投資者及理財教育委員會
Investor and Financial Education Council

A Practical Guide to Evaluating Financial Education Programmes in Hong Kong

Acknowledgement

The Investor and Financial Education Council would like to thank Adele Atkinson of the OECD International Network on Financial Education, Rebecca Kong of the Financial Consumer Agency of Canada and Professor Samson SK Tse of the Faculty of Social Sciences at the University of Hong Kong for their valuable comments and suggestions that helped to substantially improve this guide.

Table of Contents

What this guide is about

Evaluation of financial education programmes is a commitment to accountability. It enables practitioners to gauge the impact – in terms of knowledge gain, attitude shift and behavioural change – of a programme and further improve the effectiveness.

Nevertheless, it is uncommon for financial education practitioners in Hong Kong to include evaluation in their programme planning[1]. A lack of resources and technical know-how in implementing evaluation are often cited as common barriers.

In fact, programme evaluation does not necessarily take up a lot of resources. And it can be reasonably easy to implement. At the Investor and Financial Education Council (IFEC), we pride ourselves on adopting an evidence-based approach to our education work. For all our major programmes, evaluation is always considered in the development stage and we have accumulated quite some experience in conducting evaluation. To encourage fellow financial education practitioners in Hong Kong to reap the benefits that evaluation brings, IFEC has produced this practical guide to share IFEC's experience in evaluating financial education programmes - including mistakes we have made so others don't repeat them. We hope this guide can help to make things easier and enable practitioners, especially programme managers working at the frontline, to be better prepared to design and implement a programme[2] evaluation.

We believe that practitioners will be surprised by the value of their evaluation results, and we are keen to see the financial education community benefit from more shared learnings.

[1] According to IFEC's "Stocktake of Financial Education Initiatives in Hong Kong 2018", only 8% of financial education initiatives captured come with evaluation results (either reported by practitioners or available in the public domain).

[2] The word "programme", used throughout the document, essentially refers to all forms of financial education initiatives.

An overview of evaluation designs

Before we discuss how to plan and implement a programme evaluation, let's review the basics of evaluation.

Basic evaluation designs

There are three basic types of evaluation design - true experimental, quasi-experimental and pre-experimental designs.

In a **true experimental design**, participants are randomly assigned to an "intervention group" that participates in the programme, and a "control group" that does not. Because the people in the intervention group are chosen at random from among potential participants, we can assume that the only difference between them and the control group (on average) is that they have received financial education. If the evaluation measures show that they have higher levels of knowledge or change their behaviour more than the control group, it is highly like that the differences are a result of the programme.

A **quasi-experimental design** also relies on comparisons between an "intervention group" and a "control group". With this approach, as opposed to randomly assigning participants to the two groups, the researcher seeks to identify non-participants that share important characteristics with the participants to create a control group.

Pre-experimental designs are the simplest form of evaluation as only the participants of a programme who receive an "intervention", i.e., the programme, are studied. No control groups are involved.

The following table lists the major pros and cons of these three evaluation designs:

	Pros	Cons
True experimental design	It is the most stringent evaluation design.	Random assignment of programme participants is challenging and not always possible, and significant resources are usually required.
Quasi-experimental design	It is still regarded as a robust evaluation design and is more feasible for social/educational programmes compared to true experimental designs.	It can be difficult, and not always feasible, to set up a control group that has the same profile as the intervention group across key variables, and is willing to participate in the evaluation process. While demographic profile can be relatively easy to match, ensuring the two groups have similar knowledge level and interest in the subject matter as well as intention to act is not as easy.
Pre-experimental design	It is easy to set up and requires less resources.	It is considered a less robust evaluation set-up compared to true or quasi experimental designs.

In general, true or quasi experimental designs should be considered where feasible and if resources are available. Nevertheless, sometimes it is advisable and appropriate to adopt a less stringent design when the programme managers are not entirely sure about the programme under trial. Pre-experimental designs can also yield valuable evaluation data and are more prevalent for evaluation of public or social education initiatives due to practical considerations.

Pre-post vs post-only designs

Each of the three types of evaluation design can be further divided into two major[3] sub-types – post-test only and pre-post design. In a **pre-post design**, participants are assessed on key measures before and after participating in a programme. The post-tests can be repeated at different points of time, for example, immediately after completing the programme and a few months later (usually as a follow-up survey after the first post-test) to ascertain actual behavioural change and sustained motivations to better manage one's finance.

Meanwhile, in a **post-test only design**, no pre-tests are carried out, but tests are conducted after a programme. This is commonly used when it is not feasible to pre-test the participants or for single-session short programmes. Sometimes in a post-test only survey, participants are asked whether they think the programme has caused any change in their knowledge, attitude or intention to act.

It is important to note that in a true or quasi experimental design involving comparing an intervention group with a control group, surveys or measurements of the two groups should happen at the same time. For example, if participants in an intervention group is surveyed before participating in a programme in January and then after the programme in February, then non-participants in the control group should also be surveyed in January and February respectively.

Quantitative vs qualitative evaluation

Data can be quantitative (i.e., numbers such as test scores and ratings) or qualitative (i.e., descriptive texts generated from in-depth interviews or focus group discussions) in nature. The two can be supplementary to each other and it is always a good practice to include both quantitative and qualitative data in an evaluation exercise if resources allow.

Surveys provide numbers to quantify the effectiveness of a programme and make it possible to use statistical tools to identify significant changes, and are usually at the core of any robust evaluation. Paper copies of questionnaires administered among participants attending a programme are the most common tool. Online questionnaires, which can be administered on-site or via email invites, can also be considered. But the former would require all participants to have a device with mobile Internet connection while the latter would require collection of email addresses which is a piece of personal data and need to be handled with care. Also, online or mailing surveys administered after a programme generally risk low response rates.

[3] There are other variations built from different combinations of these two sub-types, such as Solomon four group design, factorial design, cross-over design, etc.

Qualitative evaluation generally involves conducting **focus groups**[4] with programme participants and/or **in-depth interviews** with partners or key stakeholders. Qualitative data is often used to provide depth of information to supplement survey findings and it is uncommon (but not impossible) to see a piece of evaluation consisting of qualitative data only. However, for programmes that cannot provide quantitative data for some reason (e.g. very small number of participants), conducting focus groups can still be a useful way to collect feedback and views which can provide valuable inputs especially in the pilot stage of a new programme. Further, conducting focus groups sometimes enables programme managers to discover unintended benefits of the programme, such as improved parent-child relationships after parents and children work together on a saving plan. Other forms of qualitative data such as diaries recording actions or thoughts on the part of programme participants are possible but less common. In some cases, observational studies – such as observing the interaction between the trainer and the participants or how a user navigates around a self-learning portal - may also provide useful insights.

[4] A focus group is a research method that brings together 6-10 people in a room to discuss views and perceptions regarding a topic. Unlike interviews, which usually occur with an individual, focus groups allow members of a group to interact and influence each other during the discussion.

Planning an evaluation

It follows that a programme evaluation can be relatively simple or complex to set up depending on the design selected, which will require varying amounts of resources. It is always the best practice to plan for evaluation at the very early stage of developing a programme, so that project timeline and allocation of resources can be aligned.

In coming up with an evaluation plan, programme managers are essentially making decisions on four major things:

1. What are the success measures?
2. Which evaluation design to adopt?
3. When to conduct evaluation?
4. Who should be conducting the evaluation?

There are different considerations for each of the above questions.

Defining a list of success measures

Broadly speaking, a financial education programme is always about knowledge gain, attitude shift and behavioural change. Depending on the programme objectives, the outcomes may be more focused on one aspect or can cover all three aspects. But sometimes the intended outcomes may not be able to be captured in an evaluation exercise due to certain limitations. For example, capturing behaviour change would require administering a survey after the participants have had a chance to take actions, which may not be feasible for single-session programmes like a day-long workshop. Instead, programme managers can think about measuring the intention to act.

It would be useful to list the top three to five intended outcomes of a programme and then determine what can be practically measured for different aspects:

- **Knowledge gain** – participants learn something new to them from the programme, for example, the concept of dollar-cost averaging in investing.
- **Attitude shift** – more participants adopt positive attitudes to money management, for example, gaining a greater appreciation of the need to plan ahead for retirement.
- **Behavioural change / intention to act** – more participants take action to better manage their finance or intend to act, such as setting a personal budget.

Measures regarding satisfaction with the programme – for example, participants are satisfied with different aspects of the programme and/or feel like recommending it to others – are also commonly included, though these are not directly related to the intended outcomes or impact of a programme. More detailed discussions on designing the survey instrument are covered in the next section.

When deciding the success measures, there must be adequate communication among the programme managers, evaluation managers/partners and key stakeholders of the programme. It is also important to recognise that while a programme can have immediate, intermediate or long-term impact on the participants' financial practices, evaluation is usually only able to gauge the immediate or at best intermediate impact due to the limitations of reaching the participants after an extended period of time.

The National Endowment for Financial Education[5] in the United States classified the levels of potential impact based on the programme types in this way:

	Potential impact		
	Immediate impact	**Intermediate impact**	**Long-term impact**
Short programmes (e.g. one-time, short seminars lasting no more than two hours)	▪ Perceived satisfaction ▪ Change of knowledge, attitudes, skills, and aspirations	▪ If the program is effective, change of financial practices and behaviour are possible	▪ Limited long-term change; possible goal achievement
Long programmes (e.g. a day-long workshop)	▪ Perceived satisfaction ▪ Change of knowledge, attitudes, skills, and aspirations	▪ Potential change of financial practices and behaviour	▪ Goal achievements and change of socioeconomic conditions are possible
Multi-session programmes	▪ Perceived satisfaction ▪ Change of knowledge, attitudes, skills, and aspirations	▪ Likely change of financial practices and behaviour	▪ High potential for goal achievement and change of socio-economic conditions

Selecting the right evaluation design

Selecting which evaluation design to adopt is all about balancing the need for a robust evaluation and the resources available as well as feasibility. There are a number of inter-related factors to consider:

1. ***Objectives and intended use of evaluation findings***

   Although the objective of evaluating a financial education programme is often to ascertain effectiveness[6], the focus can be different – for example, collecting data and participant feedback for fine-tuning a new programme, compiling an official report to seek funders' support, or even for quality control purpose when working with third-party service suppliers. Intended use of the evaluation findings determines the resources that should be allocated and the project timeline.

   In general, if the evaluation findings will be used as a measure of accountability, then it is wise to consider the most robust evaluation design possible. Also, engaging an independent assessor would be a good idea to reduce the risk of perceived bias in the evaluation results.

2. ***Potential impact of a programme***

   Intended use of the evaluation findings is often related to the potential impact of a programme. While all programmes are designed to achieve a certain level of impact, due to different programme objectives and resources available the level of impact would also

[5] Financial Education Evaluation Manual, NEFE, 2011

[6] There are also some types of evaluations that are not to do with gauging effectiveness, for example, formative evaluations that may try to ascertain efficiency, feasibility, appropriateness, etc. of a programme.

be different – for example, a short seminar versus a multi-session programme. Given the scarcity of resources, no one would disagree that programmes that carry the potential of making a deeper impact deserve to be allocated more resources in evaluation.

3. ***Potential sample size for evaluation***

   Robust evaluation requires adequate sample size for reporting the numbers and performing statistical analysis. However, it might not be easy for programmes that adopt a small-class approach for better engagement level to achieve a big sample size for evaluation. In these circumstances, a total sample size of at least 100 participants would be ideal, while samples with fewer than 50 participants are regarded as small samples (and 30 is usually the minimum sample size for reporting percentages). Of course bigger sample sizes in hundreds are even over a thousand are even better, though programme managers need to consider if they want to put a large sample through a programme yet to have confirmed effectiveness. Also, beyond a certain size, the difference in statistical power becomes insignificant.

   For programmes that risk having a small sample size for data analysis (for example, due to attrition issues), programme managers may consider whether it is feasible for data collection to cover sessions conducted over an extended period of time to accumulate a sufficiently large sample size.

4. ***Feasibility for data collection***

   On a related note to potential sample size is the feasibility for data collection. Apart from the potential reach of a programme, potential response rate to the evaluation surveys is also a key consideration factor. Programmes that require physical presence of participants have better opportunities of soliciting responses to evaluation surveys. Meanwhile, digital programmes such as online learning modules that call for use of online surveys generally see much lower response rate.

   Also, even if resources are available for a true or quasi experimental design, programme managers need to consider if it is feasible to achieve randomisation in assigning participants to intervention versus control groups, or set up control groups comparable to the intervention groups which often requires cooperation from the key programme partners.

Deciding on the timing of evaluation

How long a programme should be running before implementing evaluation is another key consideration factor. It makes sense to evaluate a new programme in the pilot stage so that the data and feedback collected can be used to improve the programme or avoid mistakes before further development or scaling up. Meanwhile, for new programmes it can also be useful, and sometimes necessary, to observe the operation in the pilot stage – such as enrolment to the programme, attrition rate, usage/take-up of tools - before deciding how best to carry out an evaluation.

In fact different evaluations can be conducted at different points in the project life cycle. A simple data collection exercise can be implemented in the pilot stage of a new programme to collect user and stakeholder feedback. And once the programme achieves stable and smooth operation, a more elaborate evaluation can be set-up for ascertaining impact. Moreover, it may be worth pilot testing the evaluation itself to ensure the evaluation design is feasible.

Engaging an independent assessor vs in-house management

Theoretically, evaluation is best conducted by an impartial third party for better credibility of the findings. But in reality, it is not always feasible due to lack of resources. Cost aside, finding a suitable consultancy with experience in evaluating an educational initiative can prove difficult in Hong Kong where expertise in this area appears to be limited.

Research consultancies and academia from relevant disciplines are both potential partners to work with, but it is important to bear in mind that while independent assessors experienced in research will likely be strong in research design and data analysis, they may not be familiar with financial education. Working with an independent assessor will require adequate communication regarding the programme objectives and evaluation goals, which often calls for a longer lead time in setting up the evaluation.

Meanwhile, it is always possible to conduct an evaluation in-house if circumstances do not allow bringing in an independent assessor. This practical guide to evaluation is intended to guide programme managers without prior experience in evaluation to conduct a programme evaluation without the support of an independent assessor.

IFEC's experience

We have dedicated staff working on evaluation and we mostly collect data and perform analysis in-house. But we also work with consultants and academic partners on selected evaluation projects.

At the IFEC, a pre-experimental and pre-post design is the most frequently adopted approach for evaluating educational programmes. We have also explored adopting a true or quasi-experimental design but so far we haven't been able to ensure true randomisation in assigning the participants or set up a satisfactorily comparable control group for meaningful comparison.

There was once when we adopted a quasi-experimental design for evaluating a workplace programme targeting pre-retirees in their late 40s and early 50s. The control group consisted of a well mix of individuals in their 40s and 50s recruited via an online survey panel (who indicated interest in a similar programme during the screening to ensure they had the same level of inclination to change). Meanwhile, it turned out that the majority of participants enrolled in the programme were in their late 50s. Even though the control group was designed to be bigger in size and covered a wider age bracket to cater to uncertainties in programme enrolment, it was still difficult to adjust to the profile of the intervention group with a decent sample size. The learning was that, for programmes that involve voluntary enrolment, the control group need to be much bigger than the intended sample size to allow for flexibility.

We will continue the exploration and hopefully will be able to share more experience in this regard soon enough.

Implementing an evaluation

After deciding on the success measures, evaluation design, timing and who will conduct the evaluation, it is time to get things rolling.

Setting up a control group

If it is decided to adopt a true experimental design, then programme managers need to recruit a pool of participants who are available to take part in the programme in two different periods of time, say in January and February. Participants are then randomly assigned (say by drawing lots) to the intervention group and control group. Participants in the intervention group will go through the programme and evaluation in January, while those in the control group will be tested at the same time as the intervention group in January but have to wait till February to take part in the programme. In this way, all participants get to take part in the programme eventually but the different timing makes it possible to test some of the participants as a control group. Apart from making it easier to solicit cooperation from those in the control group for surveys, this arrangement also ensures that participants in both the intervention group and control group have the same interest level in the programme or intention to act which is very important for a fair comparison.

School projects carry the best potential for true experimental evaluation designs as it is relatively more feasible to randomly assign students to the intervention or control groups, which can be based on schools or classes. However, it requires securing cooperation from the schools which often find the administration work challenging amid a very packed teaching schedule in Hong Kong.

For a quasi-experimental design, programme managers need to recruit a group of people who share the key characteristics as the programme participants, and who agree to be surveyed at the required timing without taking part in a programme. These non-participants can be offered chances to take part in the programme held later as incentives for participating in the study (or other forms of incentives may be required to secure cooperation). In fact if the programme involves voluntary enrolment on the part of participants (as opposed to compulsory participation say in school programmes), then programme managers need to screen for interest in taking part in a similar programme when recruiting the control group members. Other common variables to control for to ensure comparable profile between the two groups include gender, age, education level and work status (working vs non-working). For example, if it is a programme targeting pre-retirees in their 40s and 50s, then the control group should also consists of non-participants in the same age bracket. If information about the profile of participants is available early on (for example, around 40% are females and half are in their 50s), then it could be used to guide recruitment for the control group. Otherwise, the best a programme manager can do is to ensure a well mix of participants with the required characteristics or reference the proportion in the population.

When setting up control groups, it makes sense for programme managers to over-recruit non-participants. For example, if there are 100 participants in the intervention group, then the control group should ideally consists of 120-150 non-participants. This is because people in the control group may not be as cooperative to complete all surveys within the required time frame considering that they are not taking part in the programme at the same time. In a quasi-experimental set-up, a larger control group also allows flexibility to adjust the profile to match that of the intervention group in the data analysis stage.

Administering a survey

When evaluating a programme that requires physical presence of participants, surveys are commonly administered on-site by paper copy for the best response rates (pre-test administered right before starting the programme and post-test immediately after the programme is completed). Personal contacts such as email addresses may be collected on a voluntary basis[7] for administering online questionnaires as a follow-up survey sometime after the programme is completed to assess whether any change has been maintained or interrupted (incentives may be required to encourage participation). If it is decided to collect personal contacts for follow-up surveys, it is best to do so via a separate sheet of consent form (as opposed to including the information in the post-test questionnaire) so that participants can be rest assured that their responses to the survey will remain anonymous, otherwise some participants might adjust their responses if they feel they are identifiable.

When administering a survey, it is essential to communicate to the programme participants the purpose of conducting the survey and how the data collected will be used. Make sure they understand participation is voluntary. Also, it is useful to explain the whole evaluation process to the participants, especially for pre-post evaluation where they will need to fill in the post-test questionnaire that is basically the same as the pre-test one. Non-participants in the control group should be tested using the same questionnaires around the same time. Since they will not be physically present for surveys, online or mailing surveys are commonly used (which requires programme managers to obtain their contact details at the time of recruitment).

Further, for pre-post evaluation that involves repeated measures of the same pool of participants/non-participants in control groups, respondents should ideally be identified for matching pre and post surveys. This can be achieved by requesting respondents to create a reference number unique to themselves (e.g. by combining birth month and the last four digits of their mobile phone numbers). It is also possible to request respondents to provide some less sensitive personal data such as email addresses as identifiers, but again, it risks affecting how some participants answer the questions if they feel they are identifiable.

Designing the evaluation questionnaire

A survey takes up precious programme time in a face-to-face programme. Even for online or mailing surveys, a long questionnaire may cause survey respondents to drop out before completing the questionnaire. So it is important to keep a questionnaire to a reasonable length – a maximum of 10 minutes is ideal for longer programmes.

The questionnaire should cover key measures of learning outcomes that the programme sets out to achieve – in short, these would be the claims that programme managers want to make about the programme's impact. Depending on the programme objectives, these could include knowledge gain, attitude shift and behavioural change/intention to act. Each of these is illustrated with examples below:

- **Knowledge gain** – this can be tested by including a few knowledge check questions. To keep the questionnaire length manageable, true/false statements and multiple-choice questions are commonly adopted. For these types of questions, it is always a good idea to include "I don't know / not sure" as one of the answer options so that respondents who have no idea can honestly say so and not forced to guess an answer.

[7] Collecting personal data requires compliance with the Personal Data (Privacy) Ordinance in Hong Kong.

Example 1: Annuities are a type of insurance product. Is this statement true or false?

True (correct answer)
False
Not sure

Example 2: How do we calculate the price-earnings ratio of a stock?

By dividing the stock price by net asset value per share
By dividing the stock price by earnings per share (correct answer)
By dividing the market value by earnings per share
Not sure

- **Attitude shift** – this is often tested by asking participants to indicate their agreement level to selected attitude statements. The statements can be positive or negative, and pre-post comparison of agreement or disagreement to the statements can be conducted.

  Example: How much do you agree with the following statements?

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
One should always save before spend.					
Young people who just enter the workforce should also start planning ahead for retirement.					
It is alright to borrow money to enjoy life as long as one has a stable income.					

- **Behavioural change/intention to act** – although behavioural change is often regarded as the ultimate learning outcome of a financial education programme, it is not always possible to capture in evaluation surveys. For single-session programmes it would be more feasible to include questions checking "intention to act" instead.

  Even for multiple-session programmes, participants may not have a chance to take certain actions, such as settling credit card bills and conducting research before purchasing a new investment product, within the relatively short period of time between the two surveys (e.g. administered at the first and last session of a programme respectively).

  Where feasible it is always a good idea to check actions taken or intention to act simultaneously:

  Example: For each of the following items, please indicate whether (a) it is in your current practice, (b) you currently don't practise it but intend to start, or (c) if you do not think it is necessary. For items are not applicable to you, for example, you do not use credit cards, please select "NA (not applicable)".

	Currently doing	Intend to start	Don't think it's necessary	NA
Track my expenses				
Review my MPF account regularly				
Always settle the outstanding balance of credit cards in full				

In this way, programme managers can gauge changes in intention to act apart from checking actions taken. Nevertheless, depending on the intended learning outcomes of a programme and the target segment, in some cases it may be more suitable to just check whether participants have adopted certain practices without complicating the question with intention to act. Meanwhile, programme managers can also consider adding more options in the scale for better granularity, for example, to further differentiate "want to start but don't know how" and "feel ready to start".

Again, it is advisable to keep the questionnaire concise by including only the most important learning outcomes. It is also important to ensure the coverage is balanced among the different topics or areas covered in the programme. For example, a multiple-session programme might have talked about day-to-day money management, investment and insurance. Each of the areas should be covered in a balanced manner, or reflect the amount of time spent covering that topic.

Depending on the target segment, questions checking basic demographics could be included. Questions gauging participants' satisfaction with the programme may also be useful.

- **Basic demographics** – Information on the participants' key characteristics such as gender, age bracket, education level, etc. is essential for matching the profile with the control group in a quasi-experimental set-up. Even for other evaluation designs the basic demographics can provide better understanding of the participants. It also allows subgroup analysis such as checking if the programme is equally effective among males and females.

- **Satisfaction ratings** – it is common to gauge participants' satisfaction with a programme and collect feedback and suggestions.

  Example: Please indicate how strongly you agree or disagree with the following statements about the workshop.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
This workshop has strengthened my financial knowledge.					
Overall I am satisfied with this workshop.					
I will recommend this workshop to others.					

  Open-end questions asking participants what they like about the programme as well as suggestions for improvements are usually included as well.

For pre-post evaluation, basically the same set of questions on knowledge gain, attitude shift and behavioural change/intention to act should be used to ensure comparability, while the post-test questionnaire will see the addition of questions on satisfaction ratings and open-end questions for qualitative feedback. If there is a follow-up survey, say three to six months after a programme is completed, then again the same set of questions on knowledge gain (optional), attitude shift and behavioural change should be used for comparability. An example of questionnaires used for a pre-post evaluation of an IFEC programme is included in the appendix for reference.

Lastly, just as programme managers may pilot a programme before going on full-scale, it is also a good practice to pilot the evaluation questionnaire (say among colleagues who are not directly

involved in the programme) to ensure it is of a manageable length and that question wording and instructions are clear to the respondents.

Keeping administrative records

Evaluations always benefit from integrating multiple sources of data. And where possible administrative records such as programme reach/attendance and attrition rate (over multiple sessions) should be presented alongside the survey data.

Conducting focus groups or interviews

Qualitative feedback from programme participants and key personnel/stakeholders involved in a programme adds depth to survey data and is especially useful for new programmes.

Discussions should mostly cover “what’s” and “why’s” - common questions to probe in a focus group may include:

- What attracted you to join this programme?
- Did this programme meet your expectations? Why / why not?
- We covered a range of topics related to managing personal finance in this programme, which ones did you find more useful? Are there any topics that you think should be elaborated on more or can be taken out? Why did you think so?
- Did you learn anything new from the programme, i.e., things that you didn’t know before joining this programme? Do you think the depth of content is right for you?
- Did the course make you feel like taking any actions to better manage your finance? Why or why not?
- Overall, what would you say are the best things about this programme?
- Are there any areas that you think the programme can improve on?

If an independent assessor is not available, it would be ideal to arrange for a facilitator who is not heavily involved in the programme, so that participants will feel more at ease to point out areas for enhancements. Try to keep the focus group size to around six to ten individuals for the best dynamics.

When recruiting participants for focus groups, incentives (such as souvenirs or supermarket coupons) may be considered to encourage participation, especially if travel is required on the part of respondents.

Apart from programme participants, it would also be a good idea to interview key personnel involved in a programme such as instructors for group activities, employers for workplace programmes and teachers involved in coordination of school projects.

IFEC’s experience

We have experimented with different ways of administering a survey among programme participants, covering on-site surveys using paper questionnaires, online surveys, mailing surveys and phone interviews (for the latter three, prior consent and contacts are obtained from programme participants). On-site surveys mostly yield almost 100% participation, while other forms of surveys generally see less than 30% response rate even with the support of incentives like supermarket coupons. When it comes to surveying programme participants, it seems that “out of sight is out of mind”. Therefore, it is always recommended to conduct on-site surveys as much as possible and only use other means for follow-up surveys.

Analysing data and reporting evaluation findings

Once all quantitative and qualitative data are collected, programme managers can start putting together the evidence.

Processing the survey data

For many programme managers who are not familiar with working with quantitative data, the most daunting part of an evaluation may be processing the data. But in fact data processing of a short survey can be rather simple and there are two common ways (assuming most programme managers do not have access to professional statistical packages like SPSS or SAS):

1. Using Excel spreadsheets

   As most evaluation questionnaires are short and simple, it is easy to punch in the data to a spreadsheet. The questions and answer options will be the columns (answer options should be coded) while each row will be the record for one respondent. Below is an example using the sample knowledge check question about annuities:

Respondent no.	Q1 - annuities
981003	2
619508	1
972712	1
938203	2
602911	3

Coding of answer options:
1 - True
2 - False
3 - Not sure

   After all data punching has been completed, it is easy to set formulae to count the responses and come up with a simple frequency table of the questions.

2. Using online survey platforms

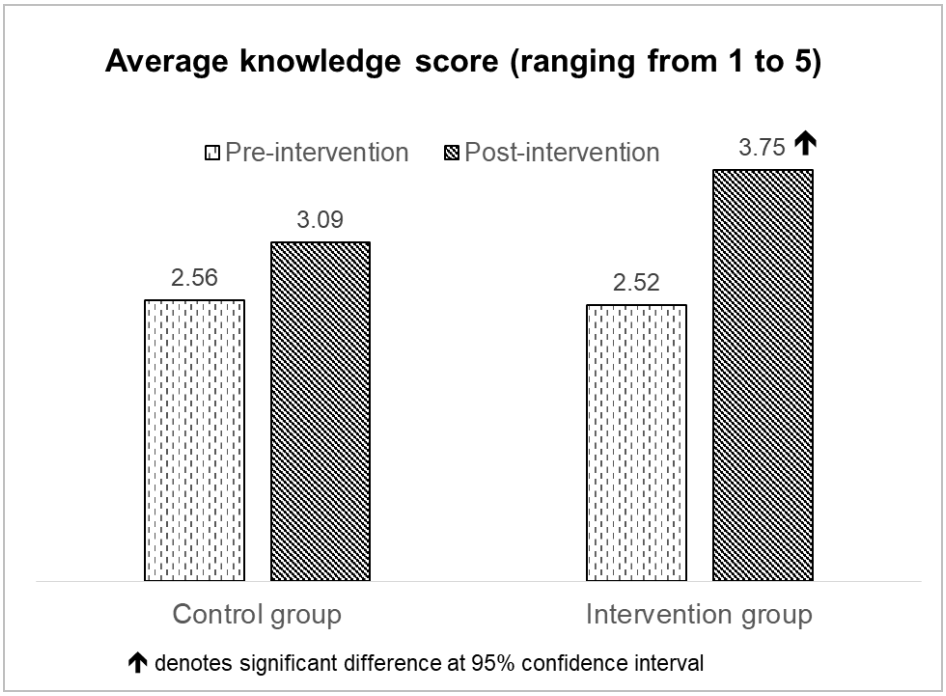
   If resources allow, an even easier option is to subscribe to the service of an online survey platform. Use it to create an online questionnaire based on the same questionnaire used and then simply punch in the data to it. Most of these platforms then automatically display the aggregated data in both counts and percentages and it is also possible to generate simple reports with charts.

In pre-post surveys, if identifiers are used for matching respondents, then the two sets of survey data need to be matched (e.g., by using the VLOOKUP function of Excel based on the respondent no. in the above example) before preparing the frequency tables. Data items that cannot be matched will have to be dropped for the pre-post comparison – which often means a smaller sample size for reporting. For example, there might be 100 responses for the pre-test, and 95 responses for the post-test (due to some participants not returning the questionnaires). As is often the case, some participants will leave the identifier blank or may not follow the instruction correctly in coming up with one. Therefore, after the matching exercise there might only be 80 matched responses for measuring change. Nevertheless, these unmatched data items can still be used in other ways (such as analysis of differences by key characteristics).

Presenting the survey findings

Survey numbers should be presented in a user-friendly format for the intended readers of the report. Depending on the reporting needs, PowerPoint presentation decks or Word document are both common report formats.

When conducting pre-post comparison or comparing results between the intervention group and control group, findings should ideally be presented side-by-side and any changes highlighted for easy reference. For example:



It would also be very useful to include a significance test to indicate whether a difference noticed between two surveys (e.g., increased knowledge score) is statistically significant. Simple significance tests such as t-test are generally available in Excel or common online survey platforms. There are also many free online statistical tools available.

Interpreting the findings

After putting together the survey findings, other sources of data such as administration records and qualitative data should be integrated to paint a complete picture before discussing the findings and drawing conclusions about the programme.

In any evaluation with a range of measures, it is likely that some measures would see significant differences while other measures may remain stagnant. Discussions should cover the strengths and weaknesses of the programme as the data indicates. Be cautious about drawing causal relationships – that any significant change is attributable to the programme – especially for pre-experimental evaluation designs without control groups. Very often a programme may not be the only cause of positive changes in participants' knowledge, attitudes or behaviour. There are always other possible factors, especially for multiple-session programmes conducted over a period of time, such as a news coverage of the financial markets or a television programme about managing finance that happen to take place around the same time. Or sometimes the

evaluation survey, for example the knowledge check questions, can trigger participants' interest and heighten their sensitivity to other sources of information. All these can also explain the increased scores/ratings among participants in the control group who have not received the programme at the time of survey. Although it is difficult to establish causal relationships, the fact that the programme is associated with significant increase in knowledge level, attitude shift and behavioural change/intention to act is already an achievement.

Negative changes in a pre-post evaluation can also occur for a wide variety of reasons. Participants may misunderstand the questions, or external factors may cause them to adopt undesirable behaviour such as going deeper into debt. In particular, it is not uncommon to see decreased confidence level in managing personal finance after attending a programme, which could be simply because some participants are over-confident before attending a programme and not aware that managing one's finance well involves a wide range of skills.

Needless to say, it is always a good practice to report and take into account the limitations of the evaluation, which usually include:

- **Limited comparability between the intervention and control group** – despite the best of efforts, programme managers may not be able to achieve truly random assignment of participants to the intervention and control group as intended (in a true experimental set-up); or the profile of the control group in a quasi-experimental set-up may turn out very different from the intervention group. Programme managers also need to be aware of the potential for information leakage across the intervention and control groups – for example, participants in the intervention group may share what they have learned with non-participants in the control group (which is more likely in school programmes).
- **Limited sample size/response rate** – sometimes the achieved sample size may be smaller than planned due to obstacles such as high attrition rate before completing the programme (e.g. among segments that may have difficulties physically attending programmes). Also, follow-up surveys generally see a fall in response rates, as the respondents are no longer engaged with the programme.
- **Self-report bias** – as the evaluation surveys are usually based on participants' self-reports, findings are subject to bias for several reasons. For example, participants may be reluctant to report that they haven't changed when someone has tried to help them. Or they may not wish to tell the truth about their actions/intention to act - perhaps because they feel such information is confidential or they feel ashamed of their intentions. Some participants may simply misunderstand the questions (especially when filling in the questionnaire in a rush after finishing a programme session).
- **Only immediate impact can be captured** – because most evaluation surveys are administered shortly after the programme, any longer-term impact would not be reflected in the evaluation findings. This is perhaps the most significant limitation of a programme evaluation.

Finally, conclusions drawn about the evaluation findings should ideally be followed by recommendations for future programme development or studies based on the evaluation results.

Sharing the learnings

The last step of an evaluation is to share the findings with the financial education community to build evidence for the effectiveness of financial education and promote learning. Dissemination of evaluation results can be as simple as circulating the report among stakeholders, or better still,

making the evaluation report available in the public domain (e.g. uploading onto the corporate website).

There was an academic paper[8] authored by two German scholars in 2017 that concluded "financial education significantly impacts financial behaviour and, to an even larger extent, financial knowledge" based on a meta-analysis of 126 financial education evaluation studies conducted during 1999 to 2015 (majority of the studies are from the U.S. and other OECD countries). This type of academic research that carries significance is only possible because many financial education practitioners made available their programme evaluation results.

[8] Does Financial Education Impact Financial Literacy and Financial Behavior, and if So, When?, Tim Kaiser and Lukas Menkhoff, 2017

On-going monitoring

Evaluation should not be a one-off exercise for any financial education programmes. After an initial evaluation that demonstrates effectiveness, on-going monitoring that often involves just a simple post-only survey among participants should be in place. And because participants' response to education tactics, along with many other things, can change over time, it is a good practice to review the need for another round of robust evaluation as and when appropriate.

And if an evaluation shows inadequacies and significant changes have been made to a programme to attempt to address the inadequacies, then another round of evaluation of the revised programme is certainly required. Subsequent evaluations can incorporate lessons learned and ask new questions.

As pointed out by the OECD International Network on Financial Education[9], evaluation is part of an on-going process of monitoring and improvement as the programme evolves. With robust evaluations, programmes become more effective leading to knowledge gain, attitude shift and ultimately, behavioural change.

[9] Detailed Guide to Evaluating Financial Education Programmes, OECD International Network on Financial Education, 2010

Appendix – Sample evaluation questionnaires

The following questionnaires were used in the evaluation of one of the IFEC's financial education programme targeting retirees, which consisted of six bi-weekly workshops addressing the key financial issues retirees face. The programme aimed at equipping the participants with the knowledge, motivations and skills to better manage their finances in retirement.
A pre-experimental and pre-post evaluation design was adopted, with the pre-test administered at the first workshop and the post-test administered at the last workshop. There was also a following-up survey (via emails and mailing) about three months after the programme was completed.

Pre-test questionnaire

****

參考編號: _____
(請填寫你的生日日期，例如你的生日日期為 3 月 25 日，你的參考編號將會是 0325)
Ref no: _____
(Please use your date of birth as the reference number, e.g., if your birthday is on 25 March, then your reference number will be 0325)

Q1. 請判斷以下句子是否正確，如你不肯定，請選擇「不清楚」。
Please indicate if you think each of the statements below is true or false. If you are not sure about a statement, please select "Not sure".

		正確 True	不正確 False	不清楚 Not sure
K1	利息上升會推動債券價格上升 An interest rate hike will push up bond prices	☐	☐	☐
K2	投資於不同行業的股票可以達到充分分散投資之目的 Investing in stocks in different sectors can sufficiently diversify investment risks	☐	☐	☐
K3	年金計劃是一種能提供現金流的保險產品 Annuities are an insurance product that generates stable income streams	☐	☐	☐
K4	透過預先簽署持久授權書，假如我在精神上變成無能力行事時，被授權人便可代為處理我的個人財務 Signing an EPA enables my appointed representative to take care of my finances in the event that I lose my mental capacity	☐	☐	☐
K5	安老按揭是一種以住宅物業作抵押的貸款，借款人可不用償還貸款，直至他/她離世 Reverse mortgage allows using a residential property as a loan arrangement that the borrower does not need to repay during his/her lifetime	☐	☐	☐

Q2. 你有多同意以下各項有關理財的講法呢？
How much do you agree with the following statements about managing personal finance?

		非常同意 Strongly agree	同意 Agree	中立 Neutral	不同意 Disagree	非常不同意 Strongly disagree
A1	我應該定期計算自己擁有多少資產，以及估算退休生活需要多少儲備 I should regularly review my financial position and estimate the amount of retirement fund I need	☐	☐	☐	☐	☐
A2	作為一個退休人士，資產越少，投資時越應進取以爭取擴大退休儲備 As a retiree, the fewer assets I owned, the more aggressive I should be in investment so as to expand my retirement reserves	☐	☐	☐	☐	☐
A3	子女成年後有責任維持財政獨立，即使我在財政上支援子女也應量力而爲 Children should be responsible for their own finances once they start working, and parents don't have the obligations to provide financial assistance	☐	☐	☐	☐	☐
A4	身心健康與個人財政健康息息相關 A healthy living style is instrumental to financial well-being	☐	☐	☐	☐	☐
A5	我有信心能够為退休生活妥善理財 I'm confident I can manage my personal finances well in retirement	☐	☐	☐	☐	☐

Q3. 就著以下每一項, 請剔選 (a)你現時或過去一年有這樣做 (b) 現時沒有這樣做但打算實行 (c) 認為沒有需要. 若有關項目並不適用於你, 例如你並沒有信用卡, 請選擇「不適用」。
For each of the following items, please indicate whether (a) it is in your current practice/you did it in the past year, (b) you currently don't practise it but intend to start, or (c) if you do not think it is necessary. For items are not applicable to you, for example, you do not use credit cards, please select "NA (not applicable)".

		我有這樣做 Currently doing/did this in past year	我沒有這樣做，但打算嘗試實行 Intend to start	我認為沒有需要 Don't think it's necessary	
B1	估算退休儲備是否足够 Estimate if my retirement fund is sufficient	☐	☐	☐	
B2	制訂一份個人日常開支預算 Set a personal budget	☐	☐	☐	
B4	檢討自己的生活習慣並嘗試建立更健康的生活 Review my habits and try to adopt a healthier lifestyle	☐	☐	☐	現時生活已很健康 Already leading a healthy lifestyle ☐
B6	訂立平安紙 Make a will	我已經訂立 Already have one ☐	☐	☐	
B7	訂立持久授權書（EPA） Set up an EPA	我已經訂立 Already have one ☐	☐	☐	不清楚什麼是 EPA Not sure what EPA is ☐

Q4. 你的性別 (Gender)： 男 女

Q5. 你的年齡 (Age bracket)： 50 歲以下 50-59 歲
60 – 69 歲 70 歲或以上

Q6. 你的教育程度 (Education level)： 小學或以下 中學/預科
大專 大學或以上

****

投委會 IFEC

Post-test questionnaire

A big part of the post-test questionnaire is the same as the pre-test questionnaire, which covers the same knowledge check questions, attitude statements and action items. There are only two differences:

1. Questions about satisfaction ratings and open-end questions are included to collect qualitative feedback
2. Questions on demographics in the pre-test are removed.

****

這份問卷跟你於第一節工作坊前填寫的一份很相似，我們是想看看你在工作坊後在個人理財方面的想法有否改變。請選出現時最符合你想法或者做法的答案，謝謝。
This questionnaire is very similar to the one that you helped to fill in at the first workshop. It's because we'd like to see if there are any changes in how you think about personal finance management after attending this programme. Please bear with us and select the answer options that best describe your current thoughts and practices.

Reference number - same as pre-test

Q1. (knowledge check questions) – same as pre-test

Q2. (attitude statements) – same as pre-test

Q3. (action items) – same as pre-test

Q4. 請問你有多同意以下各項有關這個課程的陳述？
How much do you agree with the following statements in describing this programme?

		非常同意 Strongly agree	同意 Agree	中立 Neutral	不同意 Disagree	非常不同意 Strongly disagree
S1	課程增進了我的理財知識 This programme has strengthened my financial knowledge.	☐	☐	☐	☐	☐
S2	參加課程後我有調整個人理財方式 I have adjusted the way I managed my finance since joining this programme.	☐	☐	☐	☐	☐
S3	課程令我檢討自己的理財習慣 This programme prompted me to review my financial habits.	☐	☐	☐	☐	☐
S4	整體來說我對這個課程感到滿意 Overall I am satisfied with this programme.	☐	☐	☐	☐	☐
S5	我會推介朋友參加這個課程 I will recommend this programme to others.	☐	☐	☐	☐	☐

Q5. 這個課程有什麼需要改善的地方？請儘量與我們分享你的意見。
In what ways do you think we can improve the programme?

Q6. 這個課程有什麼你欣賞而認為值得保留的地方？
And what are the things that you like about the programme that you think should be kept?

****

Follow-up questionnaire
(email/mailing addresses and consent forms were collected in a separate sheet at the last session of the programme)

Reference number - same as pre-test

(knowledge check questions were removed)

Q2. (attitude statements) - same as pre-test

Q3. (action items) – same as pre-test

Q4. 自完成這個理財課程之後，你有沒有進行以下各項？
Have you done any of the following since completing the programme?
- 瀏覽錢家有道的網站
  Browse the Chin Family website
- 下載錢家有道的收支管家流動應用程式
  Download the Chin Family Money Tracker mobile app
- 使用錢家有道的收支管家流動應用程式
  Use the Chin Family Money Tracker mobile app
- 以上皆沒有
  None of the above

References

This guide has referenced a number of useful resources about evaluation of financial education programmes, including evaluation principles, toolkits and guides. Programme managers may want to further study these resources when planning for programme evaluation:

1. A comprehensive collection of documents covering evaluation principles and practical steps prepared by the OECD International Network on Financial Education
   http://www.oecd.org/daf/fin/financial-education/evaluatingfinancialeducationprogrammes.htm

2. An online evaluation toolkit and other resources developed by the National Endowment for Financial Education
   https://toolkit.nefe.org/

3. “Financial Literacy Outcome Evaluation Tool” developed by Prosper Canada Centre for Financial Literacy
   http://outcomeeval.org/

4. Evaluation analysis tools developed by the Money Advice Service
   https://www.moneyadviceservice.org.uk/en/corporate/debt-advice-evaluation-analysis#analysing-manually-completed-paper-questionnaires

5. “Does Financial Education Impact Financial Literacy and Financial Behavior, and if So, When?”, Tim Kaiser and Lukas Menkhoff, 2017
   https://www.diw.de/documents/publikationen/73/diw_01.c.529454.de/dp1562.pdf

投資者及理財教育委員會
Investor and Financial Education Council

21/F, Cheung Kong Center, 2 Queen's Road Central, Hong Kong
T: +852 2700 6000 | F: +852 2297 3300