimental evidence from Vietnam. *Economics Letters*, *145*, 117–119. https://doi.org/10.1016/j.econlet.2016.06.003

Chowdhury, T. A., & Mukhopadhaya, P. (2012). Assessment of multidimensional poverty and effectiveness of microfinance-driven government and NGO projects in the rural Bangladesh. *The Journal of Socio-Economics*, *41*(5), 500–512.

Consultative Group to Assist the Poor. (2006). *Good Practice Guidelines for Funders of Microfinance: Microfinance Consensus Guidelines* (2nd ed.). Washington, DC: Consultative

Group to Assist the Poor/The World Bank.

Cook, R. G., Belliveau, P., & Von Seggern, K. L. (2001). A case study of microenterprise training: Beta test findings and suggestions for improvement. *Journal of Developmental Entrepreneurship*, *6*, 255–267.

Coppock, D. L., Desta, S., Tezera, S., & Gebru, G. (2011). Capacity building helps pastoral women transform impoverished communities in Ethiopia. *Science*, *334*(6061), 1394–1398.

DeMars, C. (2010). *Item Response Theory*. Oxford: Oxford University Press.

Drexler, A., Fischer, G., & Schoar, A. (2014). Keeping It Simple: Financial Literacy and Rules of Thumb. *American Economic Journal: Applied Economics*, *6*(2), 1–31.

Dumas, C. (2001). Evaluating the outcomes of microenterprise training for low income women: A case study. *Journal of Developmental Entrepreneurship*, *6*, 97–129.

Edgcomb, E. L. (2002). What Makes for Effective Microenterprise Training? *Journal of Microfinance*, *4*(1), 99–114.

Eversole, R. (2003). Help, Risk and Deceit: Microentrepreneurs Talk about Microfinance. *Journal of International Development*, *15*(2), 179–188.

Field, E. M., Jayachandran, S., & Pande, R. (2010). Do Traditional Institutions Constrain Female Entrepreneurship? A Field Experiment on Business Training in India. *American Economic Review*, *100*(2), 125–129.

Fox, J.-P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. New York, NY: Springer Science+Business Media. https://doi.org/10.1037/027997

Giné, X., & Mansuri, G. (2014). Money or Ideas? A Field Experiment on Constraints to Entrepreneurship in Rural Pakistan. *IMF Policy Research Working Paper*, *6959*.

Godquin, M. (2004). Microfinance Repayment Performance in Bangladesh: How to Improve the Allocation of Loans by MFIs. *World Development*, *32*(11), 1909–1926.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. New York, NY: Springer Science+Business Media. https://doi.org/10.1007/978-94-017-1988-9

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. London: SAGE Publications.

Hanushek, E. A., & Lindseth, A. A. (2009). *Schoolhouses, courthouses, and statehouses: solving the funding-achievement puzzle in America's public schools*. Princeton, NJ: Princeton University Press.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010a). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics*, *1*(1), 1–46.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010b). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, *94*(1–2), 114–128. https://doi.org/10.1016/j.jpubeco.2009.11.001

Heckman, J. J., Pinto, R., & Savelyev, P. (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, *103*(6), 2052–2086.

Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York, NY: The Free Press.

Holvoet, N. (2005). The Impact of Microfinance on Decision-Making Agency: Evidence from South India. *Development and Change*, *36*(1), 75–102.

Holzmann, R. (2010). Bringing Financial Literacy and Education to Low and Middle Income Countries: The Need to Review, Adjust, and Extend Current Wisdom. *IZA Discussion Papers*, *5114*.

Institute for Indian Mother and Child. (2017). *IIMC Annual Report 2016-2017*. Kolkata: Emesjee Traders Stationers & Printers.

Karlan, D. S., & Valdivia, M. (2011). Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions. *Review of Economics and Statistics*, *93*(2), 510–527.

Köhn, D. (Ed.). (2013). *Microfinance 3.0: Reconciling Sustainability with Social Outreach and Responsible Delivery*. Heidelberg: Springer.

Lascelles, D., Mendelson, S., & Rozas, D. (2014). *Microfinance Banana Skins 2014: Facing Reality*. Tonbridge: Centre for the Study of Financial Innovation (CSFI).

Ledgerwood, J. (1998). *Microfinance Handbook: An Institutional Perspective*. Washington, DC: The International Bank for Reconstruction and Development - The World Bank.

Lensink, R., Mersland, R., Vu, N. T. H., & Zamore, S. (2018). Do microfinance institutions benefit from integrating financial and nonfinancial services? *Applied Economics*, *50*(21), 2386–2401. https://doi.org/10.1080/00036846.2017.1397852

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Lord, F. M. (1952). *A Theory of Test Scores*. Richmond, VA: Psychometric Corporation. https://doi.org/10.1039/tf9524800166

Lord, F. M. (2008). *Applications of Item Responsibility Theory to Practical Testing Problems*. New York, NY: Routledge.

Mahmud, S. (2003). Actually how Empowering is Microcredit? *Development and Change*, *34*(4), 577–605.

Martin, A. D., Quinn, K. M., & Park, J. H. (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software*, *42*(9), 1–21.

Martin, A. D., Quinn, K. M., & Park, J. H. (2018). *Markov Chain Monte Carlo (MCMC) Package* (1.4-3).

Mayoux, L. (2001). Tackling the Down Side: Social Capital, Women's Empowerment and Micro-Finance in Cameroon. *Development and Change*, *32*(3), 435–464.

McKenzie, D., & Woodruff, C. (2014). What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World? *The World Bank Research Observer*, *29*(1), 48–82.

Moghadamzadeh, A., Salehi, K., & Khodaie, E. (2011). A Comparison the Information Functions of the Item and Test on One, Two and Three Parametric Model of the Item Response Theory (IRT). *Procedia - Social and Behavioral Sciences*, *29*, 1359–1367.

Morduch, J. (2000). The Microfinance Schism. *World Development*, *28*(4), 617–629.

Natesan, P., Nandakumar, R., Minka, T., & Rubright, J. D. (2016). Bayesian Prior Choice in IRT Estimation Using MCMC and Variational Bayes. *Frontiers in Psychology*, *7*(9), 1–11. https://doi.org/10.3389/fpsyg.2016.01422

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw-Hill.

Partchev, I. (2004). A visual guide to item response theory. Retrieved from https://www.metheval.uni-jena.de/irt/VisualIRT.pdf

Patz, R. J., & Junker, B. W. (1999). Applications and Extensions of MCMC in IRT: Multiple Item Types, Missing Data, and Rated Responses. *Journal of Educational and Behaviorial Statistics*, *24*(4), 342–366.

Quayes, S. (2015). Outreach and performance of microfinance institutions: a panel analysis. *Applied Economics*, *47*(18), 1909–1925. https://doi.org/10.1080/00036846.2014.1002891

Reise, S. P. (2014). Item Response Theory. *The Encyclopedia of Clinical Psychology*, 1–10. https://doi.org/10.1002/9781118625392.wbecp357

Reise, S. P., & Revicki, D. A. (2015). *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*. (S. P. Reise & D. A. Revicki, Eds.). New York, NY: Routledge.

Revelle, W. (2018). *Procedures for Psychological, Psychometric, and Personality Research: Package 'psych'* (1.8.4).

Roslan, A. H., & Karim, M. Z. A. (2009). Determinants of Microcredit Repayment in Malaysia: The Case of Agrobank. *Humanity & Social Sciences Journal*, *4*(1), 45–52.

Sayinzoga, A., Bulte, E. H., & Lensink, R. (2016). Financial Literacy and Financial Behaviour: Experimental Evidence from Rural Rwanda. *Economic Journal*, *126*(594), 1571–1599. https://doi.org/10.1111/ecoj.12217

van der Linden, W. J. (Ed.). (2016). *Handbook of Item Response Theory, Volume One: Models*. Boca Raton, FL: Taylor & Francis Group.

Wells, C. S., & Faulkner-Bond, M. (Eds.). (2016). *Educational Measurement: From Foundations to Future*. New York, NY: The Guilford Press.

Wright, B. D., & Stone, M. H. (1999a). Reliability and Separation. In *Measurement Essentials* (2nd ed., pp. 157–165). Wilmington, DE: Wide Range.

Wright, B. D., & Stone, M. H. (1999b). The Idea of Measurement. In *Measurement Essentials* (2nd ed., pp. 1–5). Wilmington, DE: Wide Range.

Zięba, A. (2013). The item information function in one and two-parameter logistic models - A comparison and use in the analysis of the results of school tests. *Didactics of Mathematics*, *10*(14), 87–96.

APPENDIX

Annex 1 - Correlations

Table A1.1 – Correlations

	age	**nchild**	**fam_me~s**	**year_s~l**	**nfam_w~k**	**mc**	**wpc**	**year_mc**	**year_wpc**	**nloans**	**savings**
age	1										
nchild	0.2760*	1									
fam_members	0,0751	0.3513*	1								
year_school	-0,1617	-0,1643	-0,0382	1							
nfam_work	0,1401	0,1569	0.3861*	-0,083	1						
mc	0,0257	0,085	0,0725	-0,1436	0,0207	1					
wpc	-0,0861	0,0064	0,0554	0.2276*	-0,1581	-0,1465	1				
year_mc	0.2347*	0,0698	-0,0342	-0,0335	0,0365	**0,1577**	-0,0837	1			
year_wpc	0,0842	-0,0939	0,1398	0.2253*	0,01	0,01	**0.3523***	0.2164*	1		
nloans	-0,0713	0,0602	0,0781	-0,0794	0,0231	0.6687*	-0,1639	0.2015*	-0,0501	1	
savings	-0,0206	-0,0112	-0,0146	-0,0499	0,0072	-0,0067	-0,0146	0,0796	-0,033	0,0809	1
delays	0,1744	0,0976	-0,0113	-0,1701	0,0046		-0,1289	0.2650*	-0,1825	0,0289	-0,0332

All pwcorr on 213 obs. except delays (102 obs.)

* 1% sign level or better

Annex 2 - Item Characteristic Curves

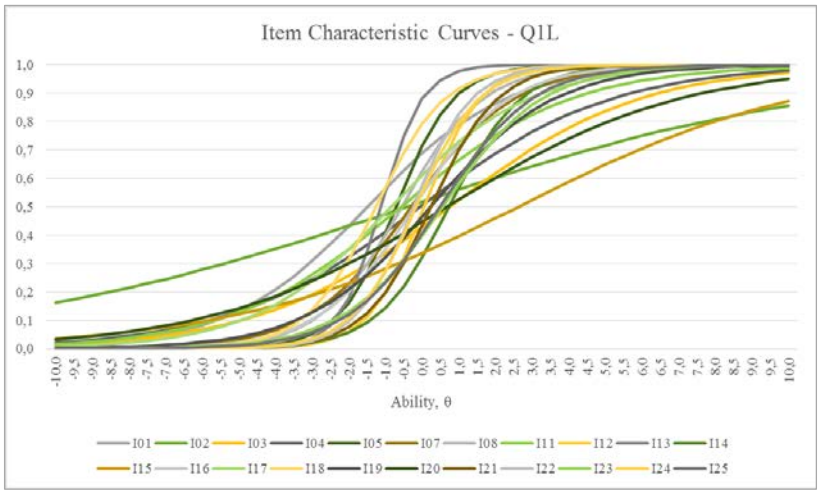

Figure A2.1 - Item Characteristic Curves – Q1

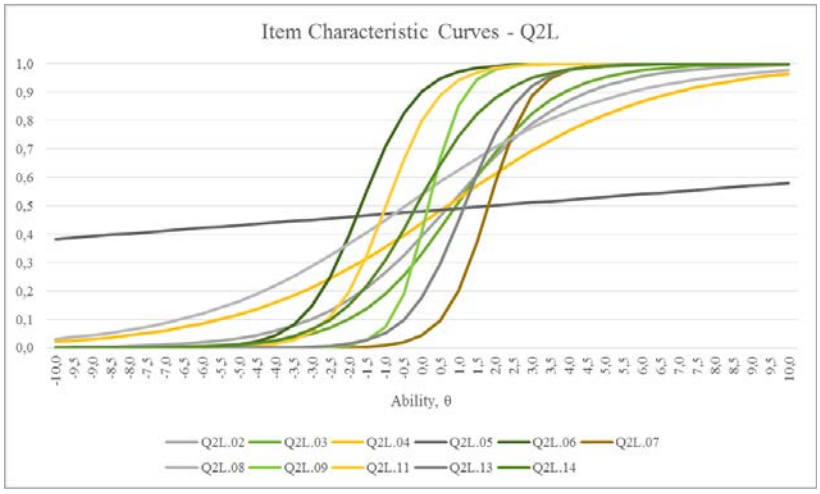

Figure A2.2 - Item Characteristic Curves – Q2L

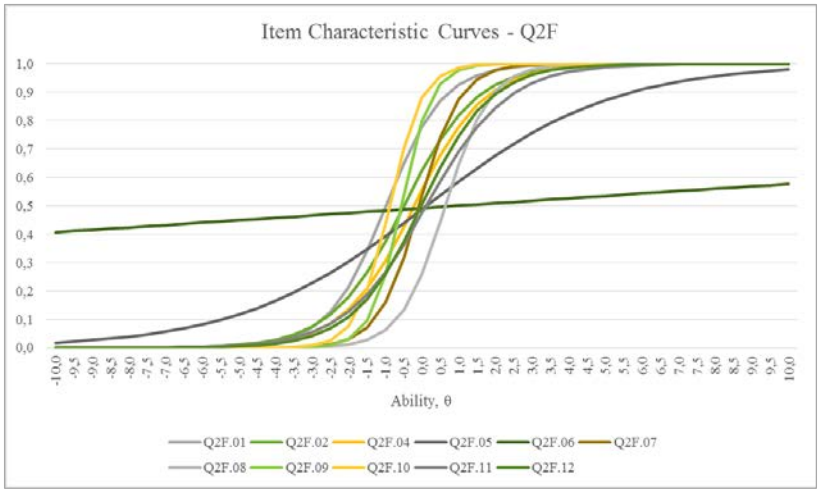

Figure A2.3 - Item Characteristic Curves – Q2F

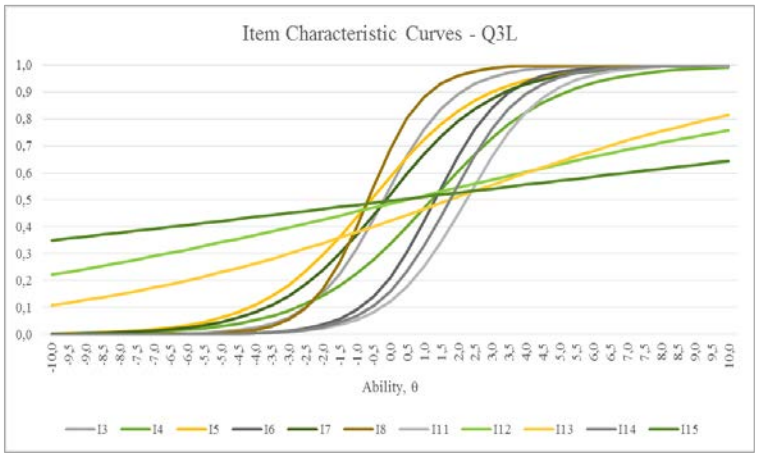

Figure A2.4 - Item Characteristic Curves – Q3L

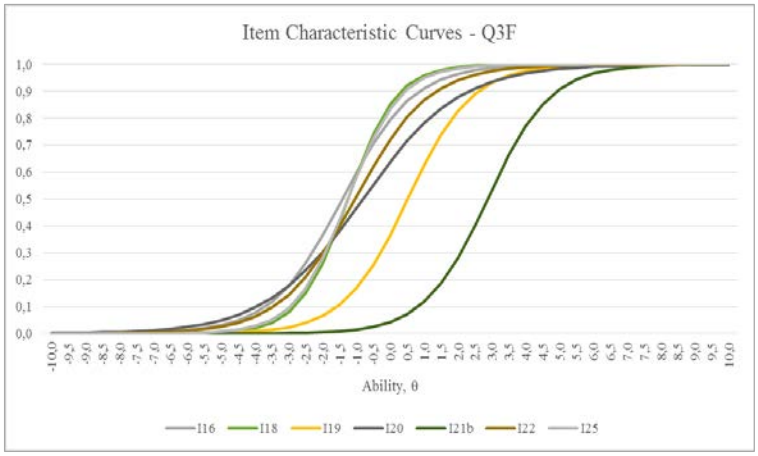

Figure A2.5 - Item Characteristic Curves – Q3F

Annex 3 - Test Information Functions

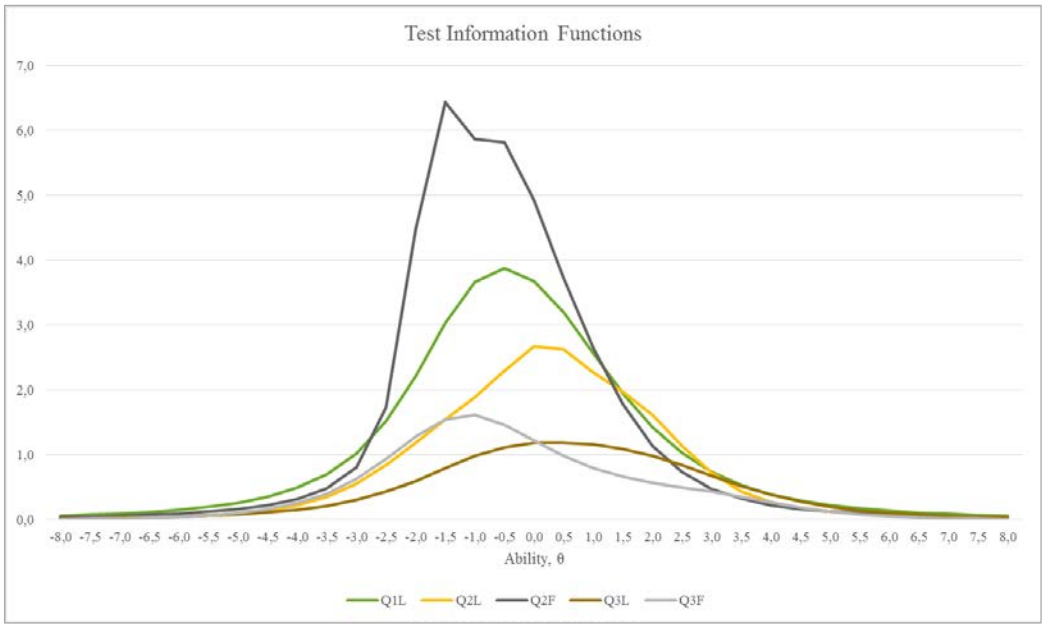

Figure A3.1 - Test Information Functions

Annex 4 - Scree Plots

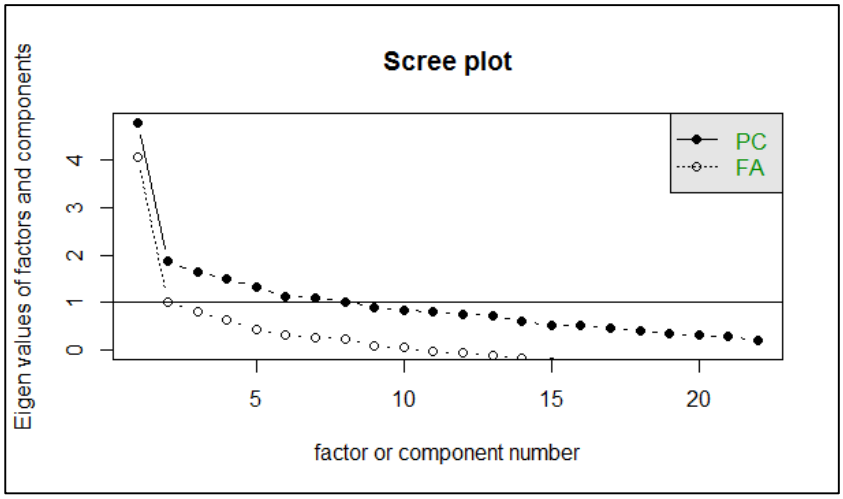

Figure A4.1 – Q1L Scree plot

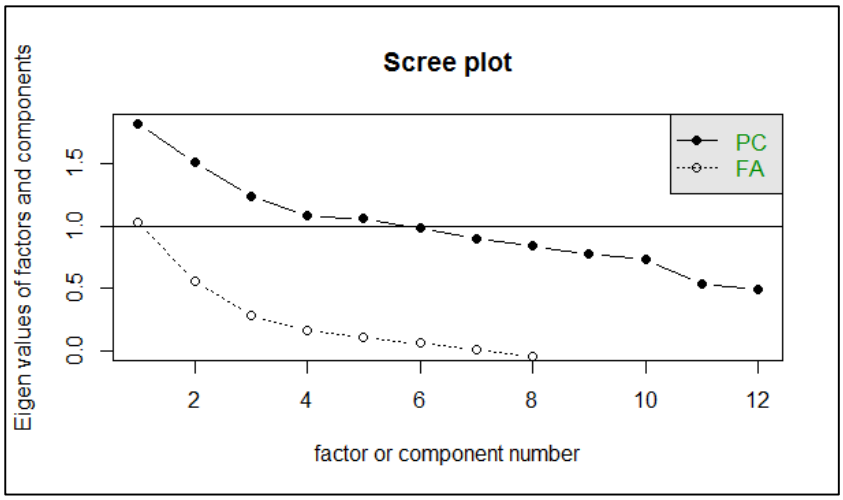

Figure A4.2 – Q2L Scree plot

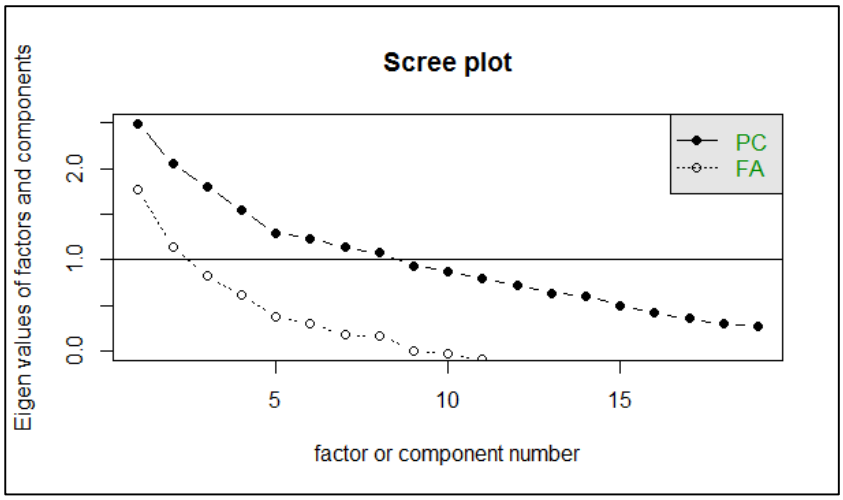

Figure A4.3 – Q2F Scree plot

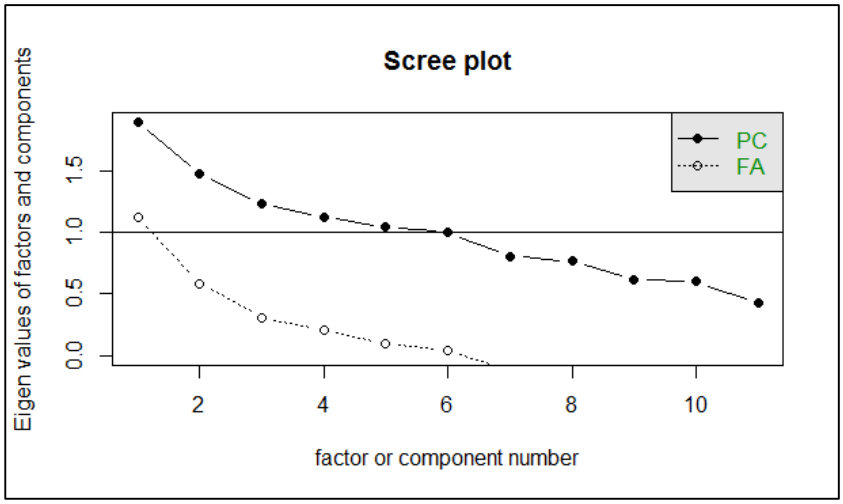

Figure A4.4 – Q3L Scree plot

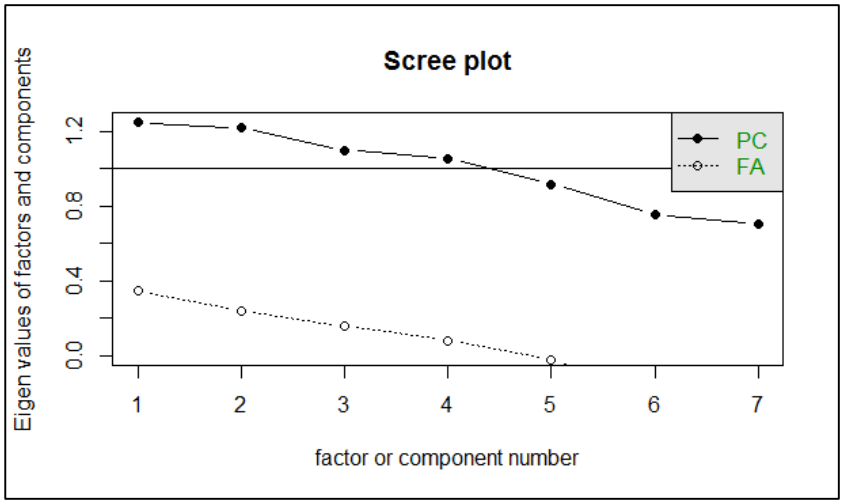

Figure A4.5 – Q3F Scree plot