Chapter 2

A Framework for Evaluating Financial Education Programmes

by
Annamaria Lusardi*

It is still relatively unusual for financial education programmes to be evaluated. In this chapter, the main benefits of programme evaluation are discussed, along with the challenges faced by evaluation designers and the resulting limitations of existing evaluations. A five tier evaluation framework is assessed as a potential solution to improve evaluation design whilst still allowing flexibility.

* Professor, Dartmouth College, USA

1. The Issue: The necessity of evaluating financial education programmes

Over the past thirty years, individuals have had to become increasingly responsible for their own financial well-being. The shift from defined benefit (DB) to defined contribution (DC) plans in many countries has meant that workers today have to decide both how much they need to save for retirement and how to allocate their pension wealth. Furthermore, financial instruments have become increasingly complex and individuals are presented with new and ever-more-sophisticated financial products. Access to credit is easier than ever before and opportunities to borrow are plentiful. Are individuals well equipped to make financial decisions? Unfortunately, many studies have documented that the majority of individuals lack the knowledge of basic financial concepts, such as interest compounding, inflation, and risk diversification; concepts that form the basis of financial decision-making. Moreover and most importantly, lack of knowledge has been found to be associated with lack of retirement planning, lower wealth accumulation, problems with debt, and poor investment choices (see Lusardi (2008, 2009) for a discussion of these issues).

Perhaps as evidence that financial illiteracy is considered a severe impediment to saving, governments, employers, and not-for-profit organisations have promoted financial education programmes. The effects of financial education programmes have not yet been precisely assessed. Several programmes provide some evidence of a general positive effect of financial education on behaviour, but the impact of specific programmes and teaching methods is still unclear. The question "what works best?" has not been clearly answered. The evaluation of these programmes is critically important.

Programme evaluation is crucial for three reasons: (1) to assess the magnitude of a programme's impacts on participants and the community/population as a whole; (2) to verify how resources and funds are spent; and (3) to ultimately improve the effectiveness of a programme. National governments and private financial education providers need to allocate resources efficiently: resources should go to the programmes that are most effective. They also need to fund programmes adequately and make sure that resources are allocated to the designated objectives. Finally, they need to find ways to improve upon existing programmes. In principle, evaluation is crucial for any type of financial education initiative, irrespective of the size of the programme. Without an evaluation, no programme can claim success. Consequently, proper evaluation should be one of the requirements for obtaining funding for both initial and repeated financial education initiatives.

Currently, not every programme performs an evaluation of its impacts, and when an evaluation is performed, different methodologies are often used. Thus, it is very hard to make consistent comparisons across programmes. Moreover, not all evaluations follow rigorous evaluation methods that allow investigators to pin down the effect due to a programme alone, rather than to other confounding factors. In order to identify the most effective ways to improve financial education, it is important to establish a rigorous evaluation methodology that can be applied to all programmes. As explained by Lyons et al. (2006): "The challenge is to create a tool that is flexible enough to meet the needs of a wide variety of individual programmes, yet standardised enough so that it can be used to make comparisons across programmes." The evaluation should be conducted following scientific conventions in order for results not to be dismissed or undervalued. Moreover, if all evaluations follow similar measurement methods, policy makers will be able to compare the results and gain insights on what is most effective. This evaluation system would give policy makers the opportunity to identify the best methods to make financial education successful and to effectively tailor programmes to specific audiences.

2. Existing Literature: Some difficulties in evaluating financial education programmes

So far, there are no commonly accepted definitions of financial education. The OECD definition of financial education provides a useful framework to refer to:

> Financial education is the process by which financial consumers/investors improve their understanding of financial products and concepts and, through information, instruction and/or objective advice, develop the skills and confidence to become more aware of financial risks and opportunities, to make informed choices, to know where to go for help, and to take other effective actions to improve their financial well-being. [1]

Evaluating financial education has proven to offer many challenges. Several scholars have examined the evaluation of financial education programmes. Reviewed below are several studies in the field, including Fox and Bartholomae (2008), Fox et al. (2005), Lusardi (2004, 2008), Lusardi and Mitchell (2007), Lyons et al. (2006), Hogarth (2006), O'Connell (2009), Crossan (2009), Cakebread (2009), and Collins and O'Rourke (2009). From these studies, the main challenges and limitations of the current state of evaluation studies are identified as follows:

- **Lack of a theoretical framework**. Some programmes target changes in behaviour that are not consistent with what prescriptive models of behaviour would predict. For example, a programme might simply target increases in savings. However, if the target population is young adults facing an upward sloping age-earning profile, economists would argue that they should be borrowing and not saving. Similarly, increasing saving may not be an optimal strategy if participants carry debt. In other words, the outcome of a programme needs to be in line with what is best for participants rather than ad-hoc indicators of behaviour.

- **Potential biases**. There are several potential biases in the evaluation process. The first is self-selection: programme participants are often not chosen randomly. Participants may choose to attend a programme because they are interested in improving their economic situation. In choosing to take advantage of a programme, they may be demonstrating motivation that other individuals do not share. Thus, the effects on this group may overestimate the effects the programme would have on other randomly selected groups of the population. Similarly, some groups may be the target of financial education programmes because of their behaviours, i.e., financial education could be *remedial* and offered to those who save the least or face financial problems. This leads to an underestimate of the effect of the education programme on random groups of the population. Another potential bias is attrition bias: participants may drop out during the programme or may not answer follow-up surveys, causing not only a loss of data but also loss of the representativeness of the sample. Low response rates to surveys conducted months after the programme is also a problem.

- **Measurement issues**. It is possible that financial education programmes may simply improve how participants report their assets and debt rather than have an effect on saving and debt behaviour. Moreover, because of data confidentiality or lack of access to administrative records, assessment is often based on surveys distributed to participants before and after financial education programmes. However, self-reported information may not always reflect actual behaviour or may measure behaviour with a lot of error.

- **Difficulty proving causation.** Because it is difficult to control for the many factors that affect behaviour, it is difficult to prove that programmes cause changes in behaviour. To do so, it is important to have a rich set of data. Due to the presence of many other factors that are difficult to control for in empirical works, many studies prove correlation rather than causation.

- **Lack of a comparison group**. Most studies do not include a comparison group; as a result, it can be hard to prove that the measured improvements are due to participation in the programme. Comparing programme participants to similar individuals who did not attend the programme provides strong evidence of the changes induced by the initiative. The presence of a comparison group therefore helps isolate the specific effects of the programme. An experimental or quasi-experimental design provides more reliable data than a "descriptive" experiment, in which the changes in participants are evaluated only through a pre-test and a post-test.

- **Lack of comparison of outcome with size of intervention**. Programmes rarely compare outcome with the size of the intervention. For example, some programmes assess the effect of one retirement seminar or the effect of sending participants to a benefits fair. Some of these interventions may simply be too small to generate any effects. If participants have very low financial literacy or face very large search and information costs, one retirement seminar may simply be insufficient to generate any effects. This is not because a financial education programme is ineffective, per se, but because the size and intensity of the programme is insufficient to generate a change in behaviour.

- **Difficulty assessing efficacy of different delivery methods** (e.g., lectures, brochures, and videos). Most financial education programmes combine two or more delivery methods, and few programmes are able to control for individual effects of the different methods. Therefore, it is very hard to determine outcomes of each method and to compare them with similar programmes that use a different number of delivery methods.

- **Practical hurdles**. Most prominent are the high cost of a thorough evaluation, the limited funding for evaluation programmes, and the lack of technical expertise.

- **Publication bias**. Evaluators and financial education advocates may not be willing to publish their studies if the results are unsuccessful. Certain non-experimental designs can allow evaluators to show better programme outcomes than do experimental techniques, consequently biasing evaluation toward that which appears most favourable.

- **Data are often not comparable**. Different studies use different methods, measures, indicators, and parameters, even if they ultimately evaluate the same thing. There is currently no agreement **on the most appropriate indicators,** outcomes, and measurable changes to use in evaluating financial literacy programmes.

Some initial Suggestions on how to improve the evaluation of financial education programmes

The difficulties mentioned above are serious and there has not been much agreement on how to address them. Some of these difficulties can be mitigated via experimental design, standardisation, as is explained below.

The importance of an experimental design

Evaluations following an experimental or quasi-experimental design, rather than merely being a "descriptive study,"[2] offer more reliable proofs of programme effects. An experimental design would use a fully randomised “treatment group” and a “control group.” A quasi-experimental design does not randomly select the participants of a programme, but uses a pre-selected group and tries to construct a “counterfactual group” that shares the features of the individuals attending the programme. The counterfactual group, which will not attend the programme, will undergo the same assessment processes and will resemble the treatment group as much as possible in characteristics such as age, sex, social class, and ethnicity. Assessment of the control group and counterfactual group help isolate the effects of the programme from other potential factors influencing the outcome(s), informing the evaluator as to whether the changes in the participants’ behaviour were *caused* by the programme or not. As Collins and O’Rourke (2009) remark, a strictly experimental design should be the “gold standard” for an evaluation study because it avoids self-selection bias. However, for practical reasons or budget constraints, it is often not feasible to follow this route. Quasi-experimental studies are easier to conduct and still provide reliable results. Financial education providers who cannot afford expensive experimental evaluations may consider turning to a quasi-experimental design.

The experimental or quasi-experimental design can address the problem of self-selection and the confounding effects resulting from external factors and better assess causality. However, it is important to note that an experimental design alone does not offer solutions for all of the difficulties mentioned above. Although experimental studies are able to control for differences between treatment and control groups, the outcomes of a financial education programme may still be hard to detect if the size of the intervention is not significant enough. For example, a study by Duflo and Saez (2003) undertook an ingenious randomised trial to assess the impact of attending a benefits fair on retirement savings for employees of a major U.S. university. They show that attendance of the fair had a rather modest effect on retirement savings. The study is a good example of a rigorously designed evaluation whose conclusions are hard to interpret: it is not clear whether the small impact on savings is due to the ineffectiveness of providing information and education to employees or to the small size of the intervention, i.e., a single benefits fairs may have minimal impact on behaviour in the face of widespread financial illiteracy and general lack of financial information.

Controlling for spillovers

An important issue when conducting randomised evaluations of large-scale financial education initiatives is the spillover of programme impacts from recipients of the programme to non-recipients, a point which has been highlighted in programme evaluation in the field of development economics. Duflo and Saez (2003) show that spillovers happen due to sharing of information and influence people’s motivation to increase their financial well-being. In their study, they observe a spillover effect in the form of information transfers between colleagues in the workplace. Spillovers may also influence the measurement of impacts in a larger-scale experiment, such as a financial education project in several communities located in the same geographical region. If the evaluation was conducted as a randomised field experiment in which a specific programme was implemented in some towns and not in others, it is very likely that the programme will find some effects (i.e., the effect will spill over) for the communities that were not “treated.” Positive spillover effects are beneficial overall as they help financial education providers indirectly reach a larger number of people. Spillover effects are a potential problem for evaluators, however, because they risk diminishment of the measured impacts of a programme: If the control group of a randomised-trial evaluation benefits from a programme attended by the treatment group, a comparison of the two groups will be less likely to show any impact. The evaluation conducted by Duflo and Saez (2003) was able to measure the degree of information transfer between employees of a U.S. university.

For programmes implemented on a larger scale (e.g., many schools or communities in an entire region), the field of development economics offers insights on how to get around these problems or to measure the spillover itself. Miguel and Kremer (2004) measure the spillover effect of deworming treatments in primary schools in Kenya by randomly selecting a third of the schools in a region. By comparing attendance rates in primary schools, they found that non-treated schools located closer to treated schools also had increased attendance rates after the administration of the treatment. Furthermore, by treating only a portion of the student body in a school, they showed that non-treated pupils in treated schools also became healthier, simply by interacting with treated children.

Programme design and evaluation design

The planning of a programme evaluation should begin with the planning of the financial education programme, should receive a budget consistent with its objectives, and should be tailored to the characteristics and the specific objectives of the initiative. Programme implementation and evaluation go hand in hand; the evaluation should be part of the programme itself and not done as an afterthought. Only in this way can pre-implementation, baseline data be collected.

Standardisation and flexibility

Financial education programmes are very diverse. They embrace a wide range of topics, objectives, audiences, and pedagogic methods. For this reason, evaluation studies should be flexible enough to be tailored to the different programmes while maintaining a determinate standard. Such standardised evaluations would allow implementers and policy makers to compare programmes and answer important policy questions:

- What and of what magnitude are the impacts of financial education programmes on the financial literacy level of the participants?
- What types of programmes are most effective? (school-based programmes, after-school programmes, workplace programmes, etc.).
- What delivery methods are most effective? (counselling, workshops, lectures, interactive exercises, etc.).
- Is financial education the most effective way to improve financial literacy? Are there other initiatives that can achieve better results?
- Which programmes should the government implement and which should it discourage?
- Given the specificity of each programme and the necessity for comparison of results, there is often a need for a compromise between flexibility and standardisation, perfect tailoring and homogeneity.

Privacy Issues and respect of confidentiality

One important concern shared by many programmes is the treatment of confidential data and respect for privacy. This is important to improve response rate, moderate attrition biases, and address some of the measurement problems discussed above. Evaluation questions may ask participants to report their annual incomes, divulge their income sources, or provide other information regarding their financial situation (e.g., whether they have a savings account and how much of their income they allocate to a pension fund). Participants may feel uncomfortable sharing this information with evaluators, particularly if they do not know who will be able to access the information they provide.

Fear of sharing confidential data may significantly decrease the response rate. For example, participants who are not citizens may be afraid to provide information regarding their sources of income, as they worry they could be reported to governmental authorities for irregular labour activities. At small retirement seminars at a single firm, employees may fear possible repercussions of sharing private financial information with their employers.

Evaluators should always respect confidentiality when conducting tests, focus groups, and follow-up surveys. Moreover, they should explain privacy policies to participants so those participants can complete the evaluation more confidently. There are several best practices to assure the respect of privacy of financial information. A primary confidentiality measure is to destroy all data, including contact information, in a timely and secure manner after the conclusion of the evaluation. In this way, participants can be sure that once they have filled out all necessary evaluation forms, including a possible follow-up survey via mail, their contact information will not be passed to anyone else. Another important practice is creation of a system of ID numbers connected to participants. The participants use ID numbers instead of names on every evaluation form they fill out, and the evaluator is not able to track a form to a specific person (see NEFE online evaluation toolkit by Lyons, Jayaratne, and Palmer for more information on the ID number practice).

Some ways to facilitate the evaluation of financial education programmes

A national Benchmark and a national strategy for financial education

Establishing a benchmark for the financial literacy level of a country can be crucial to facilitating financial education programmes and their evaluation; a national survey measuring financial literacy and financial behaviour is useful for the design and evaluation of both large-scale and small-scale financial education initiatives. A national survey can provide key insights into the state of financial knowledge and the demographic groups that are most lacking in that knowledge. Financial educations providers can then use this information to tailor programmes to the needs and characteristics of the targeted population. For evaluation purposes, a national survey can establish a baseline financial literacy level that can be used as a yardstick in assessing the effectiveness of a financial education programme. The design and evaluation of financial education programmes of any size, from local seminars to large public policies, can greatly benefit from the existence of a national benchmark.

A national strategy for financial literacy could also provide critical help for evaluation as follows:

- identify the main areas where intervention is needed (e.g., access to credit, default in loan payments, lack of information);
- determine the at-risk populations (e.g., young people, single parents, senior citizens);
- give general guidelines for smaller financial education providers; and
- identify the organisation that can assume a leading role in the financial education effort in the country, coordinating the work of financial education providers, and minimising the overlaps.

Agreement on Indicators and establishment of an international benchmark

As Crossan (2009) points out, there currently are "no proven or agreed indicators or measures for financial literacy." Such an observation calls for the need to create a standardised, possibly international, indicator of financial literacy. This would have several advantages:

- It would allow the impacts of one programme on financial literacy to be compared to those of another programme nationally and/or internationally.

- It would allow for the establishment of a national benchmark, an average financial literacy score for each country, utilising the same measures that evaluate financial education programmes, so that:

  a) nations could compare their financial literacy levels with those of other countries;

  b) each programme could compare its outcomes with the national benchmark score; and

  c) national programmes could track their impacts over the years.

The establishment of a common measure of financial literacy and a national benchmark facilitate not only the evaluation of financial education programmes but also the comparison of the effects of these programmes across countries.

3. A five-tier evaluation framework

To address the limitations of current evaluation studies, several scholars have proposed the creation of an evaluation framework that could serve as a guideline for all programme implementers and evaluators. Such a framework could help the implementers design the evaluation in a standardised way while tailoring it to their specific programme. In *Evaluating the effectiveness of financial education programmes* (2009), O'Connell proposed a new version of a five-tier framework that was first introduced by Jacobs (1988) and later elaborated upon by Fox and Bartholomae (2008).[3]

O'Connell's five-tier approach has many advantages and constitutes an important step toward a standardised but flexible scientific evaluation. In fact, it allows researchers and evaluators to follow an experimental design if desired, to tailor the evaluation to specific audiences and objectives, and to maintain a large degree of similarity with the evaluations of related programmes. The framework is rather broad and provides overall a general direction toward which evaluators can work. More discussion about this approach is reported below.

O'Connell's denomination of the five tiers is different than proposed by others; Le Brun (2009) proposes a more conventional denomination used by scholars in the field of evaluations: the so-called Traditional Model Approach. Despite the new denomination, the content and function of the tiers remain the same. Table 1 compares the two denominations.

Table 2.1 : The two denominations of the five-tiered evaluation framework

O'Connell denomination		Traditional Model
Needs		Programme Objectives
Accountability		Programme Inputs
Fine-Tuning		Programme Delivery
Micro Impacts		Programme Outcomes
Macro Impact		Programme Impact

As previously mentioned, programme evaluation should go hand in hand with programme design and implementation. As Fox and Bartholomae (2008) point out, the five-tiered approach "encourages evaluation to occur in each stage of programming."

There are three overarching themes of the evaluation that cut across the tiers. As also discussed by Hogarth (2006), these themes are: 1) objectives, 2) audience, and 3) available resources. They are crucial not only for the design of the evaluation, but also for the planning of the programme itself.

Objectives: The questions to answer are: What are the objectives of the programme? What aspects of financial literacy is the programme trying to improve? This theme is not only part of the Programme Objectives tier—where the objectives are defined—but affects all the other tiers as well. In fact, the implementer should always consider the programme objectives when identifying the input (what type of class, how many hours, what budget, how many teachers, etc.), the delivery system (what type of teaching system best suits the objectives: lecturing, interactive lecture, activities, workshops, online course, personal counselling, etc.), and the outcomes (based on the objectives that were identified, which outcomes should be expected: e.g., better control of financial transactions and bookkeeping if the objective of the programme is increasing responsibility for and care of household finances).

Audience: The questions to answer are: What type of audience is the programme targeting and who are the programme participants? These issues are first addressed in the Programme Objectives tier, as the objectives may be strictly connected to a specific group of people, but they should then be considered in all the other tiers. When designing the programme, the implementers should ask what types of input and delivery systems best suit the audience (e.g., a primary school programme would be more likely to reach a large number of students, with several hours of classes distributed throughout the year, and use interactive exercises and activities to help children learn; on the other hand, a retirement workshop for a firm's employees might aim at two or three meetings in the same week, with a restricted number of participants, and have a more straightforward, lecture-type delivery method). The outcomes and impacts will also be measured in different ways depending on the audience. For example, high school students might receive surveys that assess critical thinking and direct application of the knowledge acquired, while adult employees might receive a straightforward survey assessing knowledge, attitudes, and behaviour.

Available resources: Evaluations are costly. Some methods are more expensive than others; given the budget constraints of most programmes, not all methods will be affordable. Evaluators should carefully consider existing resources and budget constraints. Thus, objectives, inputs, and delivery methods, for example, have to be chosen in keeping with the resources that are available for the programme and the evaluation.

4. Implementation of the five-tier framework

O'Connell's five-tier approach is a sound basis on which to create a more detailed framework for the evaluation of financial education programmes. The tiers give good direction for evaluators, but more specific guidelines and information on tools, measurement methods, and indicators are needed. This chapter provides further implementation of the five-tier framework to establish a more detailed guideline for future financial education evaluations. Each tier is analysed, giving a quick definition of its objectives, and possible options are provided of what could be measured in each section and how it could be measured. For the last two tiers, where evaluation could take very different paths depending on the type of programme, some of the limitations of the potential methods are analysed. Le Brun's (2009) denomination is used but also O'Connell's (2009) tiers is included in Table 2.

Table 2.2 : The implementation of the five-tiered framework

O'Connell Denomination	Traditional Approach Model	What to Measure	How to Measure It
Need	Programme Objectives	Financial literacy scores Levels of bankruptcy High debt Low savings	Financial literacy survey to target group Existing financial literacy results Publicly published data on specific issues
Accountability	Programme Inputs	Costs of the programme Duration of the programme Size and characteristics of the target group Teaching methods/delivery	Description of teaching methods, etc. Survey at the start or end of programme Data on programme expenses
Fine-Tuning	Programme Delivery	What in the programme was effective What was not effective What can be improved What should be changed or completely eliminated	Post-survey to participants and teachers Focus groups One-on-one interview
Micro-Impact	Programme Outcomes	Knowledge Outcome	Pre- and post-test Alternative assessment Criterion reference group test Follow-up survey
		Attitude Outcome	Pre- and post-test Alternative assessment Follow-up survey Focus groups – One-on-one interview
		Behaviour Outcome	Follow-up survey Alternative assessment Focus Groups – one-on-one interview Administrative records
		Other Outcome	Follow-up survey
Macro-Impact	Programme Impact	Financial literacy scores	National survey conducted regularly
		Other indicators of financial well-being: i.e., income level, savings level, indebtedness.	National survey conducted regularly

5. Analysis of the five tiers

Programme objectives

Programme Objectives is the tier that deals with the purpose of the programme and identifies its aims. The objectives are normally addressing one or more problems related to financial literacy, which could range from general lack of financial literacy to a more specific issue such as high number of defaults on mortgage payments. The evaluator should verify the existence and relevance of such issues before the programme is implemented. O'Connell called this tier "Needs" because it identifies what should be improved in the community. In order to identify needs and objectives, the evaluator should also identify the target audience (e.g., high school students, retiring workers, low-income families).

Below are some of the possible measures to verify needs and set objectives that Fox and Bartholomae (2008) provide:

- Testing financial literacy levels in the target population in the case of a general financial education programme. The implementer can take advantage of pre-published data if they exist; otherwise financial literacy tests can be administered among a random sample of the target group.

- High rates of small-business bankruptcy, high levels of consumer debt, and low savings can be used as indicators of poor financial literacy and of poor financial management. To evaluate needs using these indicators, the implementer should find statistics and data on the target community.

- More specific data can be used for particular programmes, such as high levels of default in the repayment of mortgages for a programme on financial management for families that are planning to buy a house. Publicly available data or new survey data could be used for such issues.

Programme inputs

This tier analyses the "inputs" of the programme. It deals with the collection of information regarding the education programme itself and the service provided: its costs, its length, what it teaches, and who participates (number of participants and particular characteristics such as ethnicity, education level, and income level). As Fox and Bartholomae (2008) discuss, the goal of this tier is to assess whether the target group is affected by the programme and in what way. Precise information about the target group is also important to create a well-tailored control group.

Information on inputs can be gathered with a survey given during programme registration, at the end of the programme, or at another appropriate time.

Any programme should pay attention to the inputs both for cost considerations and for accountability.

Programme delivery

The programme delivery tier assesses the way the programme was implemented, whether and what was effective, and what should be changed to improve it. One effective way to receive feedback on programme delivery is by directly asking both participants and instructors. This can be done through specific surveys at the end of the programme. Students and instructors should be provided with different surveys, as their observations on the programme come from different perspectives. The survey could ask participants and instructors to rate the effectiveness of specific teaching methods or course material as well as ask open-ended questions (e.g., what was the most/least effective part of the course?). While participants could be asked how the programme affected them, the instructors could be asked what improvements they saw in participants. Other possible evaluation tools in this tier are focus groups and in-depth interviews with participants or instructors. These two methods can provide valuable qualitative data to better put into context the results of the quantitative data collected through the survey. They might substitute for open-ended written questions and allow for additional comments.

Programme outcomes

All direct effects of a programme on participants are considered programme outcomes. Outcomes can generally be divided into subsets: knowledge outcomes, attitude outcomes, behaviour outcomes, and practical outcomes. Each subset assesses different aspects of the overall effect of the programme and is measured with different, although often overlapping, methods.

Knowledge outcomes

Knowledge outcomes measure changes in programme participants' knowledge of specific financial concepts (e.g., how compound interest works) or programme-related concepts (e.g., how a firm's retirement plan works).

There are several methods through which to assess knowledge outcomes:

- **Pre- and post-test**. This is a test or survey administered to participants before and after a programme to measure change in knowledge of general financial literacy concepts and/or concepts specific to a programme. The pre- and post-test should be well tailored to the programme audience and objectives. The design of the pre- and post-test must account for age and education level of participants. For example, school children who participate in a course on financial responsibility could receive a test with very simple language, with the questions perhaps supported by images. Low-education employees could be given a test with language and concepts that are simpler than those directed to employees who are college graduates. The results of a post-programme test can be compared both with pre-programme and control group results.

- **Criterion Reference Group Test** (CRT). This test can be used for specific groups of individuals who may have problems with reading and writing (for example, very young or very old participants, or participants with low education). The CRT is similar to a group pre- and post-test, but is conducted orally. The instructor asks the questions out loud to the group; individuals may answer and other participants can supplement or correct those answers. Based on correctness and thoroughness, scores are assigned to the answers. This type of evaluation is also applicable to groups who are unlikely to be familiar with or who are uncomfortable with written tests.

- **Alternative assessment**. The alternative assessment presents participants with hypothetical scenarios, vignettes, and exercises to which they have to apply the concepts they have learned. The alternative assessment is mainly used to assess attitudes and expected behavioural changes rather than objective knowledge or actual behavioural changes. The alternative assessment has both advantages and limitations.

Follow-up survey. A survey conducted some time (months or even years) after the programme can measure knowledge retained by individuals long after their participation. The evaluator should be aware, however, of the many external factors that might contribute to and influence the knowledge of programme participants in the intervening time period.

Attitude outcomes

Attitude outcomes measure the effects of a programme on participants' attitudes toward financial literacy and financial responsibilities (e.g., the participant plans to keep closer track of his/her expenses in the future, says he/she will seek more information before purchasing a new financial product) or toward the topic specifically dealt with by a programme (e.g., the participant now believes it is very important to plan ahead for retirement). Attitude outcomes also assess the participant's level of self-confidence regarding general or specific issues (e.g., the participant feels more/less confident dealing with loans and mortgages) following the programme.

There are several methods through which to assess attitude outcomes:

- **Pre- and post-post survey**. The same pre- and post-test used for knowledge outcomes can include a survey to assess attitudes and self-confidence. The attitude results before and after the survey can be compared; participant responses could also be compared with those of a control group.

- **Alternative assessment.** The alternative assessment exercises used to evaluate knowledge can also be used to assess the self-confidence of participants in dealing with particular issues or handling specific concepts.

- **Follow-up survey**. Attitudes can be measured months or years after the programme via a follow-up survey.

- **Focus groups**. These groups are selected to discuss and share their experiences, attitude changes, and actions they are planning to take in the future. Focus groups should normally be held immediately after the conclusion of a programme. Focus groups scheduled for weeks or months after an initiative are less likely to be informative, as the direct effects of a programme may be disturbed by other external factors.

Behaviour outcomes

Behaviour outcomes are changes in behaviour normally associated with general economic issues (e.g., the participant has opened a savings account and/or keeps better track of his/her transactions) or specific issues addressed by a programme (e.g., the participant has changed his/her retirement plan or has bought a different financial product). Behaviour changes can be measured months or even years after a programme.

One way to assess the impact of a programme on behaviour is with follow-up surveys conducted a determinate amount of time after a programme. Done via phone or mail, or even face-to-face, such surveys assess how participants' financial behaviour has changed compared to earlier behaviours, which were identified through a pre-programme survey. Focus groups can also be used to gather qualitative data, asking participants to discuss their changes in financial behaviour. Another method is to collect data on behaviour using administrative records, for example bank records, employer records, or government records.

Other outcomes

Other Outcomes refer to specific outcomes, including changes in the financial well-being of participants. A follow-up survey can be used to determine these outcomes. Participants can be asked questions about their level of savings, investment income, retirement plans, satisfaction with their financial situation, or other indicators of financial well-being. The results should then be compared with the data collected via a pre-programme survey and with national data, if they exist. Countries can differ widely on measures of financial well-being and these differences need to be taken into account in performing an international comparison.

Programme impact

This last tier refers to the programme impact at the macro level and is relevant only for programmes that have a large scope (national or at least regional). The issue addressed in this tier is the effect of the programme on society as a whole; its impact relative to other possible initiatives to increase financial literacy and financial well-being in general. The measurements of this tier will be statistics and data at the national level regarding, for example, financial literacy, level of savings, percentage of households with checking/savings accounts, overall self-confidence of the citizens, etc.

This last tier reinforces the importance of establishing a national benchmark and keeping track of changes over time in financial literacy and other indicators. The evaluators in this case need to assess whether programmes (such as mandatory school-based financial education) create improvements over time in the entire community, not only among the people who participated in a financial education initiative.

One way to evaluate macro-effects is by conducting national surveys to analyse trends and changes in financial literacy and in other indicators. In order to do so, a standard survey with a determinate set of questions and a common measurement method should be established and maintained without alteration. In this way, the results of the surveys can be compared over time. These surveys can be conducted by the government or other agencies.

Some pitfalls in the implementation of the five-tier approach

The implementation of the five-tier approach faces many challenges. Some of these problems are again highlighted below and they can instrumental in the success or failure of the implementation of the programme.

Programme objectives

Sometimes the objective of a programme is not properly spelled out or is not well known to the evaluator or the reviewer of the programme. For example, some financial education programmes may be initiated to satisfy specific regulatory restrictions. Employers may offer retirement seminars to comply with laws. Similarly, financial counselling has become mandatory in some bankruptcy procedures. While the improvement of financial literacy may be a declared objective, it may in fact be secondary to the objective of complying with the law. Several studies have been done, for example, to assess the impact of employer-provided financial education programmes in the United States. However, the investigator often had limited or no data on the reason for initiation of the financial education programmes.

The objectives of the programmes also call into play the importance of an independent evaluation. Because the objectives of programmes are often improvements to financial literacy or to financial behaviour, there is a potential bias in reporting only the parts of the programme that work and downplaying or not reporting what does not work. However, the latter can be of great importance, too. Similarly, there is an incentive to choose evaluations that are flexible and tend to favour finding an effect, such as relying on descriptive methods rather than experimental methods.

Programme inputs

Very few studies report the costs of implementing a programme and it is consequently hard to truly assess effectiveness. Moreover, it is hard to make comparisons across programmes. One way to do so would be to rely on indicators such as the return on investment, which are, however, not applicable to every programme.

Programme delivery

Many programmes do not use more than one delivery method and it is therefore hard to disentangle whether (lack of) effectiveness is due to the programme design overall or simply to its delivery. For example, programmes relying on brochures, calculators, and heavy statistics may be unappealing to participants with low financial literacy. The programme may in fact be effective with a different audience but a mismatch between delivery method and recipient characteristics can make it ineffective.

Programme outcomes

As mentioned before, it is difficult to measure the outcome of the programme properly.

Pre-test, post-test, and survey

The main weakness of the pre- and post-test is that some of the indicators of attitude, self-confidence, and behaviour are self-reported. They can be biased indicators. The participant may also be uncomfortable reporting data on savings and debt and/or report them with error.

Follow-up survey

A follow-up survey is a good method through which to assess changes in behaviour, knowledge, attitude, and well-being. However, it faces challenges and limitations. One hurdle is **cost**. This type of survey, conducted via phone, mail, or face-to-face interview several months or years after the programme, is often very expensive.

The second limitation is the so-called attrition bias: as noted by Collins and O'Rourke (2009), many participants will not reply to a mail survey or will not be reachable by phone. This will cause a significant loss of data. There are other biases as well: the individuals who respond to a follow-up survey may be those who are more motivated to improve their financial well-being. Therefore, the results may be biased toward finding a result. Another possible bias is the fact that data on attitude, income, and financial well-being are self-reported and the evaluator often cannot prove whether the interviewed person is answering truthfully. Another limitation is that a follow-up survey cannot in any way control for other variables that might affect the results. For example, participants might have taken more financial education courses in the time period between programme and follow-up survey; they might have received a promotion at their workplace, etc. These events can affect behaviour and are hard to control.

Focus groups

Focus groups are a good method through which to obtain qualitative data on participants' opinions on the effects of a programme. They give the evaluator an opportunity to collect extended data that closed-ended questions and numerical scores of the pre- and post-test are often not able to communicate. With focus groups, specific details of the programme can be discussed in detail, including complaints and suggestions for improvement.

Programme impact

Delays in publishing data

One major limitation of assessing the impact of large programmes is the long time frame necessary to collect, elaborate, and publish data from large surveys. It may take months or sometimes years to collect data at the national level and to elaborate the results from the data. The costs of such surveys are high and often require a significant amount of resources.

Baseline and national surveys

Without a baseline to refer to in order to assess the impact of a programme, every single programme will have to collect data. Because baselines can be useful for a variety of programmes, it may be particularly valuable to centralise the collection of those data.

6. Recommendations

There are several recommendations that emerge from the five-tier approach, particularly for private and not-for-profit institutions. These institutions play a key role in promoting financial literacy, given the importance of a grassroots approach to promote financial literacy and financial education programmes.

The primary recommendation is simply a reinforcement of the importance of evaluation. Without an evaluation, no programme can be considered effective, and this can severely limit not only its adoption by other institutions, but also its funding. Thus, time, effort, and resources need to be allocated not only to programme design but also to programme evaluation. And, as mentioned throughout the report, design and evaluation should go hand in hand, and the evaluation should be part of the programme from the very beginning.

Given the inherent difficulties of measuring the effects and assessing the impacts of a programme, a variety of methods and tools should be employed. Using different methods will allow evaluators to obtain a more complete and multifaceted view of a programme's impacts. For example, both qualitative and quantitative data can be part of an evaluation. They serve different purposes and can provide complementary insights into the impact of and ways to improve upon a programme.

Given the many biases in evaluating the effectiveness of financial education programmes, experimental or quasi-experimental methods should be given priority. It is often hard - if not impossible - to get around self-selection biases and be able to assess the causality between financial education and financial behaviour. Being able to rely on a control and a treatment group both facilitates the assessment of the impact of the programme and gives the programme more credibility.

Clearly, one of the best ways to evaluate an educational initiative is to enlist an independent agency with professional expertise in the field to conduct the evaluation. However, professional agencies are expensive and their costs can be prohibitive for small, private organisations. Public and private higher-education institutions (such as colleges and universities) are perhaps an alternative to professional evaluation agencies. Higher-education institutions are frequently interested in partnerships which give them the opportunity to do research and gather new data. Small organisations could also take advantage of free (online) evaluation toolkits and resources, for example those provided by the OECD, their national governments, educational institutions, or other private organisations. Online resources provide important material that can help prevent common mistakes and propose best practices. Toolkits like the one designed by Lyons, Jayaratne, and Palmer for NEFE are user friendly and flexible, giving evaluators the ability to use established questions on a wide range of topics or create their own questions while maintaining the standardised layout of the evaluation.

Continuing on the previous point, evaluators should rely on external support (and partnerships) to conduct their analysis. Peer-reviewing of the evaluation methodologies before the implementation of the study and prior to the publication of the results is highly recommended.

Resources dedicated to financial education and its evaluation are scarce. Some coordination may be not only helpful but will likely benefit the community at large. For example, a baseline establishing the level of financial literacy and measuring at-risk financial behaviour can be done in a centralised way, without having each institution run a survey. Moreover, there may be a lot of overlap in the type of programmes and the target population. For example, many programmes have set up websites to provide information and help with financial decisions and, in doing so, institutions may end up replicating the efforts of other similar organisations.

There has been little attention to dissemination of results in the discussion of programme evaluation. However, one of the objectives of an evaluation is to prove its significance in order to be compared with other programmes or be adopted at large. Therefore, not only should the results of the evaluation be promptly made available but the evaluator should describe in detail the methodology used for the evaluation. A precise explanation of the methods is crucial to allow for their replication in other studies. If feasible, the evaluator should benchmark the outcomes of the programme with the results of earlier studies on similar initiatives. All results should be reported, not only those that provide evidence of a positive impact of the programme on behaviour. Knowing what is not effective can be as important as knowing what is effective. Results should be made available to any interested parties. For example, an evaluation study could be uploaded to the website of the organisation or agency that conducted the programme. The organisation could also share its report of the initiative with local, national, or international clearinghouses for financial education, such as the IGFE. The evaluation study should reach as many people as possible in order to share with other agencies the findings of the initiatives and allow other financial education providers and evaluators to build on previous knowledge and experiences.

Finally, the evaluation must follow proper guidelines to protect the rights of participants. Data confidentiality, proper disclosure, and securing privacy are not only necessary requirements but are also important requisites to limit attrition biases. Also, participation in the programme has to remain voluntary and evaluators have to think hard about the potential ethical issues of treating groups differently.

7. Conclusions

In this chapter a five-tier framework has been presented and discussed that is directly applicable to different types of financial education programmes. The framework is a simple guideline that financial educators can follow when designing evaluation studies. While leaving significant flexibility to the evaluators, the framework provides a high degree of standardisation, which will allow programmes to be compared both nationally and internationally.

Evaluators have the responsibility to follow the framework and apply it to their programmes in the most effective way. Their diligence in doing so will lead to progressively deeper insights into which methods are the most effective in assessing the impacts of financial education programmes, and perhaps even lead to the creation of novel evaluation methods. Therefore, the five-tier framework is just the starting point, a compass for policy makers and educators in the field.

The debate about financial education programmes and evaluation methods is far from over. The continued discussion and interaction among scholars, evaluators, and policy makers is necessary to enrich and improve upon the existing evaluation studies. Further efforts are needed to apply the five-tier framework to as many programmes as possible in order to assess the effectiveness of different initiatives and collect information on the needs of specific groups of citizens. In this way policy makers and private organisations will be able to tailor financial initiatives to specific needs, using the most effective methods. And the OECD can become an important resource by creating a supporting structure for all policy makers and financial educators.

The interest raised by financial education initiatives in many countries underlines the importance of financial literacy and its link to financial well-being. Financial education is increasingly becoming a priority among policy makers and private institutions in countries around the world. It is therefore important for institutions like the OECD to facilitate discussion and sharing of ideas among its members.

Notes

1 See OECD (2005).

2 I borrowed this definition from Collins and O'Rourke (2009), who used it to refer to those experiments that simply administer a pre-post survey to assess the outcomes of the program.

3 See O'Connell (2009) for detailed explanation of this framework.

References

Cakebread, C. (2009), "Is Financial Education in Canada Working? Research and recommended best practices for evaluating financial education programs." Working paper for the Financial Consumer Agency of Canada (FCAC).

Caliber Associates (2005), "JA Worldwide: JA Dollars and $ense pilot program formative evaluation, final report." For Junior Achievement Worldwide. *http://www.ja.org/files/afterschool/JA_Dollars_&_$ense_Final_Formative_Eval_Rpt.pdf*

Collins, J. M., O'Rourke, C. (2009), "Still holding out promise: a review of financial literacy education and financial counselling studies" a report for the Annie E. Casey Foundation.

Crossan, D. (2009), "Evaluating financial education programmes" for the Retirement Commission, New Zealand.

Duflo, E. and Kremer, M. (2008), "Use of Randomization in the Evaluation of Development Effectiveness" in Easterly, W. (ed.) *Reinventing Foreign Aid,* Cambridge, MA: The MIT Press, pp. 93–120

Duflo, E., and E. Saez (2003), "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *Quarterly Journal of Economics*, 118, pp. 815–842.

Fox, J. J., and Bartholomae S. (2008), "Financial education and program evaluation" in Xiao J. J. (ed.) *Handbook of consumer finance research*. New York: Springer, pp. 47–68.

Fox, J. J., Bartholomae S., and Jinkook L. (2005), "Building the case for financial education." *The Journal of Consumer Affairs*. Vol.39, No.1, 195–214.

Glazerman, S., Levy, D., and Meyers, D., (2002), "Nonexperimental Replications of Social Experiments: A Systematic Review." *Mathematica Policy Research, Inc*. *http://www.mathematica-mpr.com/publications/PDFs/nonexperimentalreps.pdf*

Hogarth, J. M., Federal Reserve Board, USA (2006), "Financial education and economic development". Prepared for *Improving Financial Literacy: International Conference hosted by the Russian G8 Presidency in Cooperation with the OECD*.

Jacobs, F. H. (1988), "The five-tiered approach to evaluation: Context and implementation" in H. B. Weiss & F. H. Jacobs (Eds.), *Evaluating family programs*, New York: Aldine DeGruyter.

Le Brun, M. (2009), "Analysis of O'Connell framework for evaluating financial education programmes."

Lusardi, A. (2004), "Saving and the Effectiveness of Financial Education," in O. Mitchell and S. Utkus (eds.), *Pension Design and Structure. New Lessons from Behavioral Finance*, New York: Oxford University press, pp. 157–184.

Lusardi, A. (2008), "Introduction," in A. Lusardi (ed). "*Overcoming the Saving Slump: How to Increase the Effectiveness of Financial Education and Saving Programs,* Chicago: University of Chicago Press, pp. 1–13.

Lusardi, A. (2009), "U.S. Household Savings Behaviour: The Role of Financial Literacy, Information and Financial Education Programs," in C. Foote, L Goette, and S. Meier (eds), "*Policymaking Insights from Behavioral Economics,*" Boston: Federal Reserve Bank of Boston Press.

Lusardi, A., P. Keller, and A. Keller (2008), "New Ways to Make People Save: A Social Marketing Approach," in Annamaria Lusardi (ed.), *Overcoming the Saving Slump: How to Increase the Effectiveness of Financial Education and Saving Programs,* Chicago: University of Chicago Press, pp. 209–236.

Lusardi, A., and O. S. Mitchell (2007), "Financial literacy and retirement preparedness: Evidence and implications for financial education." *Business Economics*, pp. 35–44.

Lyons, A. C., Palmer, L., Jayaratne, K. S. U., and Scherpf, E. (2006), "Are we making the grade? A national overview of financial education and program evaluation." *The Journal of Consumer Affairs*, Vol. 40, No.2, p.208–235.

Lyons A. C., Jayaratne, K. S. U., and Palmer L. "Evaluation Manual." NEFE® Financial Education Online Toolkit. (accessed 21/10/2009) *http://www.nefe.org/eval/manual_toc.html*.

Miguel, E., Kremer, M. (2004), "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, Vol. 72, No.1, pp. 159–217.

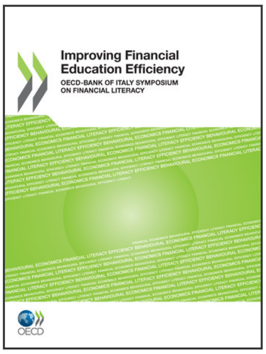
O'Connell, A. (2009), "Evaluating the effectiveness of financial education programmes" working paper for the Organisation for Economic Cooperation and Development.

Organisation for Economic Cooperation and Development (2005), "Recommendation on Principles and Good Practices for Financial Education and Awareness." *http://www.financial-education.org/dataoecd/7/17/35108560.pdf*

Organisation for Economic Cooperation and Development (2005), "Improving Financial Literacy: Analysis of Issues and Policies." *http://www.oecd.org/document/2/0,3343,en_2649_15251491_35802524_1_1_1_1,00.html#HTO*

Pritchett, L. (2002), "It Pays to Be Ignorant: A Simple Political Economy of Rigorous Evaluation" *Journal of Policy Reform*, vol. 5(4), pp. 251–269.

Retirement Commission, New Zealand. "2009 Financial Knowledge Survey." *http://financialliteracy.org.nz/sites/default/files/webfm/Research/Financial_Knowledge_Survey_2009_report.pdf*



From:

Improving Financial Education Efficiency

OECD-Bank of Italy Symposium on Financial Literacy

Access the complete publication at:
https://doi.org/10.1787/9789264108219-en

Please cite this chapter as:

Lusardi, Annamaria (2011), “A Framework for Evaluating Financial Education Programmes”, in OECD, *Improving Financial Education Efficiency: OECD-Bank of Italy Symposium on Financial Literacy*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264108219-5-en