

WP/18/151

IMF Working Paper

Evaluating IMF Training: What Can We Learn?

by Hali Edison, Dimitre Milkov, and Josephat Koima

***IMF Working Papers* describe research in progress by the author(s) and are published to elicit comments and to encourage debate.** The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

I N T E R N A T I O N A L M O N E T A R Y F U N D

IMF Working Paper

Institute for Capacity Development

Evaluating IMF Training: What Can We Learn?

Prepared by Hali Edison, Dimitre Milkov, and Josephat Koima*

Authorized for distribution by Valerie Cerra and Norbert Funke

June 2018

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

Abstract

The IMF provides training to its membership in its core areas of expertise mainly through its Institute for Capacity Development (ICD) or formerly the IMF Institute (INS). This paper looks at the methods that ICD used to evaluate this activity and analyzes the data collected over the period 2006–13. Since 2015, ICD has undertaken a review of its curriculum and revamped its courses and evaluations. Hence this paper provides a detailed analysis of the situation prior to the review. The study's novel feature is its attempt to distill information from all evaluation sources in one place. It also conducts analysis to explain the evaluation results using participant demographic information. An important message that emerges from the different surveys is that ICD's training program is well liked. Notable differences in results surface when sorting evaluations results by course type or by geographic location, and whether evaluations were filled out by participants or by their sponsoring managers.

JEL Classification Numbers:

Keywords: Evaluation, Capacity Development

Author's E-Mail Address: hedison@imf.org, dmilkov@imf.org, koimakip@gmail.com

* The authors thank Dominique Desruelle, Alan MacArthur, John Green, Ray Brooks, Norbert Funke, Robert Powell, Sean Culhane, Ellen Nedde, Charleen Gust, Martin Schindler, Adolfo Barajas, Yasemin Bal Gunduz, and Marcelo Selowsky for useful comments and Ranae Jabri for her research assistance. The authors also benefited from comments received during an ICD workshop and presentation at a technical meeting of the Committee on Capacity Building (CCB).

Abstract	2
I. Introduction	5
II. ICD Training Program	7
III. End-of-Course Surveys	11
A. Methodology	11
B. The data	12
C. What can we learn from the data?.....	13
D. Summary	19
IV. Follow-Up Surveys.....	19
A. Methodology	20
B. The data	21
C. Results	22
D. Linking follow-up surveys with end-of-course evaluations.....	25
E. Summary	28
V. Triennial Surveys.....	29
A. Methodology	29
B. The data	31
C. The effectiveness of training: Results	31
D. Summary	36
VI. Conclusions	36

Tables

1. ICD Training Program: End-of-Course Surveys by Department, FY 2007–13.....	13
2. ICD Training: Average Scores by Fiscal Year, FY 2007–13.....	14
3. ICD Training: Average Scores Across Locations, FY 2007–13	16
4. End-of-Course Survey Regression Results.....	18
5. Follow-Up Survey Regression Results.....	25
6. Follow-Up Surveys and End-of-Course Survey Results, 2007–11.....	26
7. Follow-Up Surveys and End-of-Course Survey Results for Selected ICD Courses	27
8. Follow-Up Surveys and End-of-Course Regression Results	28

Figures

1. IMF Global CD Presence.....	9
2. IMF Training.....	10
3. Histogram Average Course Ratings, FY 2007–13.....	15
4. ICD Training: Average Scores for Basic and Intermediate Courses, FY 2007–13.....	16
5. Response Rates, 2006–11.....	22
6. Follow-Up Survey Results.....	23
7. Follow-Up Survey: Robustness.....	24
8. IMF Triennial Surveys Response Rate, 2003–12.....	31
9. Effectiveness of Past IMF Training.....	32
10. Triennial Survey Results by Agency.....	33
11. Triennial Survey Results by Region.....	34
12. Triennial Survey Results by Income.....	35

Boxes

1. End-of-Course Survey Questionnaire.....	12
2. Follow-Up Surveys: Questions.....	21
3. Triennial Survey Questionnaire: 2009–11.....	30

Appendices

Appendix 1. Data Description.....	39
Appendix 2. End-of-Course Survey Regression Results.....	42

References.....	38
-----------------	----

I. INTRODUCTION

The IMF provides training to its membership in its core areas of expertise as part of its mandate on capacity building.¹ The main objective of this training is to strengthen the analytical and technical capabilities of member country officials to formulate and manage effectively macroeconomic and financial sector policies.² Most of IMF training is offered through the Institute for Capacity Development (ICD)-coordinated program of courses, called the ICD Training Program.³

This paper looks at the methods used to evaluate IMF training activity and examines the resulting data collected over the period 2006–13. The assessment tools for training used during the sample period (2006–13) were subsequently revised over 2015–18 to increase their results focus and introduce course-specific content in the surveys.⁴ Moreover, starting in 2017, the IMF implemented results-based management (RBM) to training.⁵ In addition, since 2015, ICD has undertaken a review of its curriculum and revamped its courses.⁶ Hence this paper provides a detailed analysis of the situation prior to the review.

Evaluation is important because it provides feedback on what works and what does not, and serves to ensure accountability. In general, the overarching message on IMF training that emerges from the various evaluation channels is very positive: participants and their sponsoring

¹ The Fund's core area of competency includes: macroeconomic policy frameworks, balance of payment issues, monetary policy, exchange rate policy, financial stability frameworks, fiscal policy and institutional frameworks, tax and expenditure policy, and macroeconomic and financial statistics. (See IMF 2014 for more details.)

² See the discussion on IMF training strategy in IMF (2008 and 2013).

³ The ICD training program consists of courses delivered by staff of ICD and the regional training centers (RTCs), as well as functional departments (including the Fiscal Affairs (FAD), Monetary and Capital Markets (MCM), Legal (LEG), Statistics (STA), and Research (RES)). The rest of the training offered by the IMF is more specialized and arranged and delivered independently by Fund departments or Regional Technical Assistance Centers (RTACs).

⁴ The revised survey is being piloted and will be administered in 2018.

⁵ RBM assesses two sequential outcome levels: (i) reaction, which measures how participants perceived the training based on self-reported satisfaction. The outcome indicator is a composite satisfaction rating derived from the end-of-course surveys regarding participants' assessments of the value of the course; and (ii) learning, which measures knowledge acquisition using two indicators: the increase in the average post-test score compared to the average pre-test score and the share of participants achieving 60 percent and above on the post-course test. Moreover, on-the-job behavior change is assessed through follow-up surveys sent to participants and their direct supervisors on participants' use of the knowledge/skills taught in the course and participants' use of course-specific forecasting techniques/ tools/models on the job after training (within three to six months).

⁶ For more information, please see "New Common Evaluation Framework for IMF Capacity Development", IMF (2017), and "Technical Note on Assessment Tools for Training Evaluations" (forthcoming).

government agencies value the training. For example, the 2012 triennial survey on IMF training finds an extraordinarily high level of satisfaction (98 percent) with the training programs.

Against this background, the issues addressed in this paper are the following:

- What methods were used to assess the IMF training? In particular, what types of evaluations were conducted and what types of questions were investigated before 2015?
- What key messages emerge from the data gathered through these evaluations? ⁷
- What additional information can be gleaned from the evaluation data, using filters to group the data or demographic information about course participants?

The IMF employs several evaluation instruments for training, which are linked to the well-known Kirkpatrick (1994) evaluation model. In this paper we examine the three-tiered evaluation system (end-of-course surveys, follow-up surveys, and triennial surveys) that was used to gauge participants' and their sponsors' perception of training.⁸ The scope and time frame of each of these surveys is different and used in a complementary way. ICD has long utilized these different techniques to monitor and evaluate its program.⁹ The techniques were developed in consultation with evaluation experts and other training organizations. These surveys are described in the paper.

One of the novel features of this study is its attempt to distill information from all the evaluations over a seven-year period. A second novel feature is combining the evaluation information with participant demographic information. To the best of our knowledge, there has been no previous effort to distill the information presented in the evaluations in such ways before. In 2008, there was a strategic review of training (IMF 2008) that focused on the design of the training curriculum and allocation of training resources. While ICD has systematically gathered information about the effectiveness of training, the current paper is the first effort to link the various data sources and distill lessons. Furthermore, there has been little dissemination of the key messages across the Fund or to member country authorities.

⁷ Lamdany and Stedman (2015) suggest in an IEO evaluation that greater efforts should be made to distill and disseminate lessons learned from these evaluations.

⁸ Other evaluation methods are used to assess the training program. For example, some RTCs hold periodic meetings with heads of training to solicit direct feedback from member country officials, using the results to inform future training. Another example relates to a "tracer study," which was requested by a donor, targeting officials that had attended two or more IMF courses at the now defunct India-IMF training program. More recently, the African Regional training division within ICD has piloted a program to assess learning by administering a quiz at the beginning and at the end of selected courses. The pilot found that "learning" takes place in the courses that were sampled.

⁹ This paper summarizes pre-existing information that has been gathered about IMF training. It does not conduct a separate evaluation to review the effectiveness of training on achieving its objective: to increase participants' skills to contribute to better macroeconomic policymaking as reported in World Bank (2008).

The remainder of this paper is organized as follows: Section II provides a brief description of the ICD training program as of 2014. Section III discusses the end-of-course survey methods and results. Section IV presents the methods and results of the follow-up surveys. Section V gives a summary of the triennial survey, focusing on the surveys questions regarding the effectiveness of training. Section VI offers a brief conclusion of the results and recommends possible channels to fine tune the methods used to assess it. A data appendix describes the data—the sources and content and a second appendix gives more detailed results.

II. ICD TRAINING PROGRAM

This section provides a brief overview of the training program as of 2014, describing the objective of the training program and highlighting the topics and the delivery methods as well as venues of the training program.

The objective of IMF training is to share the expertise of IMF staff in its core areas to help strengthen the skills of officials in member countries to analyze economic developments and formulate and manage macroeconomic and financial sector policies. These topics are practically oriented, focusing on knowledge essential to effective macroeconomic and financial analysis, including statistical and legal frameworks. Typically, these courses provide training that is not available at universities or similar institutions. The training integrates theoretical frameworks and real-world applications, using case studies drawn from Fund experience.

The general approach of the training is to use “hand-on” workshops and real-world policy applications rather than lecturing. Today, most courses are two weeks long and are taught by three IMF economists, often referred to as counselors.¹⁰ Participants are government officials. The selection process for participants varies, depending on the nature of the course. Participants either submit an application for training or are nominated by their organization to attend training.

ICD courses tend to focus on financial programming and policies, macroeconomic management, and finance. Courses taught by ICD account for about 55 percent of training. The Statistics Department provides another sizable share (over 20 percent) of the training. Other IMF departments and regional technical assistance centers (RTACs) provide the rest of the training.

Training is offered in multiple locations around the world and has been increasing over time (Figures 1 and 2).¹¹ Since 2006 the volume of training has increased almost 25 percent to over 12,500 participant weeks in FY 2013. ICD training program continues to adapt existing courses

¹⁰ Typically, the course has three instructors, two full-time IMF staff and one consultant, who is either a former IMF staff member or an academic.

¹¹ The RTCs are the Africa Training Institute (ATI) in Mauritius; the Joint Regional Training Center for Latin America (BTC) in Brazil; the Joint China-IMF Training Program (CTP) in China; the IMF-Middle East Center for Economics and Finance (CEF) in Kuwait; the Joint Partnership for Africa (JPA, formerly the Joint Africa Institute) in Tunisia; the Joint Vienna Institute (JVI) in Austria; and the IMF-Singapore Regional Training Institute (STI) in Singapore. The India Training Program in Pune delivered training from FY 2007 to FY 2013.

and develop new ones to stay attuned to the evolving needs of member country officials and to keep abreast with developments in economic analysis.

As of 2014, there were 4 regional training centers (RTCs): Africa Training Institute; IMF-Middle East Center for Economics and Finance; the Joint Vienna Institute; and the IMF Singapore Regional Training Institute. In addition to these training centers, as of 2014, there were three regional training programs: Joint Regional Training Center for Latin America in Brazil; Joint China-IMF Training Program; and the Joint Partnership for Africa.¹² Training also takes place at the IMF Headquarters, at the IMF RTACs, and at other venues, often in collaboration with other organizations.

The expansion of RTCs has facilitated the training effort, including by providing additional training capacity and enabling trainers to tailor courses to the needs of different regions. Each RTC covers a specific region, delivers courses within the ICD program, and is jointly managed by ICD and co-sponsors.

As a result of the expansion, the regional distribution of training recipients has changed over time. Recently, training to the Middle East and Central Asia region (MCD) has grown with the opening of a center in Kuwait. In FY 2013, the levels of training delivered to Africa (AFR), MCD, and the Asia and Pacific region (APD) were roughly comparable to each other. However, adjusting for the number of countries in the region, training to AFR stood at just three-quarters of training to MCD and APD. The recent establishment of the Africa Training Institute (ATI) in Mauritius is expected to boost the delivery of training to AFR.

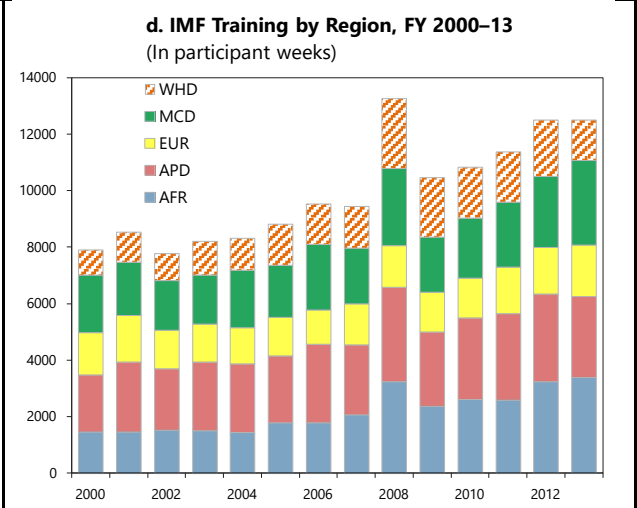
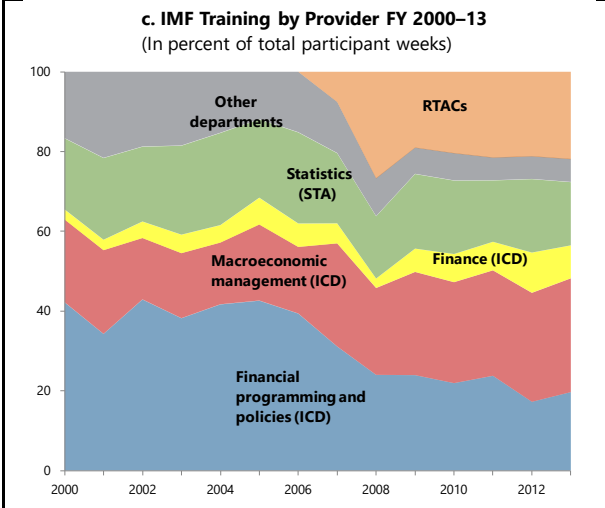
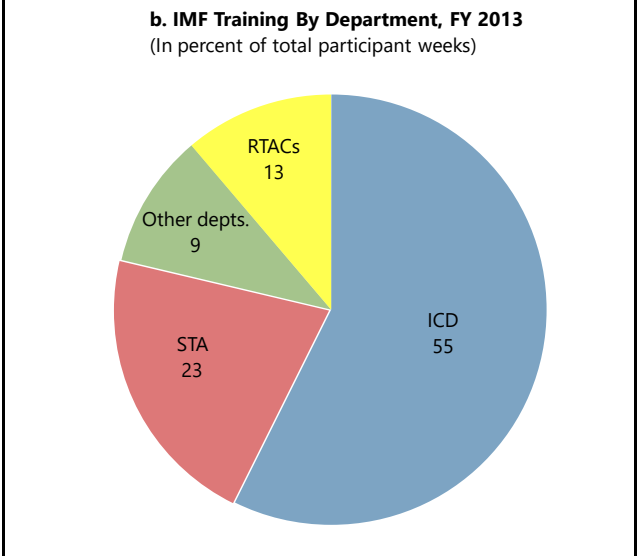
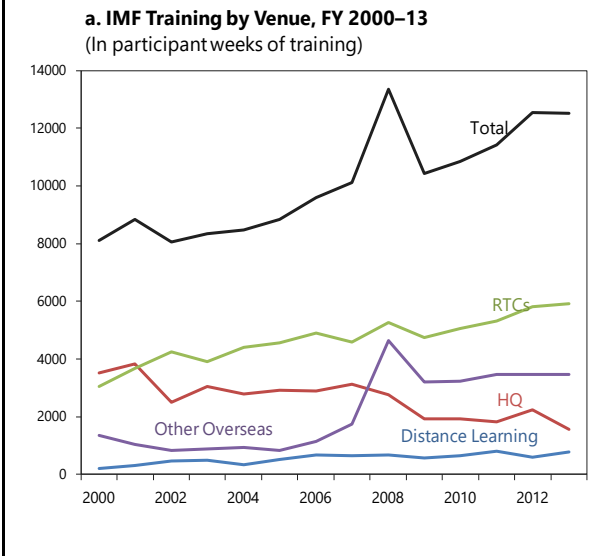
¹² As of 2018, ICD's footprint has expanded to include: The South Asia Regional Training and Technical Assistance Center (SARTTAC). The memorandum of understanding for the Joint Partnership for Africa has expired and not been renewed as of early 2018.

Figure 1. IMF Global CD Presence as of 2014



Source: Institute for Capacity Development.

Figure 2. IMF Training



Source: Institute for Capacity Development.

¹FY 2013 figures are estimates.

III. END-OF-COURSE SURVEYS

This section describes the first tier of the evaluation system: the end-of-course evaluation surveys. It then summarizes the results, focusing on the courses delivered in the ICD training program.

A. Methodology

The end-of-course surveys are conducted after each training course and are based on participants' perceptions of the training received. The results are routinely presented in a memorandum to the management of the department providing the training—a back-to-office report—following the completion of the training course. Historically, there has been no further dissemination of the results or an annual summary of all survey results, partly reflecting the relatively positive responses.¹³

The questionnaire asks the participants to evaluate the training course in three specific areas (Box 1)¹⁴: content or appropriateness of topics covered (section A); presentational methods or the effectiveness of the lectures and workshops (section B); and the overall value of training (section C).¹⁵ In addition, there is usually a fourth section where participants are asked to provide additional comments (section D).

Each section consists of two to four questions. The first three sections contain closed-ended questions. Responses are reported on a 1–5 scale. Specifically, the ratings are: 1 = strong disagreement; 2 = disagreement; 3 = neutral; 4 = agreement; and 5 = strong agreement. Historically, these questionnaires were filled out on paper; however, some courses and training centers are now gathering the information electronically. The response rate to the surveys is close to 100 percent.

¹³ Since 2014, these back-to-office reports are shared in the IMF knowledge exchange.

¹⁴ Box 1 shows a sample questionnaire from a course on macroeconomic diagnostics delivered at JVI in 2012. The format and questions on each survey are broadly similar across courses. Section D sometimes contains additional or somewhat different questions (that would be relevant for the specific course). For example, after a Financial Programming and Policies (FPP) course, participants are asked whether it has improved their understanding of the IMF macro-framework.

¹⁵ The survey distinguishes between the role of lecturer and counselor, with the latter guiding workshop sessions.

Box 1. End-of-Course Survey Questionnaire

Section A. Content

- The topics covered were appropriate.
- The contents of the lectures met expectations.
- The workshops were useful.

Section B. Presentational Methods

- The structure of the course and the mix of presentational methods (e.g., balance of lectures and workshops) were broadly appropriate.
- Lecturers were effective in delivering their presentations.
- Lecturers encouraged discussion and participation.
- Counselors contributed effectively to workshop sessions.

Section C. Value (Recorded in database)

- The knowledge/skills learned during the course will be helpful for your job. (*Skill use*)
- You would recommend the course to others. (*Recommend to others*)
- Overall you were satisfied with the course. (*Satisfaction*)

Section D. Additional Comments

B. The data

The course averages to each question in section C are typically stored in an ICD data system. For the purposes of this paper, these have been aggregated into one dataset.¹⁶ The database contains information on 863 courses taught within the ICD training program during FY 2007–13.¹⁷ Each record contains the course description, identification number, location, dates of delivery, the department teaching the course, and the average rating for each of the three “value” questions in section C of the questionnaire. The information contained in other parts of the survey is used to inform the back-to-office report and mission chiefs about the structure and functioning of the course but are not stored in the database.

About two-thirds (542) of the 863 courses have reported end-of-course survey results (Table 1). Almost all of those were courses delivered by ICD. Very few (28) non-ICD courses have end-of-course survey results stored in the ICD databank. Some departments, such as Finance and Research, have no reports. This difference reflects how the survey results are captured and recorded in the database and not whether they are conducted. However, this has an impact on our analysis going forward. As a result we limit our focus to end-of-course surveys for ICD-delivered courses (514 courses).

¹⁶ As reported below, not all questionnaires have been reported and stored in the data system.

¹⁷ The data prior to 2007 is available in paper format only and have not been included in the current analysis.

**Table 1. ICD Training Program: End-of-Course Surveys
by Department, FY 2007–13**

Department	Courses delivered	Reported Surveys	Percent
ICD	552	514	93
STA	151	17	11
LEG	78	4	5
MCM	36	4	11
FAD	31	3	10
FIN	12	0	0
RES	3	0	0
Total	863	542	63

Source: Institute for Capacity Development.

C. What can we learn from the data?

This section summarizes the main findings that emerge from the 514 end-of-course surveys, examining the robustness by disaggregating the results by time periods, locations, and course types. It also considers basic regressions aimed at explaining differences in ratings.

General findings

Overall, participants rate ICD training quite highly (Table 2). The highest average rating is *recommend to others* (4.8), followed by *overall satisfaction* (4.7). The rating for *knowledge/skills learned* is somewhat lower (4.5). This lower rating may reflect the fact that question targets only the knowledge learned that would be directly useful for one's current job, thereby ignoring other skills acquired. These differences in ratings across survey questions appear to be statistically significant, using t-tests.¹⁸

¹⁸ We perform a paired t-test, comparing the mean ratings of the three questions. The mean rating of *recommend to others* is statistically higher than the mean rating of *skills use*, $t(512) = 36.07$, $p < 0.05$, and also statistically higher than the mean rating of *overall satisfaction*, $t(512) = 22.08$, $p < 0.05$. Furthermore, the mean of *overall satisfaction* is statistically higher than the mean rating of *skills learned*, $t(512) = 15.51$, $p < 0.05$.

Table 2. ICD Training: Average Scores by Fiscal Year,¹ FY 2007–13

Year	Number of courses	Skill use	Recommend to others	Overall satisfaction
Total	514	4.5	4.8	4.7
2007	16	4.5	4.7	4.6
2008	82	4.5	4.8	4.6
2009	81	4.5	4.8	4.7
2010	80	4.5	4.8	4.7
2011	91	4.5	4.8	4.6
2012	95	4.5	4.8	4.7
2013	69	4.6	4.8	4.6

Source: Institute for Capacity Development.

¹ Courses delivered by ICD only.

Robustness of results

We further explore the pooled survey results by analyzing the distribution and by disaggregating the data across time, location, and course types.

First, we find notable variation in the distribution of the ratings (Figure 3). A few courses receive low ratings. For example, in FY 2007 one course was rated around 3.4 for each of the three questions.¹⁹ Figure 3 shows skill use has the largest dispersion of results (based on course averages) while the question of recommend the training to others has the lowest dispersion and highest ratings.

Second, we observe little variation over time (e.g., no time trend). Table 2 shows that there is no discernible difference in ratings over time. The high ratings and the relative ranking of the three questions is the same across all years.²⁰

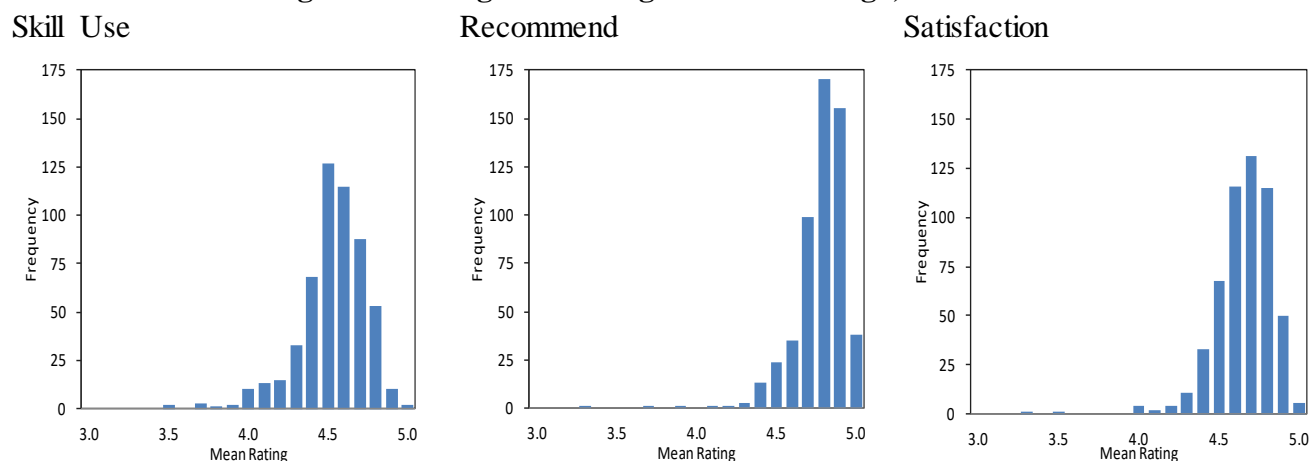
Third, we observe limited variation by delivery venue (Table 3). Small regional differences are noted as the courses in China, Kuwait, and the United Arab Emirates tend to have slightly

¹⁹ The authors reviewed a sample of BTOs of the courses that received the lowest ratings with each providing detailed explanations about the specific special problems encountered with the course. Explanations of these problems included: (i) participants' lacking sufficient language skills; (ii) participants not having technical preparation to receive training; and (iii) course not meeting participants' expectations of coverage.

²⁰ We conducted a paired t-test for each year and found that the ranking of the mean ratings of each of the three categories remained consistent (i.e. *recommend* > *satisfaction* > *skill*) and the differences in mean ratings were statistically significant at $p < 0.05$. In addition, for every two consecutive years, we compared the means of each category. The difference between the mean ratings of the courses in any two consecutive years was not statistically significant (that is, we failed to reject the *null hypothesis* that $skill(2007) = skill(2008)$, $recommend(2010) = recommend(2011)$, etc.

lower average ratings. In addition, courses in these locations accounted for six of the ten lowest rated courses, based on skill use criterion. These observed regional differences may be due to the types of courses offered, language, or other factors.

Figure 3. Histogram Average Course Ratings, FY 2007–13



Source: Institute for Capacity Development

Fourth, we find little widespread variation in the ratings across the basic and intermediate courses (Figure 4).²¹ The ratings for Financial Programming and Policies (FPP—143 course)—a basic level course—follow closely the aggregate ratings. The ratings for intermediate economic courses, such as Macroeconomic Management and Fiscal Policy (MFP) and Management and Financial Sector Issues (MMF—70 courses), we also find broad similarities in the ratings. For Macroeconomic Diagnostics (MDS—39 courses) the ratings on skill use tend to be slightly higher. The ratings for the finance courses are closely aligned with the aggregate. For example, all seven types of courses score between 4.71 and 4.79 in the *recommend to others* category.²²

²¹ In Figure 4, the curriculum is grouped into two major categories and levels. The economic courses are: Basic—Financial Programming and Policies (FPP), Intermediate—Macroeconomic Diagnostics (MDS), Macroeconomic Management and Fiscal Policy (MFP), and Macroeconomic Management and Financial Sector Issues (MMF). The finance courses are: Basic—Financial Market Analysis (FMA), Intermediate—Economic Policies for Financial Stability (EFS), Finance for Macroeconomists (FME), Core Elements of Banking (CBS/BS), Advanced—Macroprudential Policies (MPP), Financial Markets and New Financial Instruments (FMN), and Mortgage Markets and Financial Stability (MSF). For details about IMF training courses, see the training catalog contained in IMF (2014).

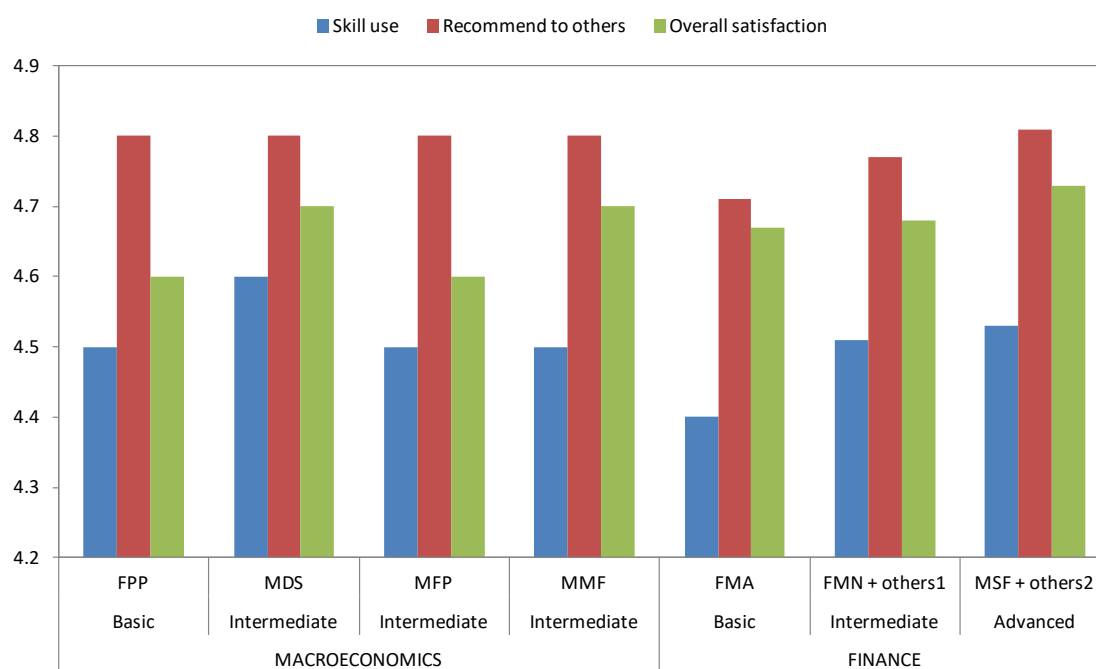
²² Using a paired t-test, we find that the differences in ratings between basic and intermediate courses are not statistically significant at a 5 percent significance level. However, at a 10 percent significance level, the mean rating of basic courses on *satisfaction* statistically lower than the mean rating of intermediate courses, $t(356) = 1.45$, $p < 0.1$.

Table 3. ICD Training: Average Scores Across Locations, FY 2007–13¹

Location ID	Location Name	Number of courses	Skill use	Recommend to Others	Overall Satisfaction
BT	Brazil	27	4.5	4.8	4.7
CE	Kuwait	14	4.4	4.8	4.7
CT	China	16	4.3	4.5	4.6
DL	Distance Learning	23	4.7	4.8	4.7
HQ	Washington, DC	51	4.6	4.8	4.6
IP	India	20	4.5	4.8	4.7
JA	Africa	12	4.6	4.8	4.6
JP	Africa	12	4.5	4.8	4.6
JV	Austria	108	4.5	4.8	4.7
OT	Other	136	4.5	4.8	4.6
RT	UAE and other	20	4.4	4.8	4.6
ST	Singapore	75	4.6	4.8	4.6
	Total	514	4.5	4.8	4.7

Source: Institute for Capacity Development.

¹ Courses delivered by INS/ICD only.

Figure 4. ICD Training: Average Scores for Basic and Intermediate Courses, FY 2007–13¹

Source: Institute for Capacity Development.

1/ BS, CBS, EFS, FFB, FME, FMN.

2/ FMN, MPP, MSF.

Explaining the ratings²³

This section explores whether it is possible to identify factors that may affect the end-of-course survey ratings, by considering demographic data of course participants. Using a simple regression analysis, the general regression specification is:

$$\begin{aligned} rating_score = & \beta_0 + \beta_1 male_{percent} + \beta_2 agency + \beta_3 education + \beta_4 age \\ & + \beta_5 national_{course} + \beta_6 APD + \beta_7 AFR + \beta_8 WHD + \beta_9 EUR \end{aligned} \quad (1)$$

where the dependent variable *rating score* is the average rating (*skill usefulness, recommend to others, overall satisfaction*) received at the end-of-course evaluation. The independent variables are: *male percent*—the percentage of male participants in each course; *agency*—the percentage of central bank participants; *education*—the ratio of participants who have at least a master’s degree; *age*—the mean participant age in the course. Regional dummies (*APD, AFR, WHD, EUR*) are included to explore if the delivery location affects the rating of a course and a dummy for national courses (*national course*) to determine if ratings vary when a majority of course participants were from the same country.²⁴

In the main text we focus on the factors affecting the skill ratings (Table 4). Column 1 reports the results for the entire sample; columns 2–4 reports results from the sub-samples of FPP courses, intermediate macroeconomic courses, and finance courses respectively. In the Appendix, we look at similar regressions for the other two ratings.

We find that the explanatory power of the demographic variables is limited, reflecting low statistical significance. The results show that the share of male participants is statistically significant at the 95 percent level. This suggests that for a given course a higher share of male participants leads to a higher reported skills rating. We also find that participants in national courses tend to report lower skills ratings, possibly reflecting that these participants are more critical or signal the lack of cross-country and peer-to-peer learning. Alternatively, this may be due to having participants for whom the course is less relevant, reflecting the smaller pool of applicants. Regional dummies show that courses delivered in AFR, APD, EUR, WHD received higher skills ratings compared to courses delivered in MCD.²⁵ The overall explanatory power of the regressions is quite low, which is consistent with the limited significance of the explanatory variables.

Comparable results emerge when we look at FPP courses only, with higher percentage of males corresponding to higher skills ratings, and national course participants giving lower skills ratings. The regional dummies, however, were not statistically significant. In the sample of intermediate courses, only the national course variable was statistically significant.

²³ The data used in this section draws on the detail data contained in ICD’s Participant Tracking Systems (PATS)

²⁴ A national course is defined as drawing participants from predominately a single country.

²⁵ These regional dummies are defined consistently with the IMF area departments.

Table 4. End-of-Course Survey Regression Results

Courses	(1)	(2)	(3)	(4)
	Skill All	Skill Basic	Skill Intermediate	Skill Finance
Percentage of Male Participants	0.18* (2.4)	0.33* (2.6)	0.07 (0.6)	0.07 (0.3)
Percentage of CB Participants	0.001 (0.03)	-0.11 (-1.1)	0.12 (1.4)	0.074 (0.4)
Percentage with Masters/Ph.D.	0.02 (0.3)	-0.003 (-0.03)	0.1 (1.1)	-0.13 (-0.8)
Mean Age of Participants	0.005 (1.6)	-0.002 (-0.3)	0.004 (0.8)	0.02 (1.9)
AFR	0.10** (2.7)	0.04 (0.6)	0.09 (1.6)	0.38* (2.6)
APD	0.09** (2.9)	-0.06 (-1.1)	0.13** (2.7)	0.48** (4.8)
EUR	0.12** (3.1)	0.11 (1.5)	0.10 (1.8)	0.53*** (4.4)
WHD	0.12** (3.8)	0.02 (0.4)	0.09 (2.0)	0.47** (4.2)
National Course	-0.12** (-4.3)	-0.18** (-3.6)	-0.17** (-3.8)	-0.10 (-0.5)
Constant	4.15** (32.3)	4.49** (18.9)	4.1** (18.1)	3.3** (7.4)
Observations	458	123	158	69
R^2	0.10	0.28	0.15	0.34
Adjusted R^2	0.09	0.22	0.10	0.24
RMSE	0.19	0.17	0.17	0.19

Note: T- statistics are in parentheses and * and ** indicate the coefficients are significant at the 95 and 99 percent level. The table reports regression results estimated using end-of-course survey data from courses delivered from January 1, 2007 through December 31, 2012. The regression includes regional dummies which indicate where the course was delivered, Africa (AFR), Asia-Pacific (APD), Europe (EUR), Western Hemisphere (WHD), and Middle East and Central Asia (MCD) – which is omitted. Basic economic courses are FPP (Financial Programming and Policies). Intermediate economic courses include Macroeconomic Diagnostics (MDS), Macroeconomic Management and Fiscal Policy (MFP), Macroeconomic Management and Financial Sector Issues (MMF). Finance courses include: Financial Market Analysis (FMA), Economic Policies for Financial Stability (EFS), and Finance for Macroeconomists (FME).

D. Summary

For each ICD course, an end-of-course survey is used to gather the immediate reaction of the participant to the course focusing on: (i) skills acquired; (ii) whether participant would recommend the course to others; and (iii) overall satisfaction. The key message that emerges is that participants are quite satisfied with the training.

In exploring the pooled survey results, we also find:

- Participants respond most positively to the question that they would *recommend the training to others*. In addition, the ranking of the ratings is broadly consistent across all courses with the ordering as follows: *recommend, satisfied, and skill use*.
- There is notable variation in the distribution of the ratings, that is, a few courses receive low ratings in one or more categories. Many back-to-office reports explain these low ratings to be related to “external factors,” such as language deficiencies of the course participants.
- There is slight variation in the ratings when we look at the pooled data in the following ways:
 - across the different course offerings: the basic courses are rated roughly the same as the intermediate and more advanced courses.
 - across the different regions: the regional differences in the ratings were small, with some differences detected in national courses.
 - across the different years: there is little notable time trend.
- The initial regressions, which focus on factors that may explain the variation of the ratings, provided limited additional insights. Three main features of the regressions results are: (i) higher percentage of males in a course leads to higher reported *skills* rating; (ii) participants in national courses tend to be more critical than non-national courses; (iii) regional variations matter for the *skills* rating.

IV. FOLLOW-UP SURVEYS

This section describes the methods used to conduct the follow-up surveys, the second tier of the evaluation system, and summarizes the results.²⁶

²⁶ The follow-up surveys were revised and administered in 2017, with its coverage expanded from about 8 course deliveries per year to all courses delivered by ICD. New questions include to what extent participants applied the knowledge/skills learned during the course, whether the participants used the tools/techniques taught at the course (for tool-focused courses and if primary functions of participants’ work units involve use of these tools), and what factors were barriers or enablers to application of skills and knowledge acquired during the training

Once again, the discussion focuses on courses delivered in the ICD training program.

This evaluation format was introduced in 2006. Its purpose was to address the sustainability of benefits of IMF training as viewed by participants and their sponsoring government agency. This survey type was driven initially by a donor request.²⁷ In 2006, one donor-financed course was selected and a follow-up survey was conducted 12–18 months after the conclusion of the course. Coverage has remained select, focusing on ICD courses.

A. Methodology

The follow-up evaluations are survey based, soliciting the views of the participants and sponsors to assess whether benefits from training are sustained long after the completion of the training. The survey is intended to address whether the courses help participants do their job better and help them in their careers, and whether the knowledge gained in the course is used and shared with others. Courses are selected to reflect the diversity of course categories, donors, and languages of instruction.

The results of each survey are summarized in a memorandum which was periodically shared with executive directors and more detailed reports were shared with donors at steering committee meetings or in project assessments. There is no annual report summarizing the overall findings of these surveys.

The survey

Questionnaires are sent 12–18 months after the completion of the course to both the participants and to the agencies that sponsored their participation in the training (central banks, ministries of finance, and other agencies). Some agencies may sponsor more than one candidate for a course so there may be fewer sponsors than participants for a given course.²⁸

The IMF manages the distribution of the survey and invitations to participate. The survey responses are gathered and compiled by an independent market research firm to ensure anonymity of the responses.

The survey contains six questions, which are the same for both participants and sponsors (Box 2). The responses are on a rating scale of 1–4 (1 = *strongly disagree*, 2 = *somewhat disagree*, 3 = *somewhat agree*, and 4 = *strongly agree*).

There are a few notable differences in the way the results are reported for this survey relative to the end-of-course questionnaire, including: (i) responses are gathered from both participant

²⁷ The initial surveys were conducted for the distant learning (DL) courses. In fact, in 2002 a similar survey was conducted on the pilot DL course, using the same methodology and questionnaire that was designed by the IMF Institute.

²⁸ As a result, their responses for the survey may refer to more than one participant.

and sponsor; (ii) the dataset contains both the average rating and the distribution of responses; and (iii) the rating scale is different.

Box 2. Follow-Up Surveys: Questions

1. The course has helped the participant to do his job better (*Performance*).
2. The course has strengthened the participant's analysis of economic developments and preparation of forecasts (*Analysis*).
3. The course has improved the participant's ability to provide policy advice (*Advice*).
4. The participant has been given added responsibility as a result of this training (*Responsibility*).
5. I believe this course training will give the participant opportunities to advance in his/her career that he/she might not have had without the training (*Career*).
6. The participant has shared with colleagues what he/she has learned in the course (*Share*).

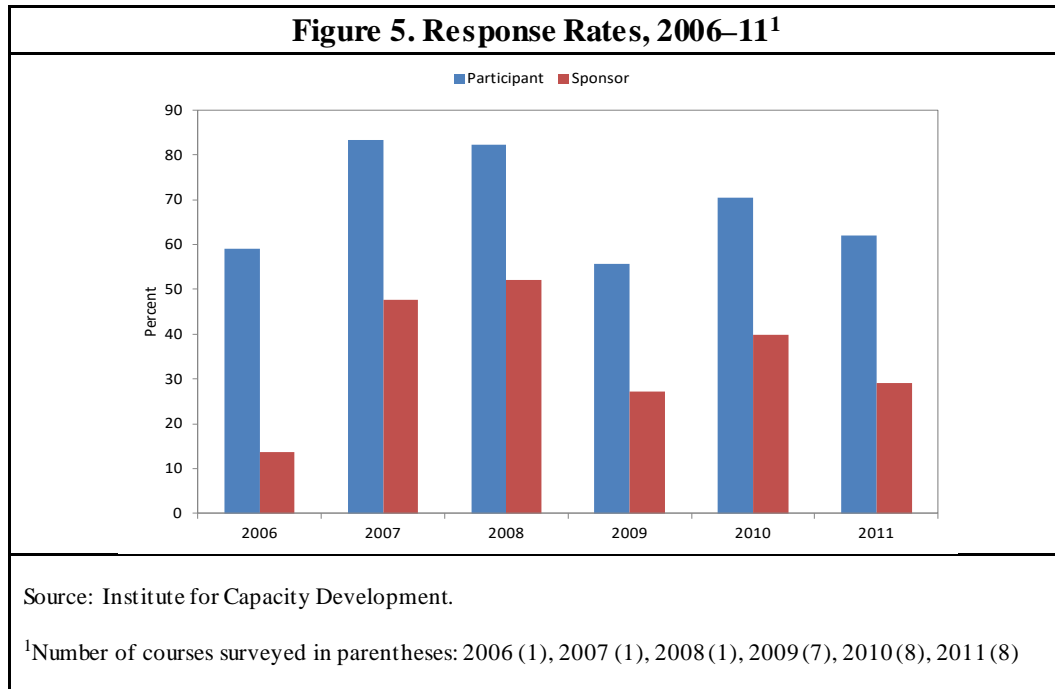
B. The data

At the time of this analysis, the total number of follow-up surveys that have been conducted is 26.²⁹ The first course in the database was taught in 2006. In the subsequent two years, follow-up surveys were conducted for an additional course each year. In 2009, the number of courses surveyed was expanded to seven courses. In both 2010 and 2011, eight courses were surveyed.

The response rate for participants is higher than those for sponsors (Figure 5). For example, in 2010, the average response rate for participants was about 70 percent while the average response rate for sponsors was 40 percent.³⁰

²⁹ In the 26 courses, there were 797 participants coming from 581 agencies, which is a relatively small sample

³⁰ The total number of surveys sent to participants was 260 of which 183 responded. The total number of surveys sent to sponsors was 163 and 65 responded.



C. Results

This section presents the pooled results of the follow-up surveys of participant and sponsoring agency.

General findings

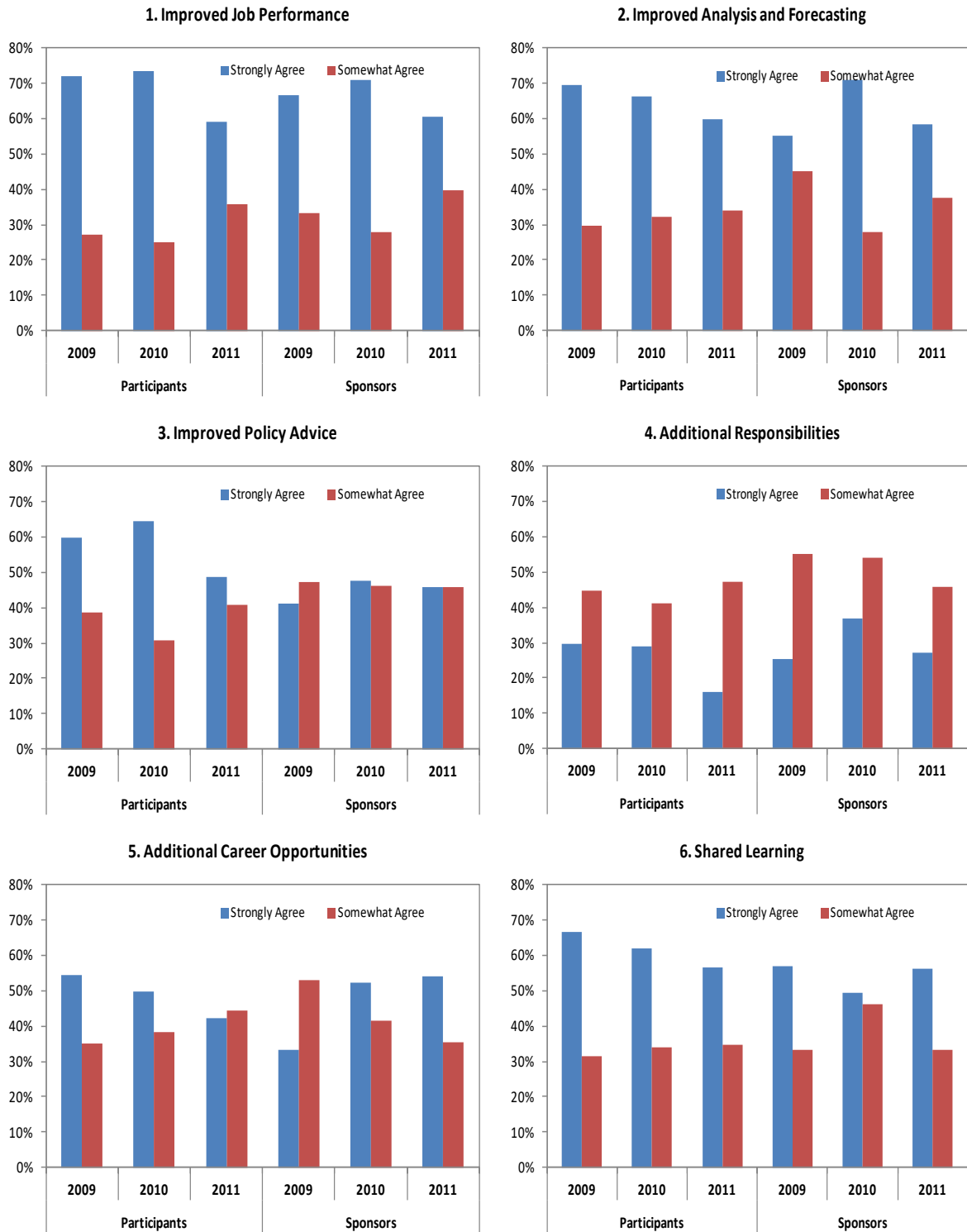
Similar to the results of the end-of-course survey, the general message that emerges from the respondents is that ICD training is perceived to be valuable by participants and sponsors. In particular, both view that the training helped participants do their job better.

The first three questions focus on whether the participant has improved their job performance/analysis/policy advice as a result of the training. Figure 6 shows that most participants strongly agree/somewhat agree (more than 95 percent) that the training has resulted in *improved job performance*, *improved analysis and forecasting*, and *improved policy advice*. This finding is consistent over all the courses over the sample period. The high ratings suggest an improved technical capacity of the participants as a result of the training (as rated by participants).

The responses of the sponsors to the first three questions are overall less positive about impact of the training on their participants, although the ratings are still quite high (Figure 6). In particular, relative to participants, sponsors are less sure that the training has resulted in improved policy advice.

The next question, question 4, focuses on whether training has enhanced the participant's job responsibility. The results are less positive than those on improving job performance. In each of the years with surveyed courses, less than 30 percent of participants strongly agreed with that statement and 25–40 percent disagreed. Sponsors rated *additional responsibilities* markedly lower than any of the other questions, yet they gave a more positive rating than participants did.

Figure 6. Follow-Up Survey Results



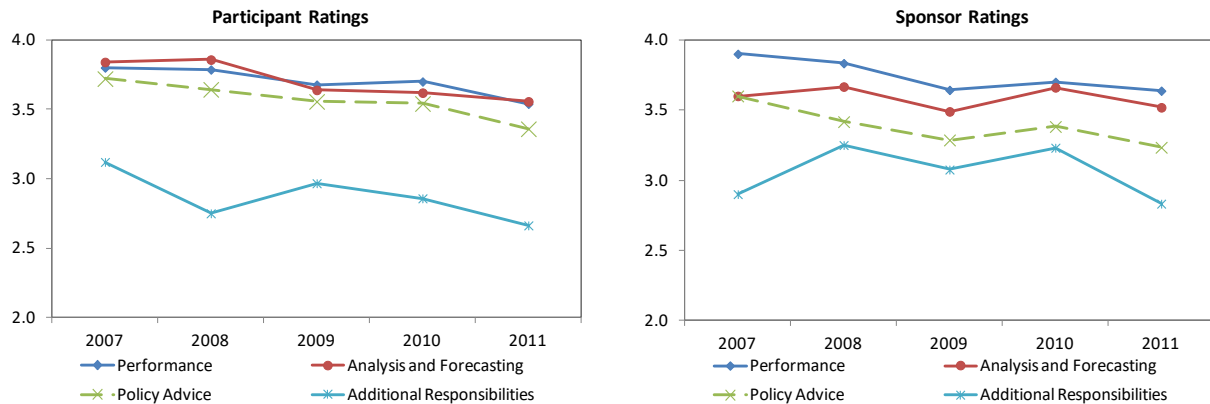
Source: Institute for Capacity Development

The question on *additional career opportunities*, question 5, is also rated quite low. Sponsors are broadly more positive about the career opportunities as a result of the training than participants. The results to this question may be reasonable as attribution to advancement is difficult, depending on many factors beyond IMF training.

The question on whether participants *shared knowledge* they gained in training with their colleagues, question 6, shows that both participants and sponsors agree that the participants regularly share what they learned with other colleagues. Consistent with other responses, participants strongly agreed more than the sponsors.

To assess the robustness of the results, we examined whether there are any noticeable time variation or display of regional differences in the results. Focusing on the first four questions, Figure 7 indicates a slight downward trend. However, we find little variation of ratings across the venues of delivery (see background chart).

Figure 7. Follow-Up Survey: Robustness



Source: Institute for Capacity Development.

Explaining the ratings

A model similar to equation (1) is used to identify factors that may explain the follow-up survey results. However, owing to a small sample size (24 observations), the dummy variables are excluded. The following specification is used:

$$rating_score = \beta_0 + \beta_1 male_{percent} + \beta_2 agency + \beta_3 education + \beta_4 age \quad (2)$$

where *rating score* is the average rating of each of the follow-up survey questions by participants (job performance, analysis and forecasting, policy advice, additional responsibilities, additional career opportunities, shared learning) and the explanatory variables are as described in equation (1).

The results for the econometric model are not very promising (Table 5). There is a positive correlation between gender and the average rating of the course, but gender is significant for

two of the six questions (*Forecasting* and *Responsibility*). The participant's work agency (central bank, ministry of finance) does not seem to influence course rating except for *Shared Learning*. The remaining variables (education, mean age) do not have any statistically significant influence on follow-up survey course ratings.

Table 5. Follow-up Survey Regression Results

	(1) Job Performance	(2) Analysis and Forecasting	(3) Policy Advice	(4) Additional Responsibilities	(5) Career Opportunities	(6) Shared Learning
Percentage of Male Participants	0.25 (0.8)	0.83** (2.9)	0.37 (1.1)	0.30 (0.6)	1.1* (2.7)	-0.32 (-1.3)
Percentage of CB Participants	0.27 (1.0)	0.09 (0.4)	-0.02 (-0.1)	0.74 (1.9)	0.19 (0.6)	0.55** (3.0)
Percentage with Masters/Ph.D.	0.14 (0.7)	0.03 (0.2)	0.23 (1.1)	0.15 (0.5)	-0.06 (-0.2)	0.17 (1.1)
Mean Age of Participants	-0.01 (-0.6)	-0.03 (-1.8)	-0.01 (-0.3)	-0.03 (-0.9)	-0.02 (-0.9)	-0.01 (-0.7)
Constant	3.62** (5.6)	4.09** (7.5)	3.39** (5.0)	3.13** (3.2)	3.37** (4.4)	3.68** (8.0)
Observations	24	24	24	24	24	24
R^2	0.09	0.32	0.10	0.18	0.28	0.43
Adjusted R^2	-0.10	0.18	-0.08	0.01	0.13	0.32
RMSE	0.21	0.17	0.21	0.31	0.25	0.15

Note: T- statistics are in parentheses and * and ** indicate the coefficients are significant at the 95 and 99 percent level. The table reports the regression results estimated using the follow-up survey data for 24 courses delivered between January 2007 and December 2012. In the follow-up survey participants were asked whether the IMF courses had contributed to each of the six work aspects listed above. The responses are on a scale of 1–4, with 4 being the highest possible score.

D. Linking follow-up surveys with end-of-course evaluations

In this section, we juxtapose the participant results of the end-of-course survey with the participants' responses from the follow-up surveys.

The follow-up surveys are linked with their respective end-of-course evaluations, comparing the participants' responses to questions on improved job performance (questions 1–3) with participants' responses to skill use in the end-of-course survey.³¹ Because the two surveys use a different rating scale, we crudely adjust the end-of-course responses by rescaling them.

The first three columns of Table 6 report the pooled follow-up survey results for participants by year. The far right-hand column gives the corresponding end-of-course ratings by

³¹ The course delivered in 2006 does not have end-of-course evaluation and has been excluded.

participants. It is clear that after 12–18 months, participants continued to perceive that IMF training has helped to improve their job performance and their technical skills such as forecasting and policy analysis. This result was less pervasive in the area of policy advice.

Table 6. Follow-Up Surveys and End-of-Course Survey Results, 2007–11

Year	Follow-up			End-of-course
	Improved Job Performance	Improved Analysis and Forecasting	Improved Policy Advice	Skill Use ¹
2007	3.80	3.84	3.72	3.92
2008	3.79	3.86	3.64	3.76
2009	3.68	3.64	3.56	3.75
2010	3.70	3.62	3.54	3.70
2011	3.54	3.56	3.36	3.62

Source: Institute for Capacity Development

¹ End-of-course survey responses have been rescaled by a factor of 0.8.

Next, we examine whether there are any variations across course types (Table 7).³² We report on 18 of the courses, covering five different topics. As reported above, participants in both the end-of-course survey and in the follow-up survey perceived that IMF training helped improve their ‘skills.’ This result appears slightly stronger for the FPP than for the intermediate MMF course. Once again, the response was uniformly lowest for improved policy advice.

³² FPP—Financial Programming and Policies; MDS—Macroeconomic Diagnostics; MF—Macroeconomic Forecasting; MFP—Macroeconomic Management and Fiscal Policy; MMF—Macroeconomic Management and Financial Sector Issues.

Table 7. Follow-up Surveys and End-of-Course Survey Results for Selected ICD Courses

Course Type	Number of Courses	Follow-up			End-of-course
		Improved Job Performance	Improved Analysis and Forecasting	Improved Policy Advice	Skill Use ¹
FPP	8	3.70	3.70	3.00	3.73
MDS	3	3.61	3.48	2.78	3.71
MF	2	3.67	3.58	3.10	3.64
MFP	3	3.69	3.55	2.76	3.65
MMF	2	3.27	3.49	2.56	3.52

Source: Institute for Capacity Development

¹ End-of-course survey responses have been rescaled by a factor of 0.8.

Do end-of-course surveys explain follow-up survey results?

To examine whether the end-of-course surveys explain the follow-up results, we estimate an equation similar to equation (2) and add the end-of course survey variable skill rating to the specification. This is considered for the first three questions of the follow-up survey questionnaire which cover improvements on job performance, analysis and forecasting, and policy advice.

$$rating_{score} = \beta_0 + \beta_1 skill_{rating} + \beta_2 male_{percent} + \beta_3 agency + \beta_4 education + \beta_5 age \quad (3)$$

where *skill_rating* is the skill use rating in the end-of-course survey and the rest of the variables are as previously defined. This regression has a small sample of 24 observations.

The results show that end-of-course ratings on skill are a strong predictor of follow-up survey ratings (Table 8). For the three rating scores (improved job performance, improved analysis and forecasting, improved policy advice) a unit increase in end-of-course rating on skill rating leads to a 0.79 increase in job performance rating, 0.47 increase in analysis and forecasting rating, and a 0.64 increase in police advice ratings. In column 2, gender has a statistically significant effect, with higher proportion of males leading to higher course ratings for the three categories that we look at. The remaining variables are not statistically significant.

Table 8. Follow-Up Surveys and End-of-Course Regression Results

	(1) Improved Job Performance	(2) Improved Analysis and Forecasting	(3) Improved Policy Advice
Skill	0.79** (3.71)	0.47* (2.23)	0.64* (2.50)
Percentage of Male Participants	0.19 (0.73)	0.79** (3.06)	0.32 (1.02)
Percentage of CB Participants	-0.06 (-0.29)	-0.11 (-0.51)	-0.28 (-1.07)
Percentage with Masters/Ph.D.	0.13 (0.81)	0.03 (0.17)	0.23 (1.14)
Mean Age of Participants	-0.002 (-0.16)	-0.02 (-1.61)	-0.0003 (-0.02)
Constant	-0.09 (-0.08)	1.88 (1.70)	0.39 (0.29)
Observations	24	24	24
R^2	0.49	0.47	0.39
Adjusted R^2	0.34	0.33	0.15
RMSE	0.16	0.16	0.19

Note: T- statistics are in parentheses and * and ** indicate the coefficients are significant at the 95 and 99 percent level. The table reports the regression results estimated using the end-of-course and follow-up survey data from the same 24 courses delivered between January 2007 and December 2012. In the follow-up survey participants were asked whether the IMF courses had contributed to each of the six work aspects listed above. The follow-up survey responses are on a scale of 1-4 while the end-of-course responses are on a scale of 1–5.

E. Summary

ICD has conducted 26 follow-up surveys on a sample of courses since 2006 addressing the sustainability of training benefits as viewed by participants and their sponsoring government agency. We find that the overwhelming majority of respondents perceive that the ICD training program is valuable in contributing to developing human capital.

Exploring the pooled results, we find that:

- Participants value the training they receive, perceiving that it has led to improvements in their job performance, analysis, and policy advice.
- Sponsors also value the training provided with their responses being somewhat more muted, particularly in the case of whether the training has improved policy advice.
- Neither participants nor sponsors perceive that the training has led to additional responsibilities. However, sponsors provided a somewhat more positive response to this question relative to participants.
- There is no discernible time trend, regional differences or variation across venue of delivery.

- In linking the participants results of the follow-up survey with the results of the end-of-course survey we observe that:
 - Participants continue to perceive the benefits from the training a year and half after the training was completed. This result is reinforced in our regression analysis: the end-of-course rating on skill use is a good predictor of follow-up survey ratings.

V. TRIENNIAL SURVEYS

This section describes the methods used to conduct the triennial survey, the third tier of the evaluation system, and summarizes the results. The results described below reflect the surveys conducted between 1995 and 2012. The triennial surveys have subsequently been revised.

A. Methodology

The triennial survey has been conducted every three years since 1995. The objective is to gather sponsoring agencies' views of the effectiveness of the ICD training program and information about future training needs.³³ The results of the triennial survey are summarized in a memorandum to management, circulated within the ICD department, and posted on the internal (ICD) and external (IMF) websites.

The triennial survey does not target information about a specific course nor does it solicit the views of participants, setting it apart from the other two evaluation methods. Rather, it seeks an aggregate evaluation from sponsors about the effectiveness of the training program.

Like the follow-up surveys, an external research firm conducts the survey on behalf of ICD to maintain confidentiality. The first survey covered the period 1995–97 and was completed in 1998. The most recent survey was completed in 2012, covering the period 2009–11. It solicited feedback from 579 central banks, ministries of finance, and statistical and other government agencies in 185 countries that had sent participants to the ICD training program.

The survey questions fall into three main categories (Box 3): (i) the effectiveness of past training; (ii) the future course demand by topic; and (iii) future course demand by course delivery mode—such as distance learning, 2-week seminars, etc. The responses are on a scale of 1–4 (1 = strongly disagree; 2 = disagree; 3 = agree; and 4 = strongly agree).

This paper focuses on the questions related to the effectiveness of past training, using the responses from the last three surveys. The four effectiveness questions repeated in each survey are questions 1.1, 1.4–1.6. The other two questions (1.2 and 1.3) were introduced in the 2012 survey.

³³ The 2012 survey also solicited sponsoring agencies views on ways in which IMF could better facilitate their collaboration with IMF country teams and contribute more to successful TA implementation.

Box 3. Triennial Survey Questionnaire: 2009–11

1. **Effectiveness of past training.** Considering our organization's experience during the 2009–11, please check the extent to which you agree or disagree with the following statements. Please select one answer for each statement.

- 1.1 Overall, our organization is satisfied with our IMF Institute training experience (*Satisfied*).
- 1.2 Our staff values IMF Institute training more than training by other providers on similar topics.
- 1.3 IMF Institute training has facilitated our collaboration with the IMF.
- 1.4 IMF Institute training has helped participants do their jobs better (*Improve*).
- 1.5 IMF Institute training has improved the way our staff formulates and implements policy (*Policy*).
- 1.6 IMF Institute participants have been given added responsibilities or promotions as a result of this training (*Promotion*).

2. **Demand for training by delivery mode.** Over the next five years, to what extent do you expect your demand for IMF Institute training to change for each of the following categories of training courses? Please select one answer for each category.

- 2.1 All courses
- 2.2 Short seminars for high-level officials
- 2.3 One-week courses
- 2.4 Two-week courses
- 2.5 Three-week courses
- 2.6 Four and five-week courses
- 2.7 Distance learning from residential segment
- 2.8 Distance learning online only (not offered now)

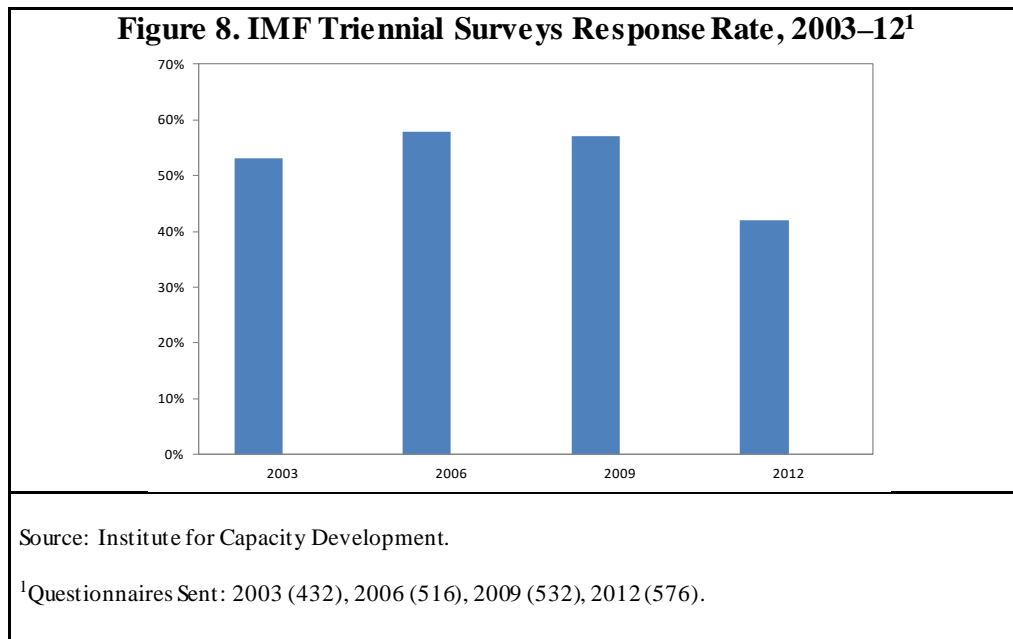
3. **Demand for training by topic.** Over the next five years, to what extent do you expect your demand for courses in the IMF Institute Training Program to change in each of the following categories? Please, select one answer for each category.

- 3.1 Courses on Financial Programming and Policies (FPP)
- 3.2 Other general courses on macroeconomic topics
- 3.3 Courses on macroeconomic statistics
- 3.4 Specialized courses on macroeconomic topics
- 3.5 Specialized courses on financial sector issues
- 3.6 Specialized courses on fiscal topics
- 3.7 Specialized courses on forecasting and macroeconomic modeling
- 3.8 Specialized courses on legal topics

B. The data

Since 1995, six triennial surveys have been conducted. The earlier surveys are not fully comparable with the more recent surveys because of sample and questionnaire design; and hence excluded in the foregoing analysis.³⁴

For this paper, the results of the surveys have been pooled together so we can consider changes in results over time. First, we consider the response rate. Figure 8 shows that response rates have varied over time, ranging between 40 percent and 60 percent. The response rate for the 2012 survey was somewhat lower than previous surveys at 42 percent, which could possibly indicate some dissatisfaction with the services provided.



C. The effectiveness of training: Results

This section summarizes the results on the effectiveness of training for the last three triennial surveys. It also examines the data by comparing the results across regions, agencies, and country income groups.

General findings

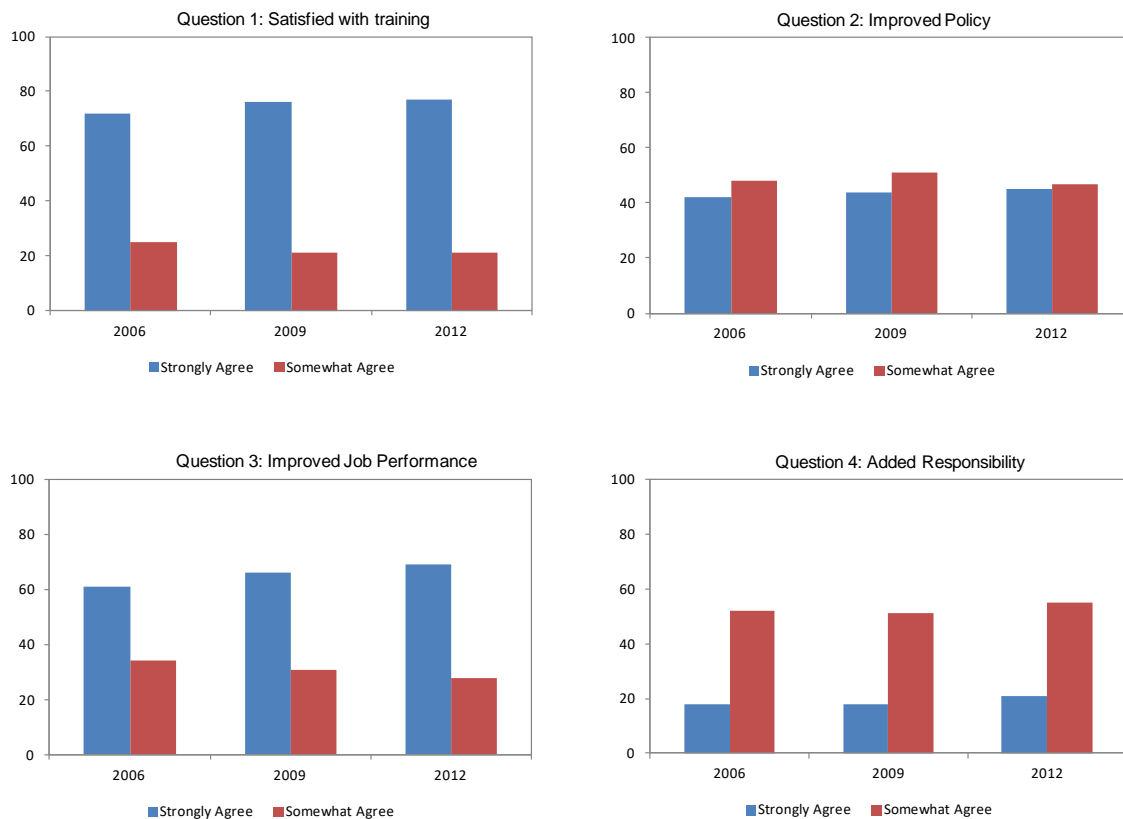
The survey results indicate that sponsoring agencies are generally quite satisfied (strongly or somewhat agree) with their training experience (Figure 9). Nearly 75 percent of the respondents strongly agreed to the question related to overall satisfaction. Both of these results were observed across the different vintages. A majority of the respondents also

³⁴ For example, one methodological difference is that the first surveys were sent to all eligible agencies while since 2003, the surveys were sent only to organizations that had sent officials for training during the reference period.

strongly agree that the training helped improved the participants’ ability to perform their job better. In contrast, the majority of respondents only somewhat agreed that the training improved policy advice or resulted in added responsibilities or promotions of the participants. This finding is consistent across the survey vintages.

The general results from the triennial survey are consistent across the different tiers of the evaluation system. For example, the findings from the triennial surveys (discussed above) are similar to those captured in the follow-up survey. In particular, for both the triennial survey and the follow-up surveys the lowest average rating is to the question on whether training contributed to additional responsibilities or promotions of the participant.

Figure 9. Effectiveness of Past IMF Training



Source: Institute for Capacity Development.

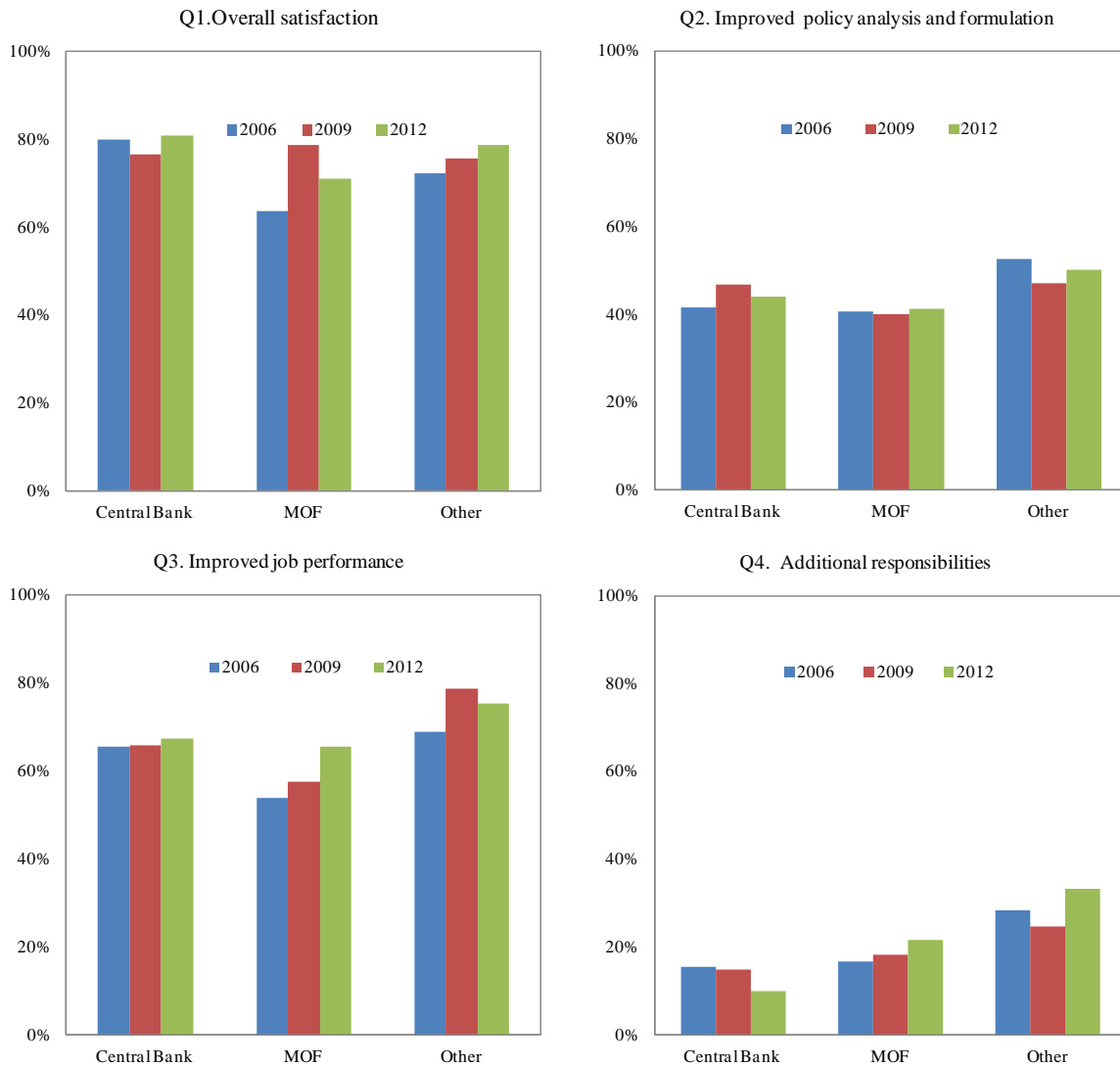
Robustness of results

The answer to each of the questions in Figure 9 can also be broken down by government agency, region, and country income classification to gain further insights. We probe the results looking for more detailed lessons by focusing on the strongly agreed responses rather than all favorable responses to explore possible differences.

Across agencies

Some variation by agency type can be detected in *strongly agreed* responses. The data are split into 3 agency groups: central banks (CB), ministries of finance (MoF), and other agencies (Other) (Figure 10). In general, the responses of CB tend to be more favorable relative to the MoF, especially for the first three questions. Another feature is that the responses of other agencies tend to be more relatively favorable for questions 2–4. There is broad agreement across the agency types that training has not resulted in participants being given added responsibilities.

Figure 10. Triennial Survey Results by Agency
(Percentage of respondents that *Strongly Agree*)



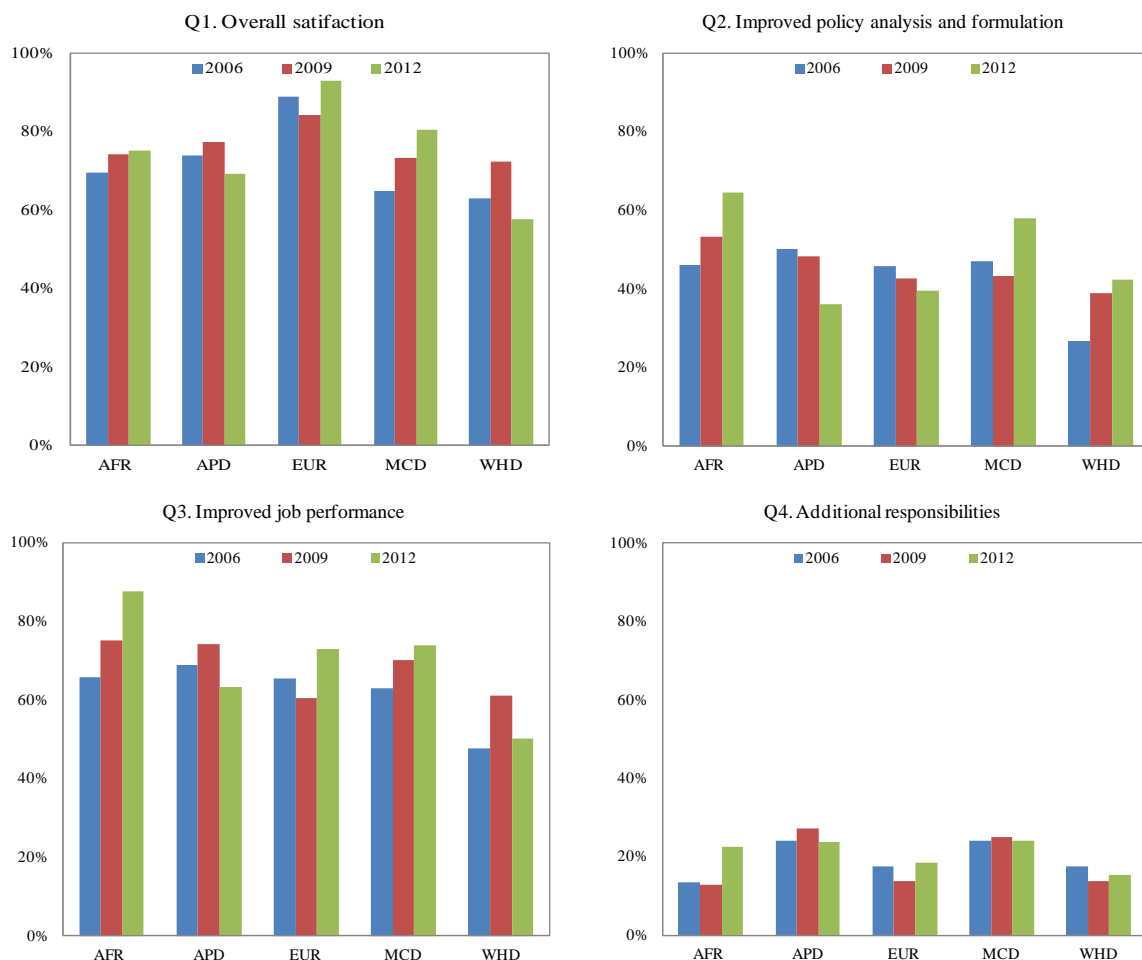
Source: Institute for Capacity Development.

^{1/}Q1. Overall, our organization is satisfied with our IMF Institute training experience. Q2. IMF Institute training has improved the way our staff formulates and implements policy. Q3. IMF Institute training has helped participants do their jobs better. Q4. IMF Institute participants have been given added responsibilities or promotions as a result of this training.

Across regions

Some regional differences emerge when considering the *strongly agreed* responses (Figure 11). For question 1 (overall satisfaction with ICD training), the ratings are generally quite strong. However, the most favorable responses were amongst agencies from the European region (EUR). This feature of the data holds for all three years of the survey. For questions 2 and 3, (on improved job performance as well as policy analysis and formulation) we observe a larger percentage of strongly agreed responses from agencies in Africa. This result has become stronger over time. Finally, we detect little regional variation on the question whether training helps provide additional responsibilities.

Figure 11. Triennial Survey Results by Region^{1/}
(Percentage of respondents that *Strongly Agree*)



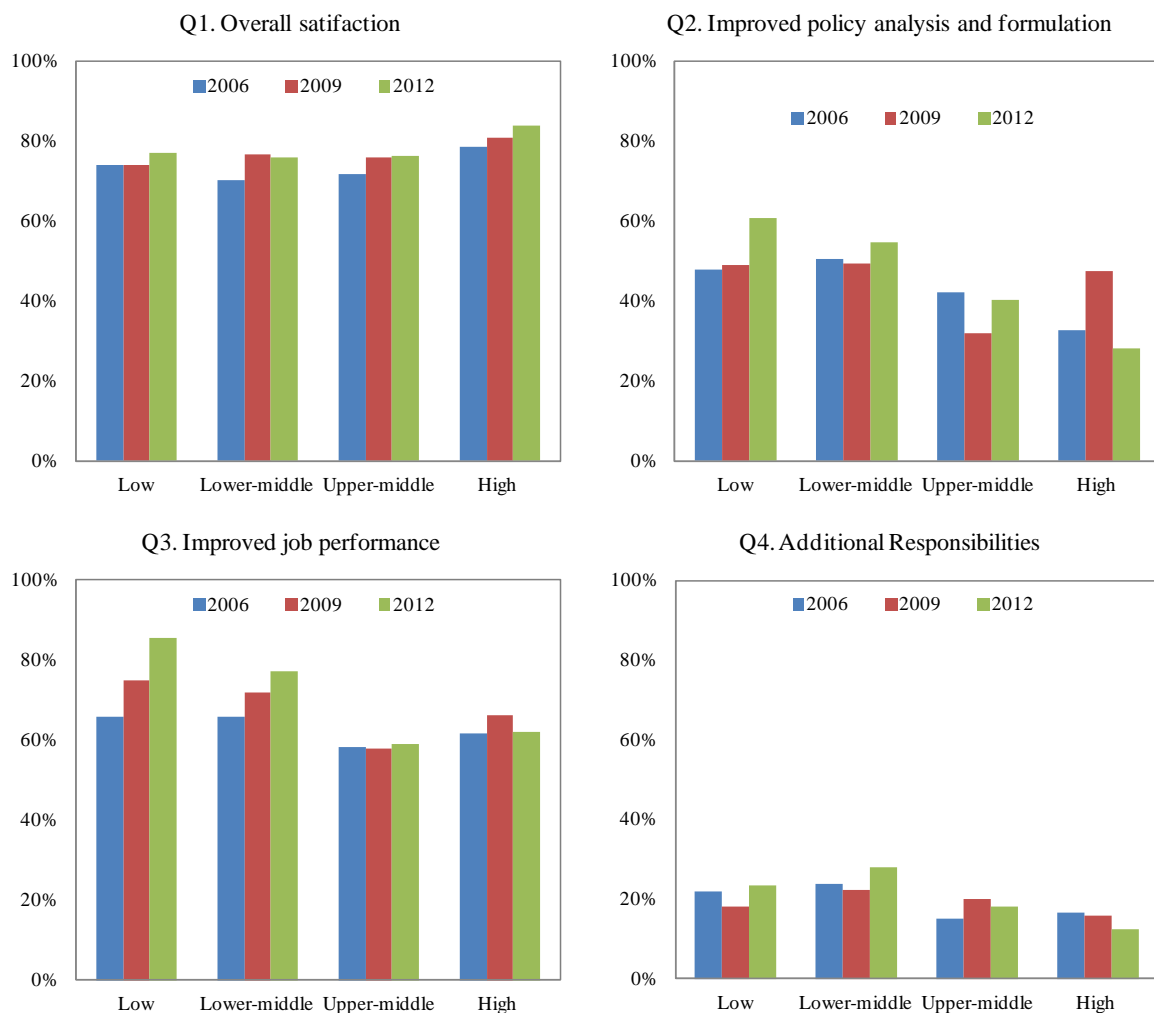
Source: Institute for Capacity Development.

^{1/} Q1. Overall, our organization is satisfied with our IMF Institute training experience. Q2. IMF Institute training has improved the way our staff formulates and implements policy. Q3. IMF Institute training has helped participants do their jobs better. Q4. IMF Institute participants have been given added responsibilities or promotions as a result of this training.

Results across income

Small differences in the percentage of respondents that strongly agree are observed when cutting the data by income levels. In this view, countries are split into four groups using the World Bank classification: High income, upper-middle income, low-middle income, and low income. The most striking differences are: (i) the respondents from high income countries respond are relatively more favorably to the question on overall satisfaction; (ii) the respondents from low-income and low-middle income countries tend to strongly agree that the training improves job performance and participants' policy formulation; and (iii) little variation on the question whether training helps add additional responsibilities.

Figure 12. Triennial Survey Results by Income^{1/}
(Percentage of respondents that *Strongly Agree*)



Source: Institute for Capacity Development.

^{1/}Q1. Overall, our organization is satisfied with our IMF Institute training experience. Q2. IMF Institute training has improved the way our staff formulates and implements policy. Q3. IMF Institute training has helped participants do their jobs better. Q4. IMF Institute participants have been given added responsibilities or promotions as a result of this training.

D. Summary

A triennial survey of government agencies that have received IMF training has been carried out by ICD since 1995. The results of this survey are very positive, including a 97 percent overall satisfaction rating.

A few notable findings emerge:

- Sponsors broadly agreed that they were satisfied with the training experience, and that the training helped participants do their job better by improving the way staff formulates and implements policy.
- Consistent with the other two questionnaires, the survey found that few participants were given promotions or added responsibility as a direct result of the training.
- There are some notable differences in responses across agencies and regions. For example, central banks tend to rate the training higher than the ministries of finance, and African sponsoring agencies tend to rate the effectiveness of training on improving job performance more strongly.

VI. CONCLUSIONS

This paper describes the main methods that ICD employed to evaluate its training program from 2006–13. These methods are: end-of-course surveys, follow-up surveys, and triennial surveys. The paper provides a brief description of the training program and then attempts systematically to distill information from the three evaluation products.

An important message that emerges from all three surveys is that the ICD training program is well-liked across both participants and sponsoring agencies. Responses to questions referring to overall satisfaction were rated highest. Similarly, all the surveys reveal that, in general new career opportunities or additional work responsibilities are not the main reason participants take ICD training, and thus rank it slightly lower compared to the other categories.

Additional results include:

- There are some geographical differences as revealed in end-of-course evaluations and the triennial survey. For example, courses taught in China and Middle East tend to have lower rating by participants relative to similar courses taught across other regions.
- In the follow-up survey, the results suggest there are different views of the ICD program among participants and the managers of the agencies that sponsored them. Specifically, managers typically had a less favorable view of the ICD program and its impact than the participants themselves. While there are many possible explanations for this, one explanation may be insufficient outreach or participation by managers (higher ranking officials) in the training program. An alternative explanation is that the sponsors in general are not the direct supervisors of the participants, therefore, they are not really in a

position to observe participants' on-the-job application of the knowledge and skills taught in the course. They may tend to give more moderate ratings (away from extremes) as they are not well informed.³⁵

- At present this three-tiered survey system (end-of-course, follow-up, and triennial) helps gauge participants' and their home agencies' perceptions of ICD training. The evaluations provide information about what works and what does not. It seems that greater distillation of the information might provide greater feedback to a variety of audiences which include staff, donors, and other stakeholders.

Looking ahead, more work on evaluating the effectiveness of IMF training might be considered covering three broad areas:

- **Document existing evaluation information.** Data on individual responses for ICD course are available and can be organized and analyzed. Efforts are currently being developed as described in Edison and Jabri (2015). Evaluation data on training conducted outside of ICD should be included in future analysis. Most training courses taught by other departments use end-of-course questionnaires, but their results are not in the public domain and hence could not be analyzed. Efforts should be made to gather this information historically and then establish an ongoing process for storing this information.
- **Revise the tools.** Our analysis points to a robust evaluation framework that needs to be more systematically utilized. Since our initial analysis, ICD has introduced a new common evaluation framework, which, among other things, implements RBM for training that will be routinely used to monitor the effectiveness of training.
- **Conduct case study evaluations.** Design an evaluation of a sample of countries that have received a lot of training to assess the extent to which IMF training has had an impact on the capacity of target institutions and organization.

³⁵ One of the revisions ICD introduced for the follow-up surveys is to request the information on the participant's direct supervisor in the application form and then ask him/her to fill out the survey after a much shorter time after the completion of the course (rather than previous 12-18 months lag).

References

- Edison, Hali J. and Ranae Jabri. 2015. "IMF Training: Does it Build Capacity?" forthcoming (Washington: International Monetary Fund).
- International Monetary Fund. 2008. "Training as Part of Capacity Building—Recent Initiatives and Strategic Considerations," SM/08/114, (Washington: International Monetary Fund).
- _____. 2013. "The Fund's Capacity Development Strategy—Better Policies through Stronger Institutions—Background Paper" SM/13/129, Supplement 1 (Washington: International Monetary Fund).
- _____. 2014. "IMF Policies and Practices on Capacity Development," (Washington: International Monetary Fund).
- _____. 2017. "New Common Evaluation Framework for IMF Capacity Development," (Washington: International Monetary Fund).
- _____. (forthcoming). "Technical Note on Assessment Tools for Training Evaluations," (Washington: International Monetary Fund).
- Kirkpatrick, Donald L. 1994. "Evaluating Training Programs: The Four Levels," (San Francisco: Berrett-Koehler).
- Lamdany, Ruben and Louellen Stedman, 2015, *Self-evaluation at the IMF: An IEO Assessment*, (Washington: International Monetary Fund).
- World Bank Independent Evaluation Group. 2008. "Using Training to Build Capacity for Development: An Evaluation of the World Bank's Project-Based and WBI Training," (Washington: The World Bank).

Appendix 1 Data Description

This data appendix describes the various data sources used in the paper.

Training data

Training is defined to include all training events within the ICD program. Training does not include single-country events that are delivered as part of a TA engagement.

ICD stores its training data in a central repository called Participant and Tracking System (PATS). It contains information on economic training delivered by the Fund to member country officials. The database contains information on courses starting in 1965 and participants starting in 1981. The extensive information on ICD program courses and participants includes course dates, delivering departments, names of participants and their educational background and agency affiliation.

Participants to ICD training are determined differently depending on the nature of the course. Participants either submit an application for training or are nominated by their organization to attend training. For the analysis in this paper a participant includes all officials who attended the course, independent on how they were selected.

End-of-course survey

The end-of-course survey data (see Box 1 for details) have been aggregated into a dataset that contains information on about 863 courses taught within the ICD training program during FY 2007–13.³⁶

Each record in the dataset contains the course description, identification number, location, dates of delivery, and the department teaching the course. The dataset contains the average rating on the three “value” questions in section C of the questionnaire. The questions relate to: (i) the knowledge/skills learned during the course will be helpful for your job (*Skill use*); (ii) whether participant would recommend the course to others (*Recommend to others*); and (iii) overall satisfaction with the course (*Satisfaction*). The responses are on a 1–5 rating scale. The ratings are: 1 = strong disagreement; 2 = disagreement; 3 = neutral; 4 = agreement; 5 = strong agreement.

Follow-up survey

Twelve to eighteen months after the completion of the course ICD conducts a survey of participant and sponsors. ICD has partnered with an independent research organization to conduct the survey on its behalf to a select sample of training courses starting in 2006. The same questionnaires are sent to participants in the selected courses and to the managers in

³⁶ The data on end-of-course evaluation prior to 2007 is available in paper format only and have not been included in the current analysis.

their agency who sponsored their participation. To ensure confidentiality, the responses are reported to the independent research organization, which transmits the data set to ICD removing traces of identification.

Survey questions include: whether the training had helped the participant do his/her job better; whether it strengthened the participant's analysis of economic developments and preparation of forecasts; whether it improved the participant's ability to provide policy advice; whether the participant receive added responsibility and/or opportunity for career advancement; and whether the participant's learning had been shared with others.

Table 1A lists the courses that have been surveyed; the number of participants and sponsors and the number of their responses to the survey.

Each record in the database contains a course type, identification number, course category, donor name, number of participants contacted, number of participant responses, number of sponsors contacted, number of sponsor responses and the number of responses of each rating by participant and sponsor.

Triennial survey

ICD has partnered with the same independent research organization to conduct a study of the ICD training program every three years. This survey includes training by the IMF Institute and other IMF departments. The questionnaire (see Box 3 for details) aims to collect information on two issues from all organizations that have sponsored participants in training during the three years prior to the survey. The two issues are on effectiveness of past training and demand for future training in terms of delivery mode and topic. The responses are on a scale of 1–4 (1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree).

Organizations were categorized by the IMF's regional departments (Africa, Asia/Pacific, Middle East and Central Asia, and Western Hemisphere). The sample was also classified by agency type, specifically by central banks, ministry of finance, and other agencies.

Table A1. Donor Course Survey Courses, CY 2006–11¹

Year	Course ID	Course Name	Participants Contacted	Part. Response	Sponsors Contacted	Sponsor Respons
2006	DL06.37/8	DFID FPP	39	23	22	
2007	DL07.37/8	DFID FPP	30	25	21	1
2008	DL08.37/8	DFID FPP	34	28	23	1
2009	JV09.06	MERP	29	14	23	1
2009	JV09.19	MFS	23	8	22	
2009	ST09.05	MDS	31	18	31	1
2009	ST09.06	MF	32	18	28	1
2009	HQ09.09	FPP	35	16	35	
2009	IP09.03	FPP	25	16	25	
2009	DL09.37/8	DFID FPP	30	24	23	
2010	ST10.17	EFS	32	21	32	1
2010	OT10.18	MMF-S	34	23	0	
2010	OT10.02	MFP	36	21	0	
2010	OT10.07	MMNR	30	18	30	
2010	JV10.26	MF	30	22	30	1
2010	IP10.05	MMF	23	17	23	1
2010	DL10.36	FPP-F	41	30	41	1
2010	OT10.19	FPP-R	34	31	7	
2011	ST11.11	FPP	31	22	31	
2011	OT11.24	BS-S	31	19	0	
2011	JV11.01	MFP-R	30	20	28	1
2011	HQ11.05	FME	36	25	35	1
2011	OT11.25	MFP	21	9	14	
2011	IP11.08	FMA	27	17	18	
2011	CE11.09	MDS	23	13	13	
2011	OT11.11	MDS	30	17	26	

Source: Institute for Capacity Development

1/ Course mnemonics are: Banking Supervision (BS), Economic Policies for Financial Stability (EFS), Financial Market Analysis (FMA), Finance for Macroeconomists (FME), Financial Programming and Policies (FPP), Macroeconomic Diagnostics (MDS), Monetary and Exchange Rate Policy (MERP), Macroeconomic Forecasting (MF), Macroeconomic Management and Fiscal Policy (MFP), Monetary and Financial Statistics (MFS), Macroeconomic Management and Financial Sector Issues (MMF), Macroeconomic Management and Natural Resource Management (MMNR)

Appendix 2 End-of-Course Survey Regression Results

This appendix presents the regression specification similar to equation (1) and Table 4, focusing on the other two ratings: *recommend to others*, and *overall satisfaction*.

Table A2 reports the results for the rating *recommend to others*. Like the results in the main text on *skill use*, four different samples of courses are considered: the whole sample, basic macroeconomic, intermediate macroeconomic, and finance. For the whole sample (column 1), the demographic variables summarizing the course participants characteristics are not statistically significant and the overall explanatory power is quite low. The only regressor of any significance is the dummy variable indicating whether the course was a “national course.” By restricting the focus of the sample of courses to a select group there is noticeable improvement in the explanatory power of the courses, especially for the FPP courses and the finance course. For the sample limited to FPP courses, we find that central bank participants as well as participants with Master’s and Ph.D. degrees are less likely to give as strong a rating as depicted by the negative coefficients on these variables. Similar to the finding for *skills* ratings, participants in national courses are less likely to recommend the course to others. Another notable result for the finance courses sample: courses offered outside of the Middle East and Central Asia region tend to give higher ratings.

Table A3 provides the results for the rating of *overall satisfaction* with the presentation following Table A2. Column 1 gives the results for the entire sample. For this regression, two coefficients on demographic variables—education and age—are small but statistically significant.³⁷ Also, the regional dummies are positive and statistically significant for APD, EUR, and WHD. The overall explanatory power of the regression is quite lower. The results for the narrower samples of economic courses are quite similar. For the finance courses, the regional dummies and gender of participants are statistically significant.

³⁷ Neither of the coefficients suggests that these are economically significant. A 10 percent increase in the number of participants with Masters/Ph.Ds. leads to a decline in overall satisfaction by 0.01 points while an increase in mean age by five years results in a 0.03-point increase in course satisfaction.

Table A2. End-of-Course Survey Regression Results: *Recommend to Others*

Courses	(1) Recommend All	(2) Recommend Basic	(3) Recommend Intermediate	(4) Recommend Finance
Percentage of Male Participants	-0.05 (-0.8)	0.10 (1.2)	-0.03 (-0.3)	-0.53* (-2.3)
Percentage CB Participants	-0.07 (-1.9)	-0.26** (-3.6)	0.01 (0.2)	0.19 (0.9)
Percentage with Masters/Ph.D.	-0.06 (-1.3)	-0.23** (-3.5)	0.01 (0.2)	-0.18 (-1.0)
Mean Age of Participants	0.004 (1.6)	-0.0003 (-0.08)	0.006 (1.3)	-0.01 (-1.0)
AFR	0.05 (1.8)	0.02 (0.5)	-0.03 (-0.6)	0.34* (2.3)
APD	-0.03 (-1.3)	-0.08* (-2.2)	-0.06 (-1.5)	0.27** (2.8)
EUR	0.05 (1.6)	0.05 (1.0)	-0.004 (-0.1)	0.34* (2.8)
WHD	0.03 (1.3)	-0.01 (-0.3)	-0.04 (-1.0)	0.45** (4.1)
National Course	-0.06* (-2.5)	-0.13** (-3.6)	-0.05 (-1.5)	-0.10 (-0.5)
Constant	4.74** (46.0)	5.06** (30.5)	4.62** (26.3)	5.16** (11.5)
Observations	458	123	158	69
R^2	0.09	0.42	0.07	0.36
Adjusted R^2	0.07	0.37	0.02	0.26
RMSE	0.15	0.12	0.13	0.19

Note: T- statistics are in parentheses and * and ** indicate the coefficients are significant at the 95 and 99 percent level. The table reports regression results estimated using end-of-course survey data from courses delivered from January 1, 2007 through December 31, 2012. The regression includes regional dummies which indicate where the course was delivered, Africa (AFR), Asia-Pacific (APD), Europe (EUR), Western Hemisphere (WHD), and Middle East and Central Asia (MCD)—which is omitted. Basic economic courses are FPP (Financial Programming and Policies). Intermediate economic courses include Macroeconomic Diagnostics (MDS), Macroeconomic Management and Fiscal Policy (MFP), Macroeconomic Management and Financial Sector Issues (MMF). Finance courses include: Financial Market Analysis (FMA), Economic Policies for Financial Stability (EFS), and Finance for Macroeconomists (FME).

Table A3. End-of-Course Survey Regression Results: Overall Satisfaction

Courses	(1) Satisfaction All	(2) Satisfaction Basic	(3) Satisfaction Intermediate	(4) Satisfaction Finance
Percentage of Male Participants	-0.11 (-1.6)	-0.03 (-0.3)	-0.09 (-0.9)	-0.65** (-2.8)
Percentage of CB Participants	0.03 (0.6)	-0.122 (-1.3)	0.10 (1.4)	0.204 (0.2)
Percentage with Masters/Ph.D.	-0.12* (-2.4)	-0.2* (-2.4)	-0.14 (-1.8)	-0.19 (-1.1)
Mean Age of Participants	0.006* (2.1)	-0.008 (-1.5)	0.007 (1.5)	-0.01 (-0.9)
AFR	0.05 (1.4)	0.02 (0.3)	-0.04 (-0.9)	0.27 (1.8)
APD	0.06* (2.0)	-0.03 (-0.61)	0.05 (1.1)	0.27* (2.6)
EUR	0.18** (5.1)	0.13* (2.0)	0.16** (3.3)	0.35** (2.8)
WHD	0.08** (2.7)	0.002 (0.1)	0.05 (1.4)	0.41** (3.6)
National Course	0.01 (0.4)	-0.03 (-0.8)	0.01 (0.4)	-0.18 (-0.9)
Constant	4.52** (39.4)	5.12** (24.2)	4.46** (23.1)	5.15** (11.2)
Observations	458	123	158	69
R^2	0.1	0.18	0.17	0.38
Adjusted R^2	0.08	0.12	0.12	0.29
RMSE	0.17	0.15	0.15	0.19

Note: T- statistics are in parentheses and * and ** indicate the coefficients are significant at the 95 and 99 percent level. The table reports regression results estimated using end-of-course survey data from courses delivered from January 1, 2007 through December 31, 2012. The regression includes regional dummies which indicate where the course was delivered, Africa (AFR), Asia-Pacific (APD), Europe (EUR), Western Hemisphere (WHD), and Middle East and Central Asia (MCD) —which is omitted. Basic economic courses are FPP (Financial Programming and Policies). Intermediate economic courses include Macroeconomic Diagnostics (MDS), Macroeconomic Management and Fiscal Policy (MFP), Macroeconomic Management and Financial Sector Issues (MMF). Finance courses include: Financial Market Analysis (FMA), Economic Policies for Financial Stability (EFS), and Finance for Macroeconomists (FME).