

**Annual Review of World
Bank Staff Learning FY04**

Cristina Ling Chard

Diana J. Arango

Shreyasi Jha

**Annual Review of World
Bank Staff Learning FY04**

Cristina Ling Chard
Diana J. Arango
Shreyasi Jha

WBI Evaluation Studies
No. EG05-104

The World Bank Institute
The World Bank
Washington, D.C.

December 2004

ACKNOWLEDGEMENTS

This evaluation report was prepared for the Learning Board under the overall guidance of Nidhi Khattri, Acting Manager, Institute Evaluation Group. The report was discussed by the Learning Board at a meeting chaired by Ms. Phyllis Pomerantz, Chief Learning Officer on November 1, 2004.

This report would not have been possible without the careful and hard work of Maria Beatriz Camargo Cardenas, and Paul Date. The authors also gratefully acknowledge the special efforts by Moria Deborah Sutherland for providing IEG with learning catalog data and Humberto S. Díaz for administering evaluation questionnaires and formatting the final report. Finally, we especially thank Marlaine Lockheed for her thoughtful review.

WBI Evaluation Studies are produced by the WBI Evaluation Group (IEG) to report evaluation results for staff, client, and joint learning events. An objective of the studies is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the name of the author and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the author and do not necessarily represent the view of the World Bank Group. WBI Evaluation Studies are available on line at:

<http://info.worldbank.org/etools/WBIEG/publications/index.cfm?pg=Home&Intro=yes>

Vice President, World Bank Institute	Ms. Frannie Léautier
Chief Learning Officer	Ms. Phyllis Pomerantz
Acting Manager, Institute Evaluation Group	Ms. Nidhi Khattri
Task Team Leader	Ms. Cristina Ling Chard

ACRONYMS

ACT	Accounting Department
ACS	Administrative and Client Support
AFR	Africa Region
ASTD	American Society for Training and Development
DEC	Development Economics
DL	Distance Learning
EAP	East Asia and Pacific Region
ECA	Europe and Central Asia Region
ESSD	Environmentally and Socially Sustainable Development Vice Presidency
ETC	Extended Term Consultant
ETT	Extended Term Temporary
EXT	External Affairs
FSE	Financial Sector
FY02	Fiscal Year 2002
FY03	Fiscal Year 2003
FY04	Fiscal Year 2004
F2F	Face to Face
GA-GD	Staff in grades A through D
GE	Staff in grade E
GF	Staff in grade F
GG	Staff in grade G
GH	Staff in grade H
GSD	General Services Department
HDN	Human Development Network
HLM	Hierarchical Linear Model
HQ	Headquarters
HRS	Human Resources
ICC	Intra-class Correlation
IEG	Institute Evaluation Group
IFC	International Finance Corporation
ISG	Information Solutions Group
INF	Infrastructure Network
IT	Information Technology
JPA	Junior Professional Associate
K&S	Knowledge and Skills
LCR	Latin America and the Caribbean
LP	Learning Plan
LEG	Legal
LMS	Learning Management System
MIGA	Multilateral Investment Guarantee Agency
MENA	Middle East and North Africa

MGT	Management
OLP	Other Learning Providers
OPCS	Operational Policy and Country Services
PI	Process and Interaction
P&T	Professional and Technical
PREMDEC	Poverty Reduction and Economic Management - Development Economics
PSD	Private Sector Development
PSI	Private Sector Development and Infrastructure
RM	Resource Management
SAR	South Asia Region
SEC	Corporate Secretariat
SLC	Strategic Learning Center
STC	Short-Term Consultant
STT	Short-Term Temporary
VPU	Vice-Presidency Unit
WB	World Bank
WBI	World Bank Institute

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
ACRONYMS.....	iii
TABLE OF CONTENTS.....	v
EXECUTIVE SUMMARY.....	vii
1. INTRODUCTION	1
2. METHODOLOGY.....	2
Selection Procedure and Evaluation Administration	2
Evaluation Instruments.....	3
Activity Drop Rates.....	3
Participant Response Rates	4
3. ACTIVITY AND PARTICIPANT SAMPLE.....	5
Activity Sample.....	5
Participant Sample.....	8
4. COMPARING FY03 AND FY04 BENCHMARKS.....	12
Comparing FY03-FY04 Benchmarks	13
Comparing Activity Sponsors Using Respondent Level Data	15
5. MODELING TRAINING QUALITY.....	21
Variables.....	22
Multilevel Analysis	24
Multilevel Modeling Results.....	27
6. CONCLUSIONS.....	32
ANNEXES	
Annex A: Activity Questionnaire.....	34
Annex B: Participant Questionnaire.....	36
Annex C: Course Audience: Distribution of Respondent Unit,By Activity sponsor.....	37
Annex D: Distribution of Respondents by Unit in Activity Content Type.....	42

Annex E: Ratings For Jointly Sponsored Courses in FY04.....	43
Annex F: Mean Ratings of Activity Characteristics, FY03-FY04 (1=low; 5=high).....	44
Annex G: Variable Definitions.....	45
Annex H: HLM Between-Group Model Testing Cross-level effects between grade level and course TOPICS.....	46
Annex I: Key Determinants of Activity Quality, by Sponsor.....	47

EXECUTIVE SUMMARY

This report evaluates staff learning activities offered in FY04. IEG compiled level one evaluation questionnaires completed by 4,750 respondents in 317 staff learning activities. We examine the results of these surveys to monitor staff training and evaluate which factors can improve it.

On the whole, training quality ratings increased markedly in FY04 from FY03. There were significant increases in participant ratings in three of the five key dimensions of training quality: (a) relevance to the Bank's mission, (b) overall quality, and (c) usefulness of training; ratings of (d) course applicability to job, and (e) perceived increases in knowledge and skills (K&S) remained statistically the same in the two fiscal years. However, it is important to note that applicability to job dropped below the ASTD benchmark.

Overall, the number of learning sponsors above the Bank benchmarks significantly improved this year, nearly doubling from 22 percent of the sponsors in FY03 to 43 percent of sponsors in FY04. In other words, in FY04, nearly half of sponsors exceeded the benchmark in four of the five key dimensions of activity quality.

A comparison between FY03 and FY04 was drawn by the type of learning provider: *Professional and Technical (P&T) Network*, *Other Networks*, *Regions*, and *Other Learning Providers*. The *P&T Networks* and *Other Networks* substantially improved across the first three dimensions in FY04 (relevance to the Bank's mission, overall quality, and usefulness of training) and *Regions* improved significantly in relevance to the Bank's mission. However, ratings of job applicability and knowledge and skills decreased significantly among *P&T Networks* and *Regions* in FY04.

In order to better understand these trends, we also examined how sponsors fared according to Bank benchmarks at the respondent level and also when controlling for class size. Thus, calculations of the five key dimensions of activity quality were based on (a) all respondents in activities offered by an activity sponsor, and (b) all courses belonging to that sponsor. The first method is the raw score based on a denominator of all respondents. The second method controls for class size and is based on a denominator of the number of activities offered by the sponsor. A comparison of both methods led to our conclusions about the individual sponsors.

Activities sponsored by ACS, ISG, and P&I were rated consistently above benchmarks in FY04. On the other hand, activities sponsored by ESSD and INF were consistently rated below the benchmarks in FY04. These differences were demonstrated to be largely due to activity features, as well as participant traits.

A series of multi-level models were conducted to examine why activities sponsored by these networks were rated significantly below benchmark in terms of perceived activity quality (a composite variable comprised of the five key dimensions listed above).

The results from these analyses showed that higher training quality is not associated with sponsor when activity features are included in the equation. Specifically, participant ratings of three main pedagogical dimensions explained most of the variance in the model of activity quality. The pedagogical dimensions were:

- (a) *Training content*- better ratings especially in larger classes, those with 21 or more participants increased training quality;
- (b) *Presentation organization*- better ratings of presentation order increased participant ratings of training quality; and
- (c) *Training materials*- higher assessments of course materials led to significantly more favorable ratings of training quality.

Hierarchical Linear Model results also showed that perceived activity quality was associated with two activity features:

- (d) *Longer courses*, those with durations longer than 2 days, on average; and
- (e) *Smaller class sizes*, those with 21 or fewer participants, on average.

Additionally, HLM results indicated that activity quality ratings were also explained by participant traits:

- Motivation to participate “to enhance current work performance” significantly increased perceived quality ratings;
- Grade level indicated that ACS staff rated activities higher, on average; and
- Participants who reported that the activity was included in their learning plans were more likely to rate activities more positively.

1. INTRODUCTION

1.1 The Learning Board reports that more than three thousand staff learning activities took place in FY04 (N=3,300), with a total cost of over \$81.7 million, almost 18 percent more than the \$69.5 million originally allotted.¹ These activities include a variety of learning opportunities such as formal face-to-face courses as well as e-learning events and informal brown-bag lunches. These also include all joint staff and client learning activities.

1.2 Given the enormous resources allocated to staff learning, the Learning Board has requested the World Bank Institute Evaluation Group (IEG) to monitor the progress of these learning activities. As part of this mandate, IEG conducts level one evaluations of staff learning on an on-going basis.² In FY04, IEG evaluated 317 courses, an increase of nearly one third from FY03 (N=242 courses). Our sample of evaluated courses in FY04 represents approximately ten percent of formal staff learning events for the year, and approximately 30 percent of unique activities delivered. In the courses evaluated in FY04, we gathered data from 4,750 respondents.

1.3 In this report, we examine participants' assessments of staff training. Specifically, we examine respondents' evaluations of training quality based on the following key dimensions:

- relevance to the Bank's mission;
- overall quality;
- usefulness of training;
- applicability to job; and
- perceived increases in knowledge and skills.

We estimate the determinants of these dimensions using multi-level hierarchical modeling. In the following section, we discuss our methodology for conducting level one evaluations, sampling, and response rates.

¹ "Building Staff Capacity for Development: Update on the Staff Learning Program- FY04." Learning Board Executive Summary, FY04.

² IEG also conducts staff learning evaluations at levels two, three and four, the results of which are published elsewhere in a series of separate reports.

2. METHODOLOGY

2.1 In FY04, we drew a random sample of courses for level one evaluation stratified by sponsor. We obtained general course information from training coordinators and the Learning Catalog. Also, we gathered data for our evaluations through standard questionnaires administered in a uniform method to course participants. We discuss the methods for data collection in detail below.

SELECTION PROCEDURE AND EVALUATION ADMINISTRATION

2.2 The selection procedure used in FY04 was to draw a random sample of four staff training events by sponsor from the learning catalog on a bi-weekly basis.³ If more than four courses were eligible for evaluation within each sponsor, the courses were selected randomly by a computerized random number generator. Course eligibility was based on the following criteria:

- Activities were registered as formal events (i.e. no brown bags or conferences) in the learning catalog two weeks before the activity began;
- Bank staff comprised at least half of attendees;
- Course delivery mode was face to face or distance learning; and
- The course had not been previously evaluated by IEG in FY04⁴.

2.3 For each selected course, we asked training coordinators to submit the course agenda and fill out a short questionnaire about the event logistics (e.g., start and end dates, number of participants expected; see Annex A). Upon receiving the information necessary for coordinating the evaluation, we scheduled the evaluation and prepared the course questionnaires with the title and dates.⁵ The survey administrator delivered the evaluations to participants on the first day of the learning event. While the standard procedure was to introduce the survey to participants in person, explain how to fill out the form, and to emphasize the importance of participant feedback, there were instances where instructors refused to comply, and the survey administrators were forced to leave the questionnaires for participants to complete without the benefit of instruction. Completed forms were collected at the end of each course by an IEG staff member (see Annex B).

³ We selected 40 courses during each selection period in order to meet our target of evaluating 20 activities (we explain later why all selected courses are not evaluated).

⁴ When sponsors offered fewer than ten courses in the fiscal year, it was necessary to evaluate additional offerings of the same course to meet the minimum requirement.

⁵ Activities for which the course administrators did not respond with the necessary information after several requests and reminders were dropped.

EVALUATION INSTRUMENTS

2.4 The data for the analyses presented in this paper are compiled from the level one evaluations and information provided by course administrators about the activity. The activity questionnaire asks training coordinators to detail the logistics of the event as mentioned above as well as give some background information on the activity such as sponsor, training content (i.e. Bank Operations, Behavioral and Social, Executive Management, Information and Technology, Professional and Technical), the learning approach (i.e. interactive versus formal), and the history, design, and delivery of the training.

2.5 The evaluation questionnaires focus on participant satisfaction with the course overall as well as with course specifics: course delivery, design, and substance. In terms of individual benefit, respondents reported the extent to which the course fulfilled their learning needs, increased their knowledge and skills, and how much it related directly to their jobs. Perceptions of course quality, usefulness, and relevance to the Bank's mission were also asked to determine participant satisfaction. Assessments of the activity include participants' ratings of subject matter, order in which the content is presented, the materials and the balance of time spent on theoretical content, practical content, instructors' presentation, audience participation and overall pace of the training.

ACTIVITY DROP RATES

2.6 We evaluated 317 activities out of a total of 480 selected for evaluation in FY04 and dropped 163 courses from the evaluation sample.⁶ The share of dropped courses in FY04 (34 percent), decreased sharply from previous years (46 percent in FY03 and 50 percent in FY02). This suggests that evaluation has become more prominent on the learning agenda.

2.7 In FY04, approximately one third of these dropped courses (N=46) were cancelled, postponed, or had inaccurate course dates published in the learning catalog. Another third (N=44) did not provide the essential information necessary for evaluation. We dropped the remaining courses (N=73) for the following reasons: the sponsor or facilitator did not agree to evaluation being administered; another previously scheduled evaluation was to be delivered that date; course was cancelled due to low enrollment; course was evaluated once already in FY; trainers opt out because activity is a pilot offering; problems were incurred during overseas mail delivery of questionnaires; the local sponsor did not administer the evaluation survey; the activity was actually for clients rather than Bank staff (advertised as both in the learning catalog); and the activity was not a formal training session but was advertised as such in the learning catalog (e.g. conference) (See Table 1).

⁶ Drop rates were highest among activities sponsored by: WBI (82 percent), ECA (76 percent), HRS (70 percent), PSI (68 percent), ESD (55 percent), and RM (52 percent).

Table 1. Evaluation Sample Drop Rates, by Reason

Reasons Courses are Dropped from Sample	No. of Courses Dropped	Percentage
Course postponed / Cancelled / Dates of course changed/ Dates of course not officially set / Dates in the Learning Catalogue incorrect	46	28%
No information received / Uncooperative contact	44	27%
Other previously scheduled evaluation is being delivered to the participants	1	1%
Course dropped due to low enrollment	13	8%
Sponsor / Facilitator did not agree to evaluation being administered	16	10%
Repeat course that has been previously evaluated during FY	2	1%
Evaluation delivery failed (e.g. overseas mail never received)	2	1%
Evaluation delivery failed: Sponsor/Trainer promises to administer survey and fails to do so	2	1%
The course is advertised as staff and client but less than half of audience is Bank staff (not evident until contacting training coordinator)	0	0%
Not a training session (e.g. Conference or brown bag lunch) - not evident until contacting training coordinator	12	7%
Other / Miscellaneous	18	11%
Pilot offering	7	4%

PARTICIPANT RESPONSE RATES

2.8 Six thousand four hundred and seven participants attended the courses evaluated in FY04. Of this total, 4,750 completed the Level-1 questionnaire, for a response rate of 79 percent. This shows great improvement in cooperation by participants in comparison to the response rate in FY03 of 67 percent.⁷ Perhaps learning sponsors are taking heed of the Chief Learning Officer's emphasis on the importance of level one evaluations and are, therefore, encouraging trainers to instruct participants to complete the survey instruments. We discuss our sample further in the following section.

⁷ This is similar to the adjusted response rate of 78 percent in FY03.

3. ACTIVITY AND PARTICIPANT SAMPLE

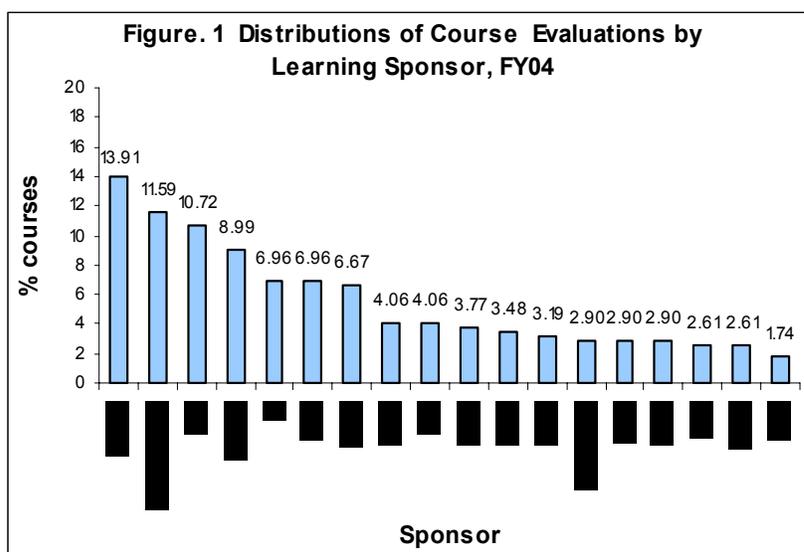
3.1 We used data at two levels of analysis for this evaluation – participant level and course level. This chapter describes the participant and course sample used for this evaluation.

ACTIVITY SAMPLE

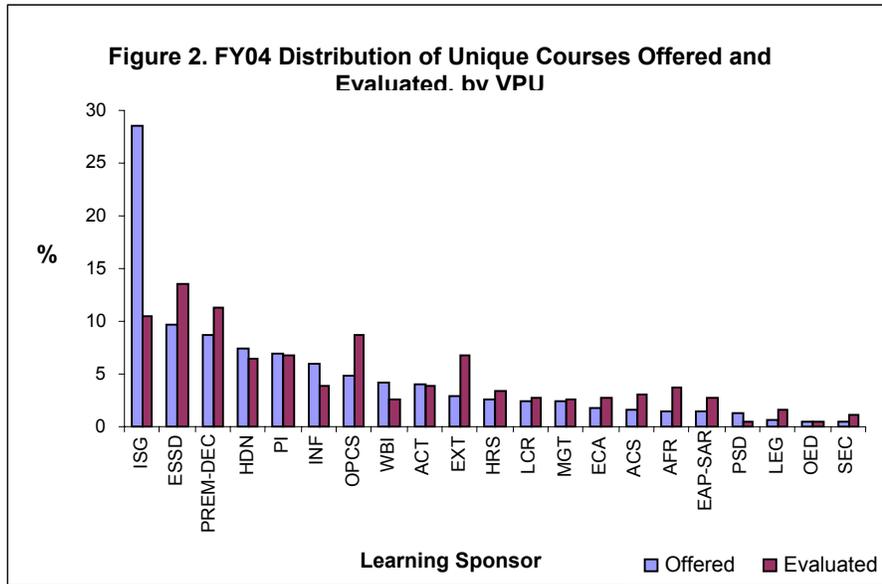
3.2 A total of 317 learning activities were evaluated in FY04. The World Bank designed (88 percent) and delivered (83 percent) the majority of these activities. External consultants designed and delivered the remaining. Activity duration ranged from one to eleven days with the average activity lasting approximately two days (mean=1.8, std dev=1.5). The following sections describe the activities’ sponsorship, general curriculum content and participant affiliation.

Courses Evaluated

3.3 Over half the evaluated staff learning activities were provided by six institutional sponsors (Figure 1). The largest proportion of courses evaluated came from ESSD (14 percent); PREM-DEC (12 percent); ISG (11 percent); OPCS (9 percent); and PI and EXT (7 percent each). Our sample also includes courses from HDN (6.6 percent), ACT and INF (4 percent each), AFR (3.7 percent), HRS (3.5 percent), ACS (3 percent), EAP-SAR (3 percent), LCR (3 percent), ECA (2.9 percent), WBI (2.6 percent), MGT (2.6 percent), and LEG (1.7 percent).



3.4 Figure 2 compares the distributions of activities evaluated with total courses offered in FY04.⁸ The evaluation sample grossly under-represents ISG courses because of the abundance of ISG courses offered throughout the year. In order to adequately represent all sponsors, we evaluate proportionately fewer ISG courses than their share in the total. However, in order to account for differences in the distributions, we weight our sample to match the distribution of courses offered in FY04. Thus, we weight respondents from ISG more heavily to match the number of participants in the population at large. Conversely, we assign smaller weights to ESSD respondents because they are over-represented in our sample.

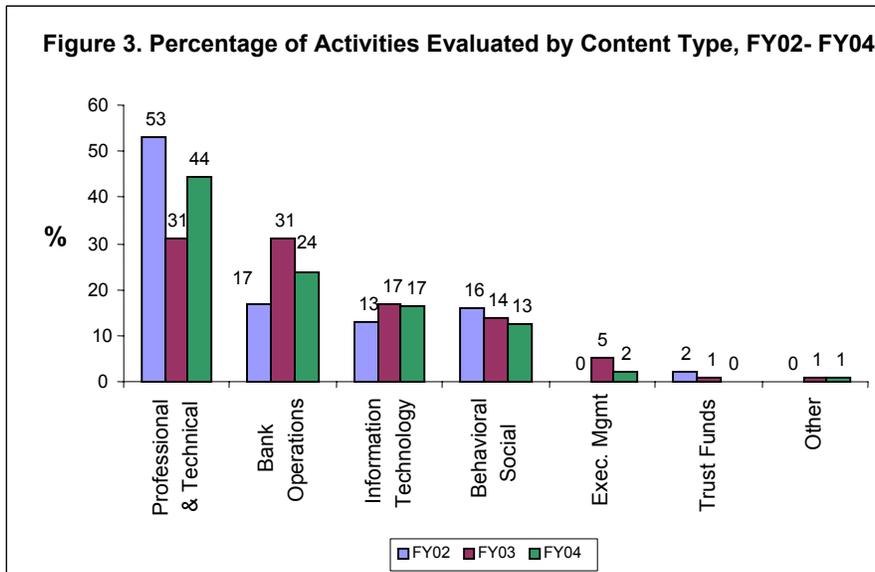


Activity Content

3.5 During FY04, the largest percentage of activities evaluated were Professional and Technical (44 percent); followed by Bank Operations (24 percent); Information Technology (17 percent); Behavioral and Social (13 percent); Executive Management (2 percent) and Other (1 percent).

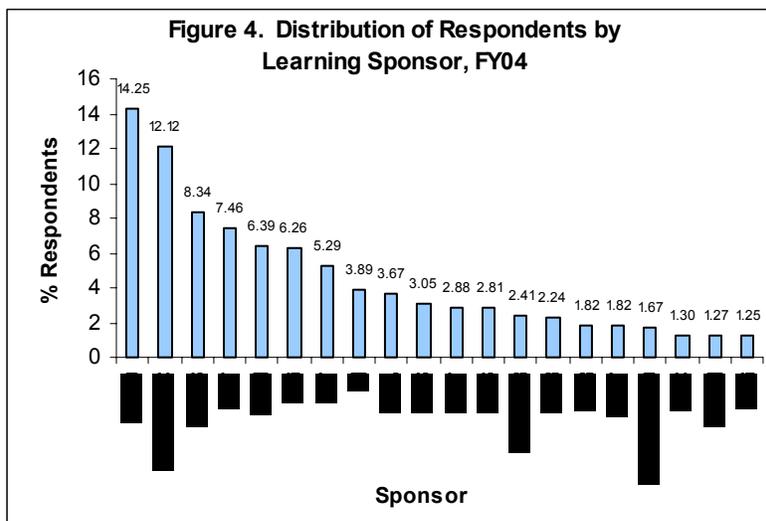
3.6 Compared to FY02 and FY03, the shares of Information Technology related activities and Behavioral and Social activities have remained constant within the range 13 to 17 percent. The share of Professional and Technical activities increased in FY04 (44 percent) compared to FY03 (31 percent), but was less than in FY02 (53 percent). By comparison, the share of Bank Operations related activities declined to 24 percent in FY04 compared to FY03 (31 percent) but was above the FY02 level (17 percent). The share of Executive Management activities also declined from 5 percent in FY03 to 2 percent in FY04. Figure 3 shows the percentage of activities evaluated by content type in FY02, FY03 and FY04.

⁸Multiple sponsors are factored into the distributions by treating each sponsor in jointly delivered activities equally. In other words, each sponsor in a jointly delivered activity is given equal credit.



Activity Respondents

3.7 Twenty-six percent of respondents attended activities sponsored by ESSD (14 percent) and PREM-DEC (12 percent). OPCS, EXT, HDN, ISG and INF accounted for 8 to 5 percent of the respondents. PI, ECA, ACS, ACT, HRS, EAPSAR, AFR, LCR, MGT, SEC, EXTPI and LEG accounted for fewer than 4 percent of the total respondents. Figure 4 shows the distribution of respondents by sponsors with one or more percentage of respondents in FY04 respondent sample.⁹



3.8 Not surprisingly, the two largest sponsors in terms of courses evaluated, ESSD (14 percent) and PREM & DEC (12 percent), also trained the largest number of

⁹ Staff learning in FY04 included both single sponsored activities and multi-sponsored activities.

respondents (14 percent and 12 percent respectively).¹⁰ In contrast, while ISG sponsored more courses (11 percent) than OPCS (9 percent), OPCS trained more respondents (8 percent) than ISG (6 percent). Likewise, while PI represented the same percentage of courses as EXT (7 percent each), EXT registered significantly more respondents (7 percent) than PI (4 percent). These findings show that PI and ISG register fewer participants in their courses even though they offered more or equal number of courses as some other sponsors. This is just what we would expect given that they are forced to control for course size (e.g., ISG training rooms can only hold a maximum of 16 participants.)

3.9 ESSD accounted for the most number of respondents in FY04 (14 percent). This was a significant gain compared to FY03 when it was in the fifth place with only 8 percent of the total respondents. In both FY02 and FY03, PREM accounted for the most respondents (15 percent in FY02 and 22 percent in FY03). However, in FY04, PREM-DEC was in the second place with 12 percent of the total respondents. Participation in OPCS sponsored courses also declined in FY04 (8 percent) compared to FY02 (14 percent) and FY03 (11 percent). On the other hand, EXT and ECA increased participation significantly in FY04 accounting for 7 percent and 4 percent of the respondents compared to FY03 when they both had less than 1 percent of the total share of respondents.

3.10 As emphasized in the FY03 evaluation report, an important aspect of the Learning Board's activities is to encourage participants to enroll in courses across units and not necessarily be restricted to taking courses offered by their units. In Annex C, we present the distribution of participants by units and course sponsors. Similar to our findings in FY03, we find that participants from regions are more likely to take courses from their own unit. Among the networks, P&T networks such as HDN, ESSD and PREM courses are more likely to have participants from their units. This suggests that region-sponsored courses and P&T network courses are designed specifically for participants working in their areas.

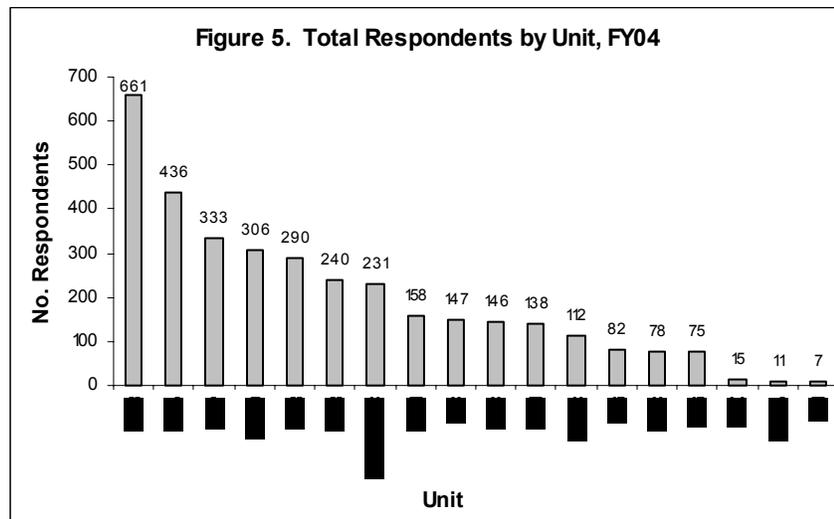
PARTICIPANT SAMPLE

3.11 Our participant sample consists of 4,750 respondents. Most of the respondents were World Bank Staff (87 percent), of whom 71 percent were based in headquarters and 29 percent were stationed in the field.

3.12 In both FY03 and FY04, the majority of the respondents come from the regional VPUs. Figure 5 illustrates the total number of respondents by the VPU in which they work. In FY04, the Africa region accounted for the largest total number of respondents (n=661); followed by ECA (n=436), EAP (n=333), LCR (n=290), and SAR (n=240). The only non-regional unit with a large number of participants was ESSD with the fourth largest number of respondents (n=306). Other than ESSD, the other units with large number of respondents were PREM (n=231), HDN (n=158), IFC (n=158), HRS (n=146),

¹⁰ As participant data is not available for the activities that are not evaluated, we rely on the evaluation sample of respondents to make inferences about the population at large.

WBI (n=138) and OPCS (n=112). Units with less than 100 respondents include: ISG, DEC, LEG, FSE, MIGA and PSI.¹¹



Participation Motivation

3.13 More than half of staff (59 percent) cited “enhance performance in current/planned assignment” as their primary motivation for taking the training course. The second most important motivation for taking a course was cited as “professional interest and growth” (32 percent). Only five percent of the respondents cited “networking” as being their main motivation. These findings are similar to those reported in FY03. These findings suggest that for the majority of the participants the main motivation for selecting a course is guided by their current work as opposed to future career opportunities that may be enhanced by internal networking within the Bank.

3.14 While the majority of respondents indicated that their primary motivation was to enhance their current work performance, only 33 percent reported including these activities in their individual learning plans.

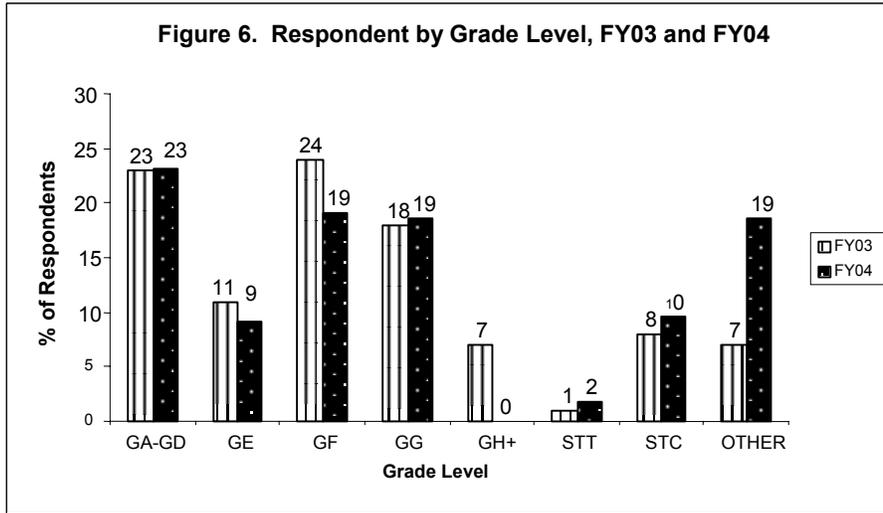
Grade Levels

3.15 Figure 6 represents respondents’ grade levels in FY04. Staff in grades GA-GD comprised nearly a quarter (23 percent) of the respondents; 9 percent came from GE level; 19 percent each from GF and GG and none from GH and above level. In addition, our sample included short term consultants (10 percent), short term temporary (2 percent) and respondents who identified themselves as “other” (19 percent).¹²

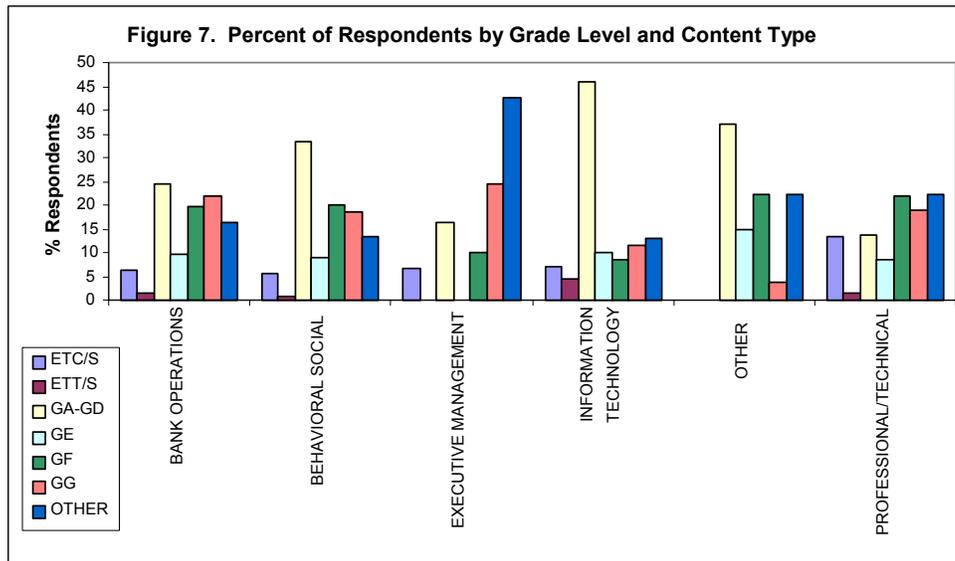
¹¹ The “Other” category refers to respondents who selected “other” on the survey for their work unit. There were 207 non-staff respondents who stated they were “clients” on the questionnaire and participated in WBI courses that were jointly offered to staff and client.

¹² Respondents in the “other” category include JPAs etc.

3.16 The distribution of the respondents by their grade level in FY04 is very similar to FY03. In FY03, the largest percentage of participants came from grade GA-GD (23 percent), GE (11 percent), GF (24 percent) and GG (18 percent). However, unlike in FY03 when 7 percent of the participants were from GH grade or above, in FY04 there were no respondents from this grade. The number of STC taking courses in FY04 increased slightly, as compared to FY03.



3.17 The types of courses attended by respondents differed according to their grade level (Figure 7). Information Technology courses attracted the highest proportion of ACS staff (46 percent) of all types of courses. Behavioral and Social courses also had the largest proportion of participants with grades A through D (33 percent). Bank Operations and Professional Technical courses attracted a generally even mix of grade levels with slightly more representation from higher level staff (51 and 50 percent, respectively). Executive management courses included mostly participants who were GG and higher or those who checked off the “other” category.



Participant Unit and Course Type

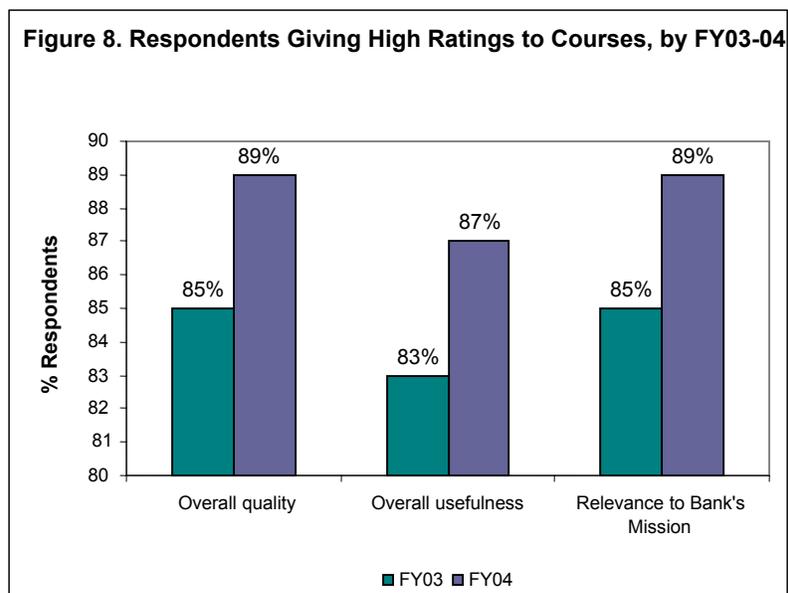
3.18 We also analyzed the types of courses attended by participants from various VPUs in order to examine the type of audience for each course type. Annex D presents the percentages of respondents in various types of courses by units. Bank Operations attracted maximum number of participants from regions: Africa (16 percent), ECA (15 percent), EAP (10 percent), LCR (8 percent) and SAR (6 percent). Behavioral/Social courses were most popular with participants from Human Resources (11 percent) followed by the Africa region (10 percent). A large proportion of participants from ISG (13 percent) attended Executive Management courses. Information Technology courses were attended most often by staff from the Africa region (13 percent) followed by IFC (10 percent). Profession/Technical courses were also very popular with participants from the Africa region (16 percent). It is not surprisingly that a majority of participants in various types of courses are from the Africa region since participants from the Africa region account for the largest share of participants overall.

4. COMPARING FY03 AND FY04 BENCHMARKS

4.1 We established new Bank benchmarks for FY04 using respondent level data (n=4,555) and course level data (n=317).¹³ These benchmarks are based on the average scores given by participants on: (a) assessment of the course's applicability to the participant's job; (b) perception of the degree to which knowledge and skills increased and the percentage of respondents giving high ratings (4-5); (c) course relevance to Bank's mission; (d) overall course quality; and (e) usefulness of the course.¹⁴ We draw comparisons both longitudinally between FY03 and FY04, and in relation to the industry standards set forth by the American Society for Training and Development (ASTD).

Comparing overall quality, usefulness and relevance in FY03 and FY04

4.2 The courses improved in FY04. Figure 8 compares the percentage of respondents giving high ratings to courses (i.e., percentage of respondents giving a rating of 4-5) on overall quality, usefulness and relevance in FY03 and FY04. We find that the percentage of respondents giving high ratings on overall quality, usefulness and relevance of courses significantly increased in FY04 compared to FY03 (quality: $\chi^2 = 26.88^{**}$; usefulness: $\chi^2 = 54.94^{**}$; and relevance: $\chi^2 = 70.48^{**}$).

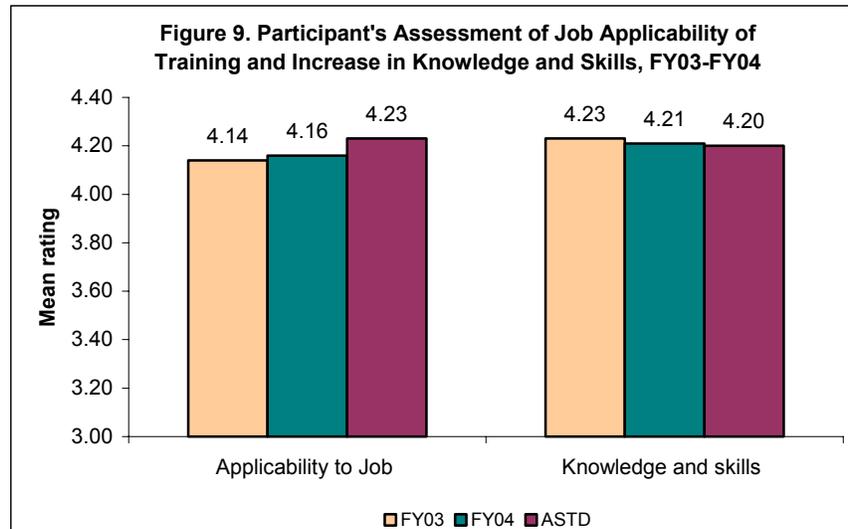


¹³In the FY04 questionnaire, participant ratings ranged from 1 (low) to 7 (high). In order to get comparative figures between FY04 and FY03, we rescaled the ratings from a 7 point scale to a 5 point scale as follows: 1→1; 2&3→2; 4→3; 5&6→4; and 7→5.

¹⁴ Annex F presents the weighted average ratings of activity characteristics: (a) material used during the course; (b) the order in which the content was presented; (c) training content quality; (d) the extent to which the training achieved its desired objective; and (e) the extent to which the training fulfilled respondent's learning needs.

Comparing applicability to job and increase in knowledge and skills

4.3 Respondents were also asked to assess the course's applicability to their job and the degree to which their knowledge and skills increased. Figure 9 presents the FY03-04 mean ratings of course applicability to job and perceptions of increases in knowledge and skills side by side to the ASTD benchmarks. There were no significant differences between the two fiscal years. The mean rating for job applicability was 4.16 in FY04 and 4.14 in FY03. Average perceived increase in knowledge and skills was 4.21 in FY04 compared to 4.23 in FY03.



4.4 While statistically significant differences between FY03 and FY04 do not exist, the Bank's FY04 average for job applicability is significantly lower than the most recently published ASTD benchmark¹⁵ 4.23, ($t=-6.69^{**}$). However, the Bank's FY04 average rating of increases in knowledge and skills was statistically the same as the ASTD benchmark, 4.20.

COMPARING FY03-FY04 BENCHMARKS

4.5 To gain insight into differences in perceived training quality across the Bank, we compared benchmarks in FY03 and FY04 across: *Regions*, *P&T Network SCLs*, *Other Networks* and *Other Learning Providers (OLP)*¹⁶. Table 2 shows the ratings on overall

¹⁵ Van Buren, Mark E. and William Erskine 2002. *Learning Outcomes: ASTD's Fourth Annual Report on Standards for Evaluating Organizations' Investments in Education and Training* Report 2002.

¹⁶ Each category includes the following sponsors: *Professional and Technical Networks* (PREM & DEC, INF, HDN, ESSD); *Other Networks* (PI, OPCS, MGT, ISG, ACS); *Regions* (MENA, LCR, ECA, EAP, SAR, AFR), and *Other Learning Providers* (WBI, SEC, LEG, HRS, GSD, EXT, ACT).

quality, usefulness, relevance, applicability to job and increase in knowledge and skills across categories for FY03 and FY04.¹⁷

Table 2: FY03-FY04 Ratings of training quality by learning provider

	FY03	FY04	t-value
P & T Networks			
Overall quality	0.82	0.84	2.42*
Overall usefulness	0.81	0.86	3.16**
Relevance to Bank's Mission	0.86	0.88	3.07**
Applicability to Job	4.11	4.01	-5.69**
Knowledge and skills	4.15	4.06	-6.18**
Regions			
Overall quality	0.92	0.93	0.87
Overall usefulness	0.89	0.87	0.66
Relevance to Bank's Mission	0.90	0.92	2.20**
Applicability to Job	4.32	4.09	-8.08**
Knowledge and skills	4.31	4.25	-2.25*
Other Networks			
Overall quality	0.90	0.94	5.33**
Overall usefulness	0.88	0.91	3.87**
Relevance to Bank's Mission	0.86	0.89	4.00**
Applicability to Job	4.24	4.29	2.68*
Knowledge and skills	4.36	4.36	0.29
Other Learning Providers			
Overall quality	0.91	0.88	-3.2**
Overall usefulness	0.87	0.87	0.5
Relevance to Bank's Mission	0.87	0.88	1.62
Applicability to Job	4.23	4.23	0.08
Knowledge and skills	4.25	4.2	-2.35*

** p<0.01; * p<0.05

4.6 Overall the networks show marked improvement in FY04. Both *P&T Networks* and *Other Networks* show significant improvement in overall quality, usefulness and relevance to Bank's mission in FY04. *Other Networks* showed significant increase in applicability to job.

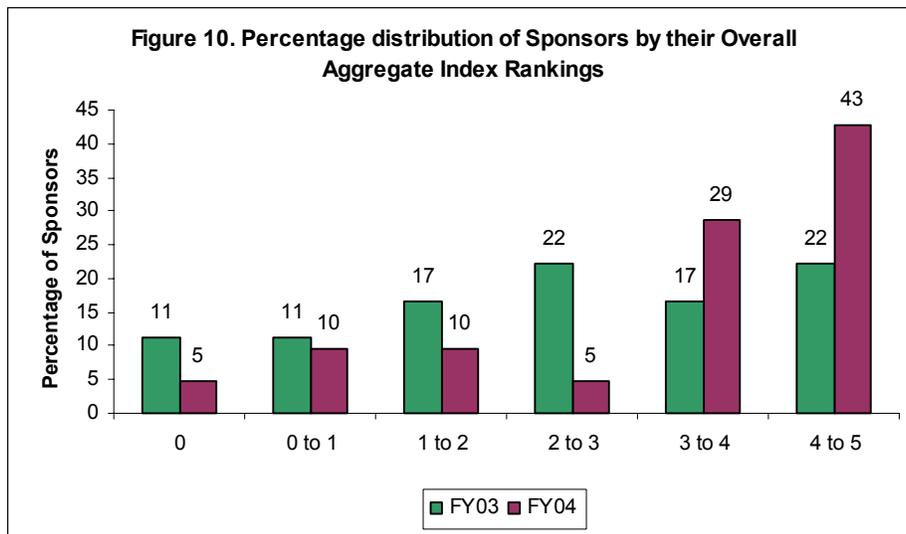
4.7 On the other hand, *P&T Networks* showed a substantial decrease in both applicability to job and knowledge and skills. Likewise job applicability and knowledge and skills declined among *Regions* in FY04. However, *Regions* improved significantly in relevance to the Bank's mission. Other learning providers decreased in overall quality and K&S.

¹⁷ We also report the t-statistic which indicates whether the difference between FY03 and FY04 ratings is statistically significant.

4.8 These results suggest that *P&T Networks* and *Regions* should pay particular attention to two dimensions of training: (a) making courses more applicable to people’s jobs, and (b) increasing participant knowledge and skills.

COMPARING ACTIVITY SPONSORS USING RESPONDENT LEVEL DATA

4.9 We examine how learning sponsors fared vis-a-vis the benchmarks in FY04 compared to FY03. We calculate an overall aggregate index across all sponsors in the five key training dimensions: knowledge and skills (K&S); job applicability; relevance; use and overall quality. We assign a score of one when a sponsor rated above Bank average, 0.5 when a sponsor rated same as Bank average, and a score of zero when a sponsor rated below Bank average. Thus, the maximum score a sponsor could achieve was five, indicating the sponsor was consistently above the benchmark, and the minimum score was zero, if they were always below the benchmark. We then calculated the percentage of sponsors that fell into the following ranges across the spectrum from high to low: 4-5, 3-4, 2-3, 1-2, 0-1 and 0 in the two fiscal years. Figure 10 illustrates the two distributions.



4.10 We find significant improvement in the proportion of learning sponsors that were above the Bank benchmark in FY04 compared to FY03. Nearly half of sponsors in FY04 (43 percent) rated above the Bank benchmarks in more than four criteria, compared with only 22 percent of sponsors in FY03. Likewise, in the next highest category (3 to 4), 29 percent of sponsors rated higher than the Bank benchmark in FY04, as compared to 17 percent of sponsors in FY03. On the other hand, the proportion of sponsors in the lower categories (2-3, 1-2, 0-1 and 0) was higher among FY03 than FY04.

4.11 Thus, overall there has been marked improvement in the proportion of learning sponsors meeting and surpassing the Bank benchmarks in FY04 as compared to FY03. To gain insight into which training criteria learning sponsors can improve, we depict the sponsors’ ratings in Table 3 with respect to training quality dimension.

4.12 Table 3 illustrates the benchmarks for single sponsors. We evaluated courses from a wider spectrum of learning providers in FY04 by including courses from: MENA, SEC, EAP, and PI. Columns one and two show whether respondents' ratings of K&S and applicability are significantly above, below, or the same as the bank average and the ASTD benchmark. Columns three, four, and five show whether or not 85 percent of respondents gave high ratings (4 or 5) for quality, usefulness, and relevance (with a dark circle indicating that 85 percent or more respondents gave scores of 4 and 5). Annex E presents the ratings by joint activity sponsors in FY04.

Table 3: Respondent level ratings of FY04, Q1-Q4, learning programs, by activity sponsor

Single Sponsors	Increase Knowledge and Skills ¹⁸	Job Applicability ¹⁹	Overall Quality ²⁰	Overall Usefulness*	Relevance to The Bank's Mission *	No. of Respondents
ACS	●	●	●	●	●	139
HRS	●	●	●	●	●	127
ISG	●	●	●	●	●	285
PI	●	●	●	●	●	177
EAP-SAR	●	◐	●	●	●	154
ECA	●	◐	●	●	●	167
EXT	◐	●	●	●	●	340
GSD	◐	●	●	●	●	8
LGR	●	◐	●	●	●	69
AFR	◐	◐	●	●	●	137
LEG	◐	◐	●	●	●	57
MENA	◐	◐	●	●	●	19
MGT	◐	◐	●	●	●	90
OPCS	◐	◐	●	●	●	390
ACT	●	◐	●	●	○	131
HDN	○	○	●	●	●	306
PREM-DEC	○	○	●	○	●	551
WBI	◐	◐	○	○	●	24
SEC	◐	◐	○	○	○	59
ESSD	○	○	○	○	●	626
INF	○	○	○	○	○	241

For columns 1 and 2, ●= higher than Bank/ASTD average, ◐= Bank/ASTD average, ○= below Bank/ASTD average; for columns 3, 4 and 5, ●= meets or exceeds Bank benchmark, ○= does not meet Bank benchmark

¹⁸ The benchmark is the Bank average rating (4.20), statistically the same as the ASTD standard (4.21).

¹⁹ The benchmark is the Bank average rating (4.18), statistically the same as the ASTD standard (4.21).

²⁰ The benchmark is 85% of respondents in a course gave ratings of 4 or 5.

Knowledge and Skills

4.13 The first column of Table 3 shows the ratings on knowledge and skills by learning sponsors. Comparing K&S in FY04 to the Bank/ASTD benchmark of 4.20, we find that 38 percent of sponsors (ACS, HRS, ISG, PI, EAP-SAR, ECA, LCR, and ACT) rated higher than the Bank average. Forty-three percent of sponsors (EXT, GSD, AFR, LEG, MENA, MGT, OPCS, WBI, and SEC) were rated statistically the same as the Bank average, while 19 percent of sponsors (HDN, PREM-DEC, ESSD, and INF) rated statistically below the Bank average.

4.14 ACS, ACT, EAP-SAR, LCR, and ISG continued to surpass the Bank benchmark in FY04. Sponsors who rated below the FY03 benchmark, EXT, GSD, ECA, and LEG improved their ratings in FY04 to either above or equal to Bank average. HRS improved its rating to above the FY04 benchmark from just meeting the benchmark in FY03.

4.15 Courses sponsored by AFR, OPCS, and WBI remained equal to the Bank benchmark in FY04. Although, ESSD and HDN met the Bank benchmark in FY03, they fell below the FY04 benchmark. PREM and INF remain below benchmark in FY04.²¹

Job Applicability

4.16 The second column of Table 3 shows the assessments of training applicability to job across course sponsors. We find that 28 percent of sponsors (ACS, HRS, ISG, PI, EXT and GSD) were statistically higher than the Bank benchmark of 4.18. Fifty-two percent of learning sponsors (EAP-SAR, ECA, LCR, AFR, LEG, MENA, MGT, OPCS, ACT, WBI and SEC) met the Bank benchmark while 19 percent of the single sponsors (HDN, PREM-DEC, ESSD, and INF) were below.

4.17 ACS was the only sponsor consistently above the Bank benchmarks in FY03 and FY04. ISG, GSD, HRS, and EXT improved their ratings in FY04 to above the Bank average. LEG also improved by meeting the Bank average in FY04 after rating below average in FY03. AFR, ECA, OPCS, and WBI remained equal to the FY04 benchmark.

4.18 On the other hand, some sponsors rated lower on job applicability in FY04 compared to FY03. EAP-SAR, ACT and LCR dropped from being above the benchmark in FY03 to just meeting the FY04 benchmark. ESSD, INF, and HDN were equal to the FY03 benchmark but decreased below Bank benchmark in FY04. PREM continued to remain below the benchmark in FY04.

Quality of Training

4.19 The third column in Table 3 illustrates whether or not 85 percent of respondents in a course gave high ratings (4 or 5) for quality (with a dark circle indicating that 85 percent or more respondents gave scores of 4 and 5). We find that 81 percent of sponsors (ACS, HRS, ISG, PI, EXT, GSD, EAP-SAR, ECA, LCR, AFR, LEG, MENA, MGT, OPCS, ACT, PREM-DEC, and HDN) rated above the Bank benchmark in FY04.

²¹ The following learning sponsors were evaluated for the first time in FY04: MENA, MGT, and SEC.

Nineteen percent of sponsors (WBI, SEC, ESSD, and INF) rated below the Bank benchmark.

4.20 In FY04, EXT, GSD, LEG, and PREM-DEC improved their ratings to meet the benchmarks for overall quality of training. On the other hand, ESSD and INF remained below the benchmark. WBI dropped from meeting the benchmark in FY03 to falling below in FY04.

Usefulness of Training

4.21 The fourth column of Table 3 presents the ratings of overall usefulness across course sponsors. Three-quarters of sponsors (76 percent) were equal to or higher than the Bank benchmark (ACS, HRS, ISG, PI, EXT, GSD, EAP-SAR, ECA, LCR, AFR, LEG, MENA, MGT, OPCS, ACT, and HDN). Only 24 percent of sponsors (PREM-DEC, WBI, SEC, ESSD, and INF) rated significantly below the Bank benchmark in FY04.

4.22 GSD and LEG improved in FY04 by meeting the benchmark. However, PREM-DEC and INF remained below the FY04 benchmark. WBI and ESSD decreased in FY04 after meeting the benchmark in FY03.

Relevance to Bank's Mission

4.23 The fifth column of Table 3 shows the respondent's ratings across sponsors on the basis of the relevance of their training to the Bank's mission. We find that 86 percent (ACS, HRS, ISG, PI, EXT, GSD, EAP-SAR, ECA, LCR, AFR, LEG, MENA, MGT, OPCS, PEMDEC, WBI, ESSD, and HDN) of the sponsors rated above the Bank benchmark and only 14 percent (ACT, SEC and INF) rated below the benchmark.

4.24 Relevance to the Bank's mission improved in courses sponsored by ISG, LEG, and GSD. On the other hand, INF remained below the Bank benchmark in FY04 and ACT dropped from meeting the benchmark in FY03 to below in FY04.

Overall Quality

4.25 The average ratings of activities of two sponsors fell below Bank or ASTD benchmarks in at least four of the above five categories: ESSD and INF. We explore the reasons for this pattern of lower ratings in subsequent analyses.

Comparing Activity Sponsors Using Course-Level Data

4.26 At the request of the Learning Board in FY04, we also calculated the benchmarks for learning programs across activity sponsors at the course level. In order to compute the benchmarks using course level data, we first calculate the average respondent ratings in each of the five categories at the course level. We then aggregate the course averages by sponsors to obtain the benchmarks across sponsors. This formula controls for the class size thereby reducing the influence of larger classes on the overall rating. Further, all sponsors are represented equally for jointly sponsored courses.

4.27 Table 4 illustrates sponsor ratings based on course level data. Columns one and two show whether average course ratings of K&S and applicability are significantly above, below, or the same as the bank average and the ASTD benchmark. Columns three, four, and five show whether 80 percent of courses meet the benchmark of 85 percent of respondents giving high ratings (4 or 5) for quality, usefulness, and relevance (with a dark circle indicating that 80 percent or more of courses met the benchmark).

Table 4: Course Level Ratings of FY04, Q1-Q4, Learning Programs by Activity Sponsor (where course N>=5)

Single Sponsors	Increase Knowledge and Skills	Job Applicability	Overall Quality	Overall Usefulness	Relevance to Bank's mission	No. of Courses
PI	●	●	●	●	●	24
ACS	●	●	●	●	●	11
LEG	●	●	●	●	●	6
ISG	●	●	●	●	○	37
EAP-SAR	●	●	●	●	●	10
LCR	●	●	●	○	●	8
ECA	●	●	●	●	○	10
HRS	●	●	○	●	●	12
OPCS	●	●	○	○	●	31
ACT	●	●	○	○	○	14
PREM-DEC	○	○	○	○	●	40
HDN	○	○	○	○	●	23
EXT	●	●	○	○	○	24
WBI	●	●	○	○	○	9
MGT	●	●	○	○	○	9
AFR	○	●	○	○	○	13
ESSD	○	○	○	○	○	48
INF	○	○	○	○	○	14

Note: Bank benchmark= 80% of activities are courses with 85% of participant ratings 4 and 5

For columns 1 and 2, ● = higher than Bank/ASTD average, ● = Bank/ASTD average, ○=below Bank/ASTD

average; for columns 3, 4 and 5, ● = meets or exceeds Bank benchmark, ○ = does not meet Bank benchmark.

Knowledge and Skills

4.28 The first column compares sponsors' mean ratings of knowledge and skills to the Bank benchmark.²² Only two sponsors (ISG and ACT) rated above the benchmark while two-thirds (61 percent) of the sponsors (PI, ACS, LEG, EAP-SAR, LCR, ECA, HRS, OPCS, EXT, WBI and MGT) met the Bank benchmark in increasing knowledge and skills. The remaining third (or 28 percent, including: PREM-DEC, HDN, AFR, ESSD,

²² A total of 18 sponsors that offered 5 or more courses were included in this comparison.

and INF) were rated below the Bank benchmark in terms of increasing knowledge and skills.

Job Applicability

4.29 The second column in Table 4 presents the assessment of job applicability across course sponsors. Two sponsors (PI and ISG) of the total of 18 sponsors with 5 or more courses were rated above average in terms of the job applicability of their training courses. Nearly two-thirds of the sponsors were rated statistically same as the Bank benchmarks (ACT, ACS, AFR, LEG, EAP-SAR, LCR, ECA, HRS, OPCS, EXT, WBI and MGT). Twenty-two percent of the sponsors offered courses that were rated below the Bank benchmark in terms of their job applicability (PREM-DEC, HDN, ESSD, and INF).

Quality of Training

4.30 Majority of the sponsors (61 percent) offered courses that were rated below the Bank benchmarks in terms of their overall quality of training. These sponsors are: ACT, AFR, HRS, OPCS, EXT, WBI, PREM-DEC, HDN, MGT, ESSD, and INF. Only 39 percent of the sponsors were higher than the Bank benchmark (PI, ACS, LEG, ISG, EAP-SAR, LCR, ECA).

Usefulness of Training

4.31 The fourth column of Table 4 shows the overall usefulness of courses across sponsors. Similar to the rating for overall quality, we find that 61 percent of course sponsors were rated below Bank benchmark for overall usefulness of their training (LCR, ACT, OPCS, AFR, EXT, WBI, PREM-DEC, HDN, MGT, ESSD, and INF). Thirty nine percent of the sponsors were higher than the Bank benchmark (PI, ACS, LEG, ISG, EAP-SAR, ECA and HRS).

Relevance to Bank's Mission

4.32 In contrast to the course level ratings across sponsors for overall quality and usefulness, we find that half the sponsors were rated above the bank benchmark in terms of the relevance of their courses to the Bank's mission. The sponsors that were rated above the Bank benchmark are: PI, ACS, LEG, EAP-SAR, LCR, HRS, OPCS, PREM-DEC and HDN. Fifty percent of sponsors were rated below the Bank benchmark in terms of relevance (ISG, ECA, ACT, EXT, WBI, MGT, AFR, ESSD, and INF).

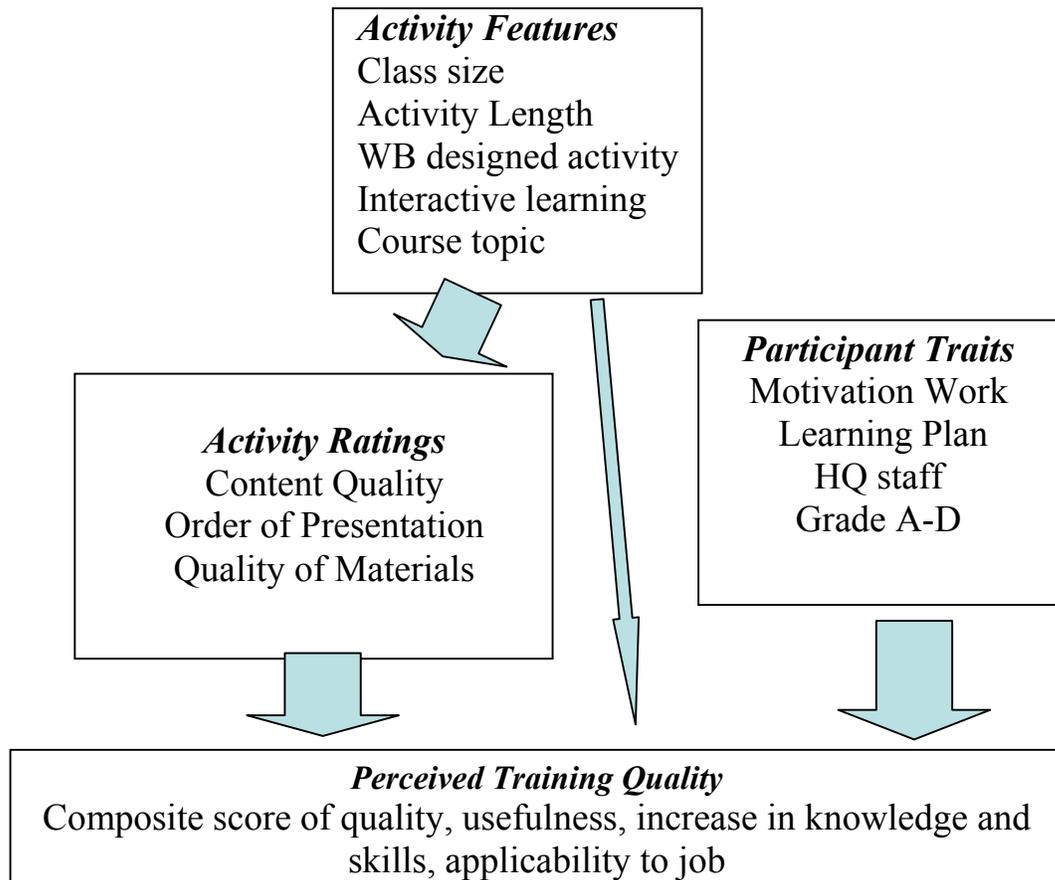
Overall Quality

4.33 Again, with respect to overall quality, two sponsors receive below standard ratings across all five categories: ESSD and INF. We use HLM analysis to determine what causes these sponsors to be rated significantly below the others, and to explain how quality can be improved based on activity features and participant characteristics.

5. MODELING TRAINING QUALITY

5.1 Some activities are more highly rated by participants than others, and both activity features and participant characteristics could account for these differences in ratings. In this chapter we estimate a series of models that (a) test whether low-rated activities are a function of participant characteristics or activity features, and (b) provide information on how to improve participants' experiences in the learning activity. More generally, we seek to understand the determinants of overall satisfaction with training quality as a function of both participant and course characteristics (see Figure 11). The approach in this chapter builds on findings from past research (Chard and Arango 2003); but utilizes a multi-level approach more suitable for these data.²³ We describe the variables and descriptive statistics below (also see Annex G).²⁴

Figure 11. Determinants of Perceived Training Quality



²³ This approach avoids the biased estimation of standard errors which results from using ordinary least squares approaches with unbalanced nested data. The data are weighted to reflect the actual distribution of the courses delivered.

²⁴ We imputed missing data with the mean values at the participant level.

VARIABLES

Individual Ratings of Training Quality

5.2 The dependent variable in our analysis, overall training quality, is a composite measure based on four participant-reported assessments of the activity for its quality, usefulness, effectiveness in increasing the participant's knowledge and skills, and applicability to the participant job (questions 13, 14, 16, and 17 in Annex B). All items had a seven point response option scale, which were recoded from zero (low quality) to one (high quality). We combine these variables both because they are theoretically measuring the same underlying dimension of overall perceived quality and because they are statistically correlated ($\alpha=.89$).

5.3 It is important to note that although values for overall quality range between zero and one, the data are strongly and positively skewed ($\text{mean}=.81$, $\text{std dev}=.15$). We tried various types of transformations (i.e. natural log, squaring, cubing), none of which mitigated the skewedness of the data; we therefore left it in its most readily interpretable state. The reliability score of .668 indicates solid reliability in the outcome variable (Raudenbush and Bryk 2002).²⁵

Participant Predictors

5.4 We include four individual participant characteristics as predictors of overall quality rating; all are “dummy” variables, where one indicates the participant has the characteristic and zero indicates absence of the characteristic. These are:

- **Motivation Work:** the participant enrolled in the activity to learn skills to enhance their work productivity. Fifty-nine percent of participants indicated work was their primary motivation for taking the course.
- **HQ staff:** the participant is based in headquarters. Seventy-one percent of staff are located in Headquarters.
- **Administrative grade level (Grade AD):** the participant is administrative staff in grade levels A through D; ACS staff represent one quarter of our sample.
- **Learning Plan:** the participant registered the course in their learning plan; thirty-three percent of respondents reported that they attended the course as part of their learning plan.

Activity Features

5.5 Indicators of activity features were developed from participant “Level-1” survey responses, a questionnaire completed by the activity task manager, and data in the Learning Management System (LMS). We include eight activity features in our analysis:

²⁵ Raudenbush, Stephen W. and Anthony S. Bryk. 2002. Heirarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks: Sage.

- **Sponsor:** a dummy variable identifying the baseline comparison group of activities that were rated below the benchmarks at both the respondent and course levels (i.e. INF and ESSD).²⁶ These activities represent twenty percent of our sample.
- **Course type:** three dummy variables identifying each of three major types of content included as controls (professional/technical, Bank operations and behavioral).
- **Training content:** the rating given by activity’s participants for the subject matter and training content, based on a seven-point scale recoded between zero (very poor) and one (very good) (question 8 in Annex B). The mean of .80 (sd=.18) indicates that content is rated highly.
- **Presentation order:** the rating given by the activity’s participants for the logical sequence of the presentations, based on a seven-point scale recoded between zero (very poor) and one (very good) (question 9 in Annex B). The mean of .80 (sd=.17) indicates that presentation organization was assessed as good to very good, on average.
- **Course materials:** the rating given by the activity’s participants for the materials used in the course, based on a seven-point scale recoded between zero (very poor) and one (very good) (question 10 in Annex B). The mean of .77 (sd=.19) shows that the materials were perceived to be of good quality.
- **Class size(Totpart):** a continuous variable measuring the total number of participants in the activity. Class size ranged from four to 105, with an average of 21 participants (sd=13.5).
- **Course duration (Ndays):** a continuous variable indicating the length of the activity in days. Duration ranged from a minimum of one day to a maximum of eleven days, with the average duration around two days (mean is 1.8 with a standard deviation of 1.5).
- **Designed by WB (WBdesign):** a dummy variable where a one indicates that the activity is designed by the World Bank. Activity task managers reported that eighty-eight percent of activities were designed internally.
- **Interactive learning (Interactive):** a dummy variable where one indicates that the activity used action-oriented learning approaches versus a solely traditional lecture format. Activity task managers reported that fifty-eight percent of activities used interactive learning.²⁷

²⁶ While INF and ESSD received the lowest scores in our sample, their evaluations are relative to the remainder of the sample, and overall, evaluations are skewed to the positive.

²⁷ Interactive Learning question: “Will the training approach primarily employ “Interactive “ Learning? (Interactive Learning is a process of learning and reflection that happens with the support of a group of peers working together with real business problems with the intention of accomplishing tasks while enhancing the learning of the group at the same time.)”

MULTILEVEL ANALYSIS

5.6 Hierarchical Linear Modeling (HLM) is used to estimate the effects of participant characteristics and activity features on overall quality ratings of activities. This multilevel modeling technique estimates fixed and random effects, and seeks to explain the apparent lower participant ratings given to activities of selected sponsors.

5.7 First, we partition the variation in the dependent measure of activity quality into components residing at the participant and activity level. This is accomplished by estimating an unconditional two-level HLM analysis where the random error components signal the relative influence of factors at both these levels on overall quality. We calculated the Intra-Class Correlation (ICC) for the dependent variable based on the fully unconditional model. The ICC of .16 [$ICC = \tau / (\tau + \sigma^2) = .00339 / (.00339 + .01819)$] indicates that 16 percent of the variance in training quality can be explained by activity level differences. Second, we use HLM to examine activity features that can explain this variance, in four models.

Model One

5.8 In our first model, we estimate a simple bivariate model to confirm that activities offered by the baseline sponsors are indeed statistically significantly below other evaluated activities. We predict overall training quality (participant ratings of: quality, usefulness of training, perceived increases in knowledge and skills, and applicability to job) as a function of sponsor. Equation 1 illustrates Hierarchical Model 1²⁸:

Level-1

$$Y(\text{Overall Training Quality}) = \beta_0 + \beta_1(\text{Sponsor}) + r$$

Level-2

$$\beta_0 = \gamma_{00} + \gamma_{01}(\text{Sponsor}) + u_0$$

Eq. 1

Model Two

5.9 In a second model, we estimate a between group model where we predict training quality not only from sponsor, but also participant characteristics. In other words, we test whether the low scoring courses are rated poorer because the participants in these courses are somehow different from participants in other activities. These characteristics are: participation motivation to improve current work, HQ based staff (versus field staff), ACS grade, and whether or not the activity was registered in participants' learning plans. All participant variables are grand mean centered. Equation 2 illustrates our second model:

²⁸ In this chapter, the terms "Level 1" and "Level 2" refers to participant and activity measures, respectively, and should not be confused with "Level-1 Evaluations".

Level-1

$$Y(\text{Overall Training Quality}) = \beta_0 + \beta_1(\text{Sponsor}) + \beta_2(\text{Motivation work}) + \beta_3(\text{HQ}) + \beta_4(\text{gradeAD}) + \beta_5(\text{Learning Plan}) + r$$

Level-2

$$\begin{aligned}\beta_0 &= \gamma_{00} + \gamma_{01}(\text{Sponsor}) + u_0 \\ \beta_1 &= \gamma_{10} \\ \beta_2 &= \gamma_{20} \\ \beta_3 &= \gamma_{30} \\ \beta_4 &= \gamma_{40}\end{aligned}$$

Eq. 2

Model Three

5.10 In our third model, we test whether adding activity characteristics to the model predicting course quality from participant characteristics helps to explain why the ratings of activities of selected sponsors are rated lower and if the addition of these activity features diminishes the effect of sponsor. In other words, we examine whether activity features (class size, course length in days, activity design by World Bank, interactive learning format, training content, materials and presentation organization) can explain the variance in quality ratings better than the simpler model comprised of sponsor and participant characteristics. Equation 3 illustrates our third model.

Level-1

$$Y(\text{Overall Training Quality}) = \beta_0 + \beta_1(\text{Motivation work}) + \beta_2(\text{HQ}) + \beta_3(\text{gradeAD}) + \beta_4(\text{Learning Plan}) + \beta_5(\text{Content}) + \beta_6(\text{Order}) + \beta_7(\text{Material}) + r$$

Level-2

$$\begin{aligned}\beta_0 &= \gamma_{00} + \gamma_{01}(\text{totpart}) + \gamma_{02}(\text{ndays}) + \gamma_{03}(\text{designWB}) + \gamma_{04}(\text{Interactive}) + \gamma_{05}(\text{Sponsor}) + \gamma_{06}(\text{P\&T}) + \gamma_{07}(\text{Bank Operations}) + \gamma_{08}(\text{Behavioral}) + u_0 \\ \beta_1 &= \gamma_{10} \\ \beta_2 &= \gamma_{20} \\ \beta_3 &= \gamma_{30} \\ \beta_4 &= \gamma_{40} \\ \beta_5 &= \gamma_{50} + \gamma_{51}(\text{totpart}) + \gamma_{52}(\text{ndays}) + u_5 \\ \beta_6 &= \gamma_{60} + \gamma_{61}(\text{totpart}) + \gamma_{62}(\text{ndays}) + u_6 \\ \beta_7 &= \gamma_{70} + \gamma_{71}(\text{totpart}) + \gamma_{72}(\text{ndays}) + u_7\end{aligned}$$

Eq. 3

5.11 These equations predict overall quality as a between group hierarchical linear model. Participant motivation (to enhance current work), staff status (HQ), grade level (A through D) and learning plan are fixed, while individual ratings of content, material, and order are free to vary at level one. Activity duration, class size, World Bank design, interactive learning, sponsor, and training content (professional/technical, Bank operations, behavioral) are included at level two. We also model the slopes for the key instruction variables (content, material, order) as a function of activity features (i.e. class

size and activity duration).²⁹ Our expectations for the impact of activity features are as follows:

- *Ratings of course features will significantly influence overall quality.* As ratings of content, materials and presentation order become more positive, individual overall satisfaction and ratings of activity quality will also increase.
- *Longer courses will increase ratings of training quality.* Past research shows that as course duration increases, quality ratings become more positive.
- *Larger class sizes will decrease overall quality ratings.* Previous findings indicate that larger courses were rated less positively by participants.
- *Courses designed by the World Bank will be rated higher in quality.* Courses designed by the Bank will be perceived more positively by staff because Bank designed courses would be more relevant to staff.
- *Courses utilizing interactive learning approaches will be perceived more favorably.* Activities that use interactive learning such as small group exercises and interactive instructional approaches encouraging participation will be rated significantly higher than activities that use only traditional lecture approaches.
- Content type and sponsor are included as controls.

Our expectations for participants' characteristics are:

- *Participants who are motivated to attend learning activities to improve their work performance will rate courses more positively.*
- *Likewise, participants who register activities in their individual learning plans will rate the course more positively.*
- *Staff working in administrative positions will rate training they attend significantly higher in quality than professional level staff.* This theory is based on past research which showed that administrative staff rate courses more positively than participants in higher level grades (Ouichi and Le Rouzic 2002).
- A variable for HQ versus field based staff is included as a control variable.

²⁹ We arrived at this model after trying various specifications (i.e. we tested whether gradeAD varied at the course level in a within model; we re-centered ndays as grand centered in order to help make interpretation of the intercept meaningful; and we estimated the model using only a between group model without modeling the slopes). We found no effect for course development (training having been previously offered 2 or more times), course topic, nor participant affiliation with region/network/other.

Model Four

5.12 In our fourth model, we present a general model of training quality. We remove the sponsor variable from the equation and test whether the relationships found in Model three are generalizable to all activities. Equation 4 illustrates Model four.

$$\begin{aligned}
 & \textbf{Level-1} \\
 Y(\text{Overall Training Quality}) &= \beta_0 + \beta_1(\textit{Motivation work}) + \beta_2(\textit{HQ}) + \beta_3(\textit{gradeAD}) + \\
 & \beta_4(\textit{Learning Plan}) + \beta_5(\textit{Content}) + \beta_6(\textit{Order}) + \beta_7(\textit{Material}) + r \\
 & \textbf{Level-2} \\
 \beta_0 &= \gamma_{00} + \gamma_{01}(\textit{totpart}) + \gamma_{02}(\textit{ndays}) + \gamma_{03}(\textit{designWB}) + \gamma_{04}(\textit{Interactive}) + \gamma_{05}(\textit{P\&T}) + \\
 & \gamma_{06}(\textit{Bank Operations}) + \gamma_{07}(\textit{Behavioral}) + u_0 \\
 \beta_1 &= \gamma_{10} \\
 \beta_2 &= \gamma_{20} \\
 \beta_3 &= \gamma_{30} \\
 \beta_4 &= \gamma_{40} \\
 \beta_5 &= \gamma_{50} + \gamma_{51}(\textit{totpart}) + \gamma_{52}(\textit{ndays}) + u_5 \\
 \beta_6 &= \gamma_{60} + \gamma_{61}(\textit{totpart}) + \gamma_{62}(\textit{ndays}) + u_6 \\
 \beta_7 &= \gamma_{70} + \gamma_{71}(\textit{totpart}) + \gamma_{72}(\textit{ndays}) + u_7
 \end{aligned}$$

Eq. 4

MULTILEVEL MODELING RESULTS

5.13 Table 5 presents the HLM results for the four models. The first six columns present the results for models one, two, and three, which successively build on each other.³⁰ The last two columns present our fourth and final model which explains perceived training quality, more generally.

5.14 Model one confirms the findings reported in the previous chapter that participants in activities belonging to the baseline group of low rated sponsors gave statistically significantly lower quality ratings ($b = -.05$, $p < .01$). Thus, we go on to examine why these activities are rated lower.

5.15 In Model two, we ask whether this is due to participant differences. Are staff who are interested in these topics and who participate in related activities inherently different from participants in other courses? Are they perhaps more critical of training? The results from Model two indicate that this is not the case. While participant characteristics are predictors of perceived training quality, they do not explain away the negative effect of sponsor. In other words, even when we control for participant motivation, HQ based status, grade level, and learning plan, a significant and negative impact from sponsor remains ($b = -.03$, $p < .01$). This could be due to unmeasured participant characteristics or

³⁰ Our analysis is based on questionnaires using the seven point scale and staff only (respondent N= 3980, course N=309).

could be a consequence of activity features. As we cannot investigate the former, we turn to the latter.

Table 5. HLM Between-Group Model Predicting Training Quality

Fixed Effects	Model 1		Model 2		Model 3		Model 4	
	Gamma Coefficient	Reliability estimate						
Intercept	0.8300**	.629	0.8246**	.599	0.8226**	.817	0.8210**	.818
Sponsor	-0.0461**		-0.0329**		-0.0140		--	
Participant Motivation (work)			0.0327**		0.0254**		0.0255**	
Based in HQ			0.0146*		0.0050		0.0049	
Grade A -D			0.0249**		0.0112**		0.0117**	
Learning Plan			0.0220**		0.0072*		0.0073*	
Total Participants					-0.0015**		-0.0015**	
Course Days					0.0138**		0.0144**	
Designed by WB					0.0088		0.0075	
Interactive Learning					-0.0012		-0.0022	
Professional Technical					-0.0129		-0.0133	
Bank Operations					-0.010		-0.0082	
Behavioral					0.0131		0.0157	
Training Content						.224		.224
Intercept					.3767**		0.3762**	
Total Participants					0.0045**		0.0046**	
Course Days					-.01401		-0.0137	
Presentation Order						.289		.289
Intercept					0.1944**		0.1946**	
Total Participants					-0.0017		-0.0017	
Course Days					0.0023		0.0021	
Course Materials						.170		.170
Intercept					0.1178**		0.1178**	
Total Participants					-0.0012		-0.0012	
Course Days					-0.0139		-0.0139	
Random Effect	Standard deviation	Variance						
Mean quality	0.0567	0.0032	0.0523	0.0027	0.0548**	0.0030	0.0550**	0.0030
Training Content					0.1370**	0.0188	0.1371**	0.0188
Presentation Order					0.1756**	0.0308	0.1755**	0.0308
Course materials					0.0989 †	0.0098	0.0990 †	0.0098
R	.12336	.0153	.12239	.0150	0.01733	0.0051	0.0713	0.0051

**p<.01, *p<.05, †p<.10;

BOLD italics indicates grand mean centered; BOLD indicates group mean centered

5.16 The results from Model three show that activity features are in fact significant predictors of training quality, and their inclusion in the model reduces to non-significance the coefficient for the effect of sponsor. Five activity features -- training content, presentation order, training materials, class size, and course duration -- explain participants' ratings of training quality. Model four, the same in all respects to Model three other than the non-inclusion of the sponsor category variable, shows the same relationships. This suggests that all activities, not only those in the baseline category, can improve their quality ratings by concentrating on enhancing the key aspects of training, prominent in both Models three and four.

5.17 Because the findings for both models are consistent, we discuss the results with respect to the coefficients reported in Model three.³¹ Overall, the reliability estimates in the third column show that we can have reasonable confidence in our estimates for the random variables (.82 for the intercept, .22 for content, .29 for order, and .17 for materials). Our model explains 72 percent of the variance in the intercept and within variance.³² The model results demonstrate the importance of five key activity features and three participant characteristics.

Activity features

5.18 Four of the eight activity features discussed above were significant predictors of participants' ratings of overall activity quality, and in some cases cross-level effects were also found.

Training Content

5.19 Good training content significantly increased quality ratings by 38 percent ($b=.38$, $p<.01$). In activities where participant rating of training content was positive, individual ratings of overall quality were also high. In examining the cross level effects however, we find that the effect of training content is mediated by class size ($b=.01$, $p<.01$). Content ratings are more important in larger classes.³³ As class size increases (average=21), the effect of subject matter becomes more prominent in overall quality ratings. For example, increasing class size from average, by one standard deviation, increases the importance of content ratings by 14 percent, thereby increasing the effect of training content to 52 percent. This suggests that trainers need to pay even more attention to content as their audiences increase. This is perhaps because larger courses do not lend themselves as easily to other important aspects of learning such as class discussion. Qualitative feedback suggests that respondents enjoy participating in class discussion in smaller classes.³⁴

Presentation Order

5.20 Likewise, presentation order being rated positively also increased perceived quality ratings by 19 percent. In other words, activities in which presentation organization is logical received significantly higher quality ratings. This indicates that paying closer

³¹ Differences in the results of the two models are limited to few instances in the third or fourth decimal place, which do not change the substantive interpretations.

³² The within model reduction in sigma squared is 72 percent. The random effect estimates for the intercept (illustrated in the lower part of the table) that the adjusted variance for quality ratings is .003. This is the amount of variance in quality ratings unaccounted for by the model, $p<.01$, indicating that there is still variance to be explained. The amount of within-group variance unaccounted for by the model is .01. In other words, after accounting for participant characteristics and assessments of course features, the estimate of within group variance is .01.

³³ We did not find any other cross-level effects between course features and course topic (i.e. P&T, Bank operations, Behavioral). We also tested whether the influence of training content, presentation order, and course materials are mediated by course characteristics such as course length, interactive learning format and activity topic.

³⁴ A summary of the qualitative analysis will be reported in a separate evaluation comparing the highest and lowest rated activities.

attention to the order in which content is presented is critical to successful learning activities, as deemed by participants. For example, this perhaps means strict adherence to the agenda and avoiding last minute schedule changes.

Course Materials

5.21 Activities where participants rated the course materials positively also received higher ratings from participants, by 11 percent.

Class Size

5.22 Smaller classes received higher ratings for training quality. Increasing the average total number of participants in an activity decreases quality ratings significantly ($b=-.002$, $p<.01$). For instance, increasing the average class size by one standard deviation from the average of 21 to 35, decreases quality ratings by three percent. Further increasing the class size to the maximum in our sample, 105, results in a 17 percent decrease in perceived quality ratings.

Course duration

5.23 The data support our hypothesis that longer courses are perceived more positively. On average, increasing course length by one day, increases quality ratings by one percent. Increasing course length from average (nearly two days) to the maximum in our sample (eleven days) increases quality ratings by nine percent.³⁵ Again, however, since increasing duration yields only a one percent increase per day in overall quality, it may be cost-effective to retain shorter courses.

Insignificant activity-level predictors

5.24 Contrary to our expectations, we found no effects in the HLM model for the following activity features: World Bank design, interactive learning, or type of course (professional/technical, Bank operations, behavioral). That is, courses designed by Bank staff, using interactive learning methods, or with specific training content were no more or less positively rated than average.

5.25 Perhaps the most counterintuitive finding is for interactive learning approaches in staff training; there is no evidence in our analyses to support the theory that interactive learning enhances adult learning. There are at least two reasons for this result. The first is that, in fact, interactive learning methods are no more effective than didactic ones. However, since a broad body of literature suggests that genuine interactive learning is effective with adults, this is not an appealing explanation. Another explanation is that there may be some social desirability bias in our interactive learning measure, which was collected from the course administrators. Since “action learning” is encouraged at the Bank, course administrators may over-report “interactive learning”. We note that the share of activities identified as using “interactive learning” is quite high in FY04: 54

³⁵ We tested activity length for a quadratic relationship but did not find one.

percent. We do not have confirmatory data on this measure from participants for staff learning, but such data are available for WBI activities. Quizon, Chard and Lockheed (2004)³⁶ found that sixty-one percent of task managers reported that action learning was utilized in their training events, but only thirty-three percent of participants indicated that the activity included action learning exercises such as action plans. When a measure of action learning that combined both staff and participant views was used, it was found to be strongly associated with more positive ratings of the activity.

Participant characteristics

5.26 Our model also controls for various participant characteristics such as motivation, grade level, staff status, and whether or not the activity was registered in participants' learning plans. Staff located in Headquarters did not rate activities more or less positively than did staff located in the field. However, participant's grade level, learning plan status, and motivation for taking the course were significantly related to assessments of training quality.

ACS grade level

5.27 Administrative staff rated training quality one percent higher on average than did professional level staff ($b=.01$, $p<.01$). We tested this for random effects and found no significant differences by course type or sponsor category (see Annex F). The null findings indicate that administrative staff are neither more, nor less likely to rate certain types of courses more positively.

Motivation for participation

5.28 Participants' motivation for taking a course was also related to training quality. We find that participants whose primary motivation was to enhance performance in their current work program rated the courses more favorably by three percent, on average, than participants with some other motivation for learning (i.e. professional growth/interest, networking, or "other", $b=.03$, $p<.01$).

Learning plan

5.29 Registration of the activity in participants' individual learning plans can be considered another indicator of motivation. We found that participants who registered their activities in their individual learning plans rated activities higher by one percent, on average ($b=.01$, $p<.05$).

³⁶ Quizon, J. B.; C. L. Chard; and M. Lockheed. "The Effectiveness and Use in FY03 of WBI FY01-02 Activities: A Baseline Assessment in Five Countries." WBI Evaluation Studies, No, EG04-86, May 2004.

6. CONCLUSIONS

6.1 Overall, staff learning quality has improved significantly in FY04. On the whole, the Bank has considerably increased the quality, usefulness, and relevance of its learning programs in FY04. Participant ratings significantly increased from FY03 in three of five key dimensions measuring the value of training: (a) overall quality (b) usefulness, and (c) relevance to the Bank's mission. A comparison of various types of learning providers showed specifically, that *P&T Networks* and *Other Networks* markedly improved across all three dimensions in FY04. Further, the *Regions* improved significantly in making training more relevant to the Bank's mission.

6.2 Conversely, there were no significant changes between fiscal years in the remaining two dimensions of training value: (d) training applicability to job, and (e) perceived increases in knowledge and skills. While participant ratings of both items remained statistically the same in FY03 and FY04, in FY04 average training applicability to job fell below the most recent ASTD benchmark. In particular, ratings of job applicability and knowledge and skills were significantly lower among *P&T Networks* and *Regions* in FY04.

ACTIVITY FEATURES EXPLAINING TRAINING QUALITY

6.3 We analyzed the determinants of training quality in order to make more precise recommendations for improving staff training. Our model results indicate that perceived training quality is highly affected by instructional design and implementation: training content, order of presentation and training materials. The higher the activity was rated on these features, on average, the more positively participants rated the quality of the course they attended. The effect of training content was stronger among activities with larger class sizes.

6.4 Two other course features were significantly related to perceived training quality: course length and class size. On average, we found that training was rated more positively among participants in learning activities with greater than two days duration and courses with fewer than 21 participants. These findings confirm results reported in the annual review of staff learning for FY03.³⁷

6.5 In contrast to some of the literature on adult learning, we did not find any evidence to support the claim that "action learning" enhances participant satisfaction with training. That is, courses described by the course task manager as using interactive learning techniques were not more highly rated by participants. This may be partially due to social desirability bias and over-reporting of interactive learning. A comprehensive summary of IEG evaluations of WBI client training shows that action learning is a significant predictor of training effectiveness when the sources of information about

³⁷ See Annex I for activity means for key determinants of activity quality by training sponsor.

action learning are course participants, rather than course task managers. Reported use of action learning approaches was found to be significantly higher among task managers than participants who recalled using action learning.

PARTICIPANT VARIABLES RELATED TO TRAINING QUALITY

6.6 Finally, we found that participant demographics and characteristics were related to perceived training quality: (a) staff grade level; (b) motivation for taking the course; and (c) learning plan status. ACS staff were more likely to rate courses higher in quality than were higher level professional staff. This effect varied neither by course nor learning provider category. Participants who reported their primary motivation for taking the course as enhancing their performance in their current work program rated the courses more positively than participants with some other motivation for learning (i.e. professional growth/interest, networking, or “other”.) Participants who registered their activity in their individual learning plans rated training quality significantly higher than their counterparts who did not report the activities as part of their learning plan. This result also underscored findings from FY03.

ANNEX A: ACTIVITY QUESTIONNAIRE

Add a course to the Level 1 database

Event Title

Task Manager(s) for learning event

Learning Coordinator

Learning Catalog Code

Format: LLLNNN-NN-NNN (L= letter ; N= number)

Event Dates and Times

Start Date

Start Time:

End Date

End Time:

Activity Location (Building / Room)

Number of Days

Number of participants expected to attend learning event

Are at least half of the expected participants World Bank staff members?

Yes

No

Would you like a separate evaluation of the instructor?

Yes

No

If you answered "yes" to the optional instructor's evaluation, please provide the names of the instructors to be evaluated:

Order of Instructors by appearance in the course	Name of instructor to be evaluated	Title of corresponding session
1		
2		
3		
(Please add rows as needed)		

Please select the sponsor(s) for this learning event

Regional Vice - Presidencies:

AFR

EAP/SAR

ECA

LCR

MENA

Network Anchors:

ACS ESSD FSE HDN INF ISG OPCS PREM/DEC

Other:

MGT SLC EXT PI GSD HRS LEG OED ACT WBI

Other, specify:

In which one of the following categories does the training content best fit?

- Information & Technology
- Executive Management
- Professional & Technical
- Trust Funds
- Behavioral & Social (interpersonal communication skills)
- Bank Operations

Will the training approach primarily employ “Interactive “ Learning? (Interactive Learning is a process of learning and reflection that happens with the support of a group of peers working together with real business problems with the intention of accomplishing tasks while enhancing the learning of the group at the same time.)

Yes No

How will this event be delivered?

- Face to Face
- Distance Learning
- Combined (F2F and DL)

Was this event designed by the World Bank?

Yes No

Will the event be facilitated by the World Bank?

Yes No

Combined (Bank Staff and Outside Consultants)

Was this course designed especially for Managers?

Yes No

How many times has this event been delivered in the past?

- First time this event will be delivered
- Second time this event will be offered
- Two or more times



ANNEX B: PARTICIPANT QUESTIONNAIRE

Level 1 Evaluation Questionnaire for Staff Learning

Please assess the course to help to improve future training. To answer, please fill in the circle completely (●). If you wish to change an answer, fully erase it or draw an X over the unwanted mark and fill in the circle indicating your preferred answer. Please choose **only one answer** per question.

Training Title:

Training Dates:

1. What was your **main reason** for taking this training? (Fill **only one** circle.)

To enhance performance in current/planned assignment

To network and share information

For professional interest and growth

Other

2. Do you work for the **World Bank**? ? Yes, at HQ ? Yes, in the Field ? No *If no, go to Q6.*

3. For what **part** of the World Bank Group do you primarily work? (Fill **only one** circle.)

<i>Regional Vice-Presidencies</i>		<i>Network Anchors</i>		<i>Other</i>	
<input type="radio"/> AFR	<input type="radio"/> LCR	<input type="radio"/> ESSD	<input type="radio"/> OPCS	<input type="radio"/> DEC	<input type="radio"/> IFC
<input type="radio"/> EAP	<input type="radio"/> MENA	<input type="radio"/> FSE	<input type="radio"/> PREM	<input type="radio"/> HRS	<input type="radio"/> MIGA
<input type="radio"/> ECA	<input type="radio"/> SAR	<input type="radio"/> HDN	<input type="radio"/> PSD	<input type="radio"/> LEG	<input type="radio"/> EXT
		<input type="radio"/> ISG	<input type="radio"/> INF	<input type="radio"/> WBI	<input type="radio"/> Other

4. What is your **grade** level?

GA-GD GE GG ETC / STC

ETT / STT GF GH or above Other

5. Was this training agreed upon in your **Learning Plan** (LP)?

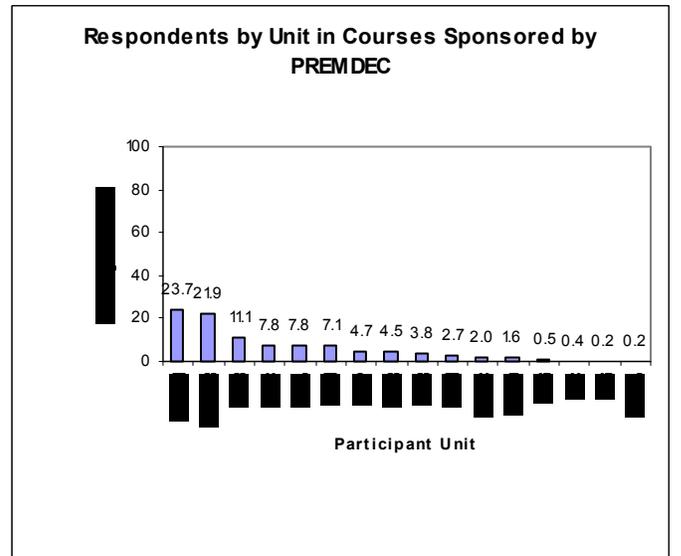
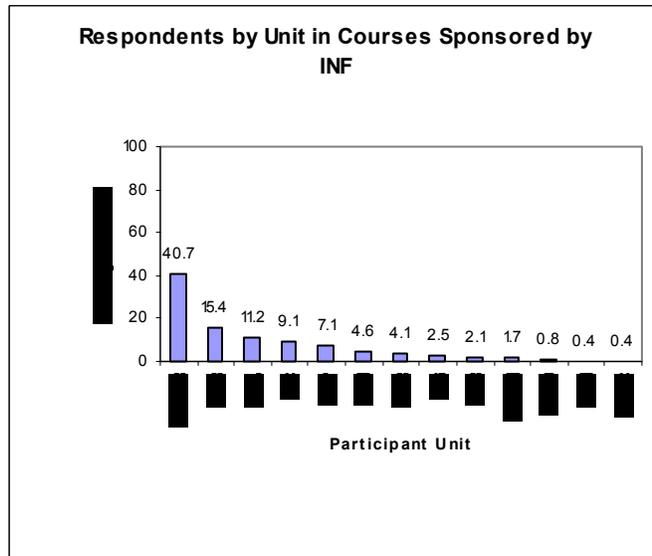
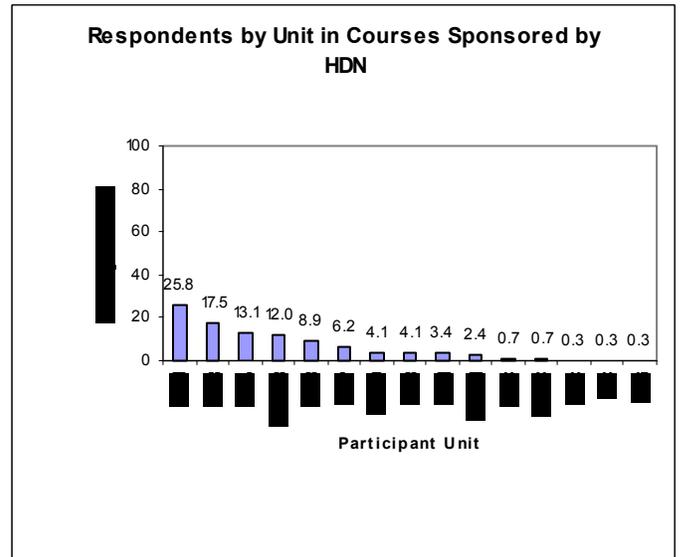
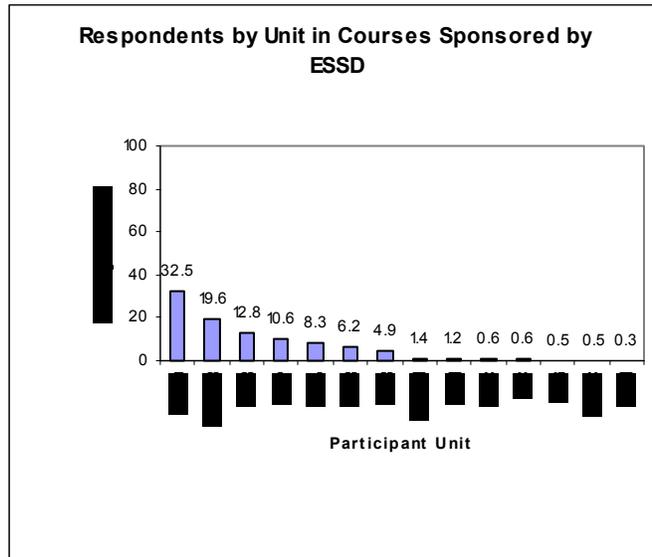
? Yes ? No I don't have a LP I don't know

Using the scale on the right, please rate each question/statement below.

	Not at all						Very much
6. To what extent did the training fulfill your learning needs ?	①	②	③	④	⑤	⑥	⑦
7. To what extent did the training achieve its announced objectives ?	①	②	③	④	⑤	⑥	⑦
	Very poor			Average			Very good
8. How would you rate the training content or subject matter?	①	②	③	④	⑤	⑥	⑦
9. How would you rate the order in which the content was presented ?	①	②	③	④	⑤	⑥	⑦
10. How would you rate the materials used during the training?	①	②	③	④	⑤	⑥	⑦
	Very low			Medium			Very high
11. How would you rate the overall quality of the training?	①	②	③	④	⑤	⑥	⑦
12. How would you rate the overall usefulness of the training?	①	②	③	④	⑤	⑥	⑦
13. How would you rate the relevance of this training to the Bank's mission ?	①	②	③	④	⑤	⑥	⑦
	Strongly Disagree			Neither			Strongly Agree
14. My knowledge/skills increased as a result of this training.	①	②	③	④	⑤	⑥	⑦
15. The knowledge/skills gained through this training are directly applicable to my job .	①	②	③	④	⑤	⑥	⑦

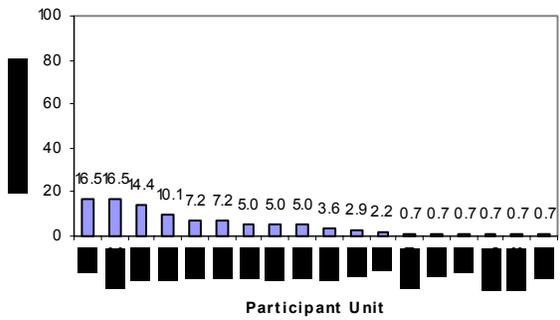
ANNEX C: COURSE AUDIENCE: DISTRIBUTION OF RESPONDENT UNIT, BY ACTIVITY SPONSOR

P&T NETWORKS

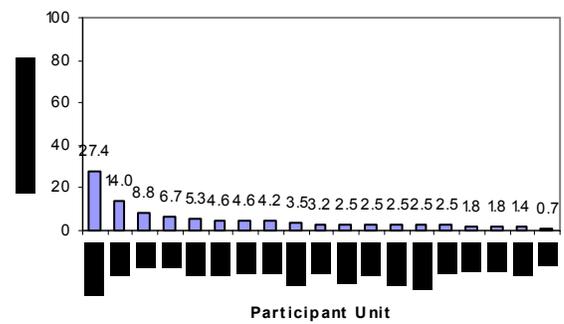


OTHER NETWORKS

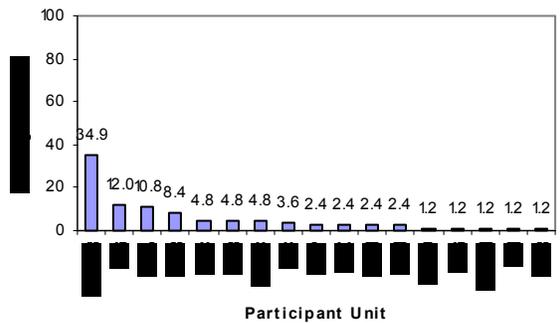
Respondents by Unit in Courses Sponsored by ACS



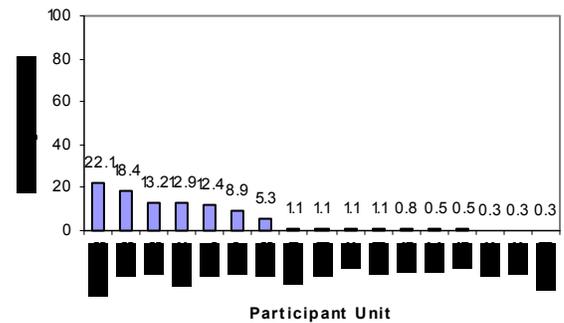
Respondents by Unit in Courses Sponsored by ISG



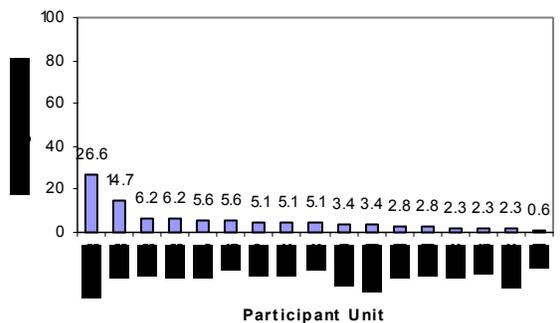
Respondents by Unit in Courses Sponsored by MGT



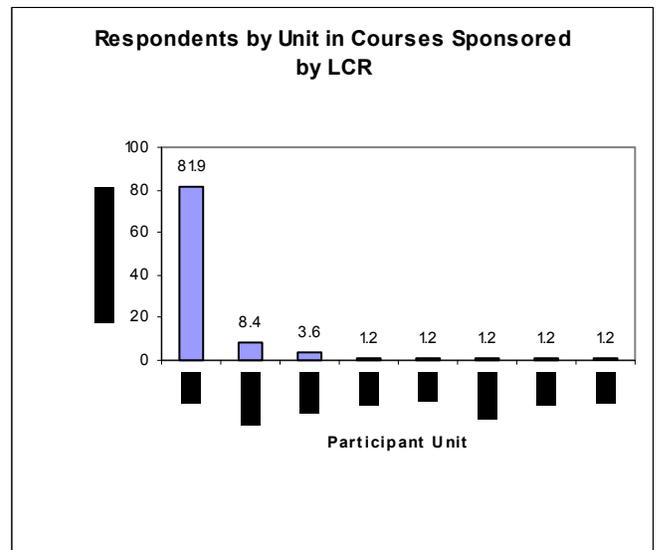
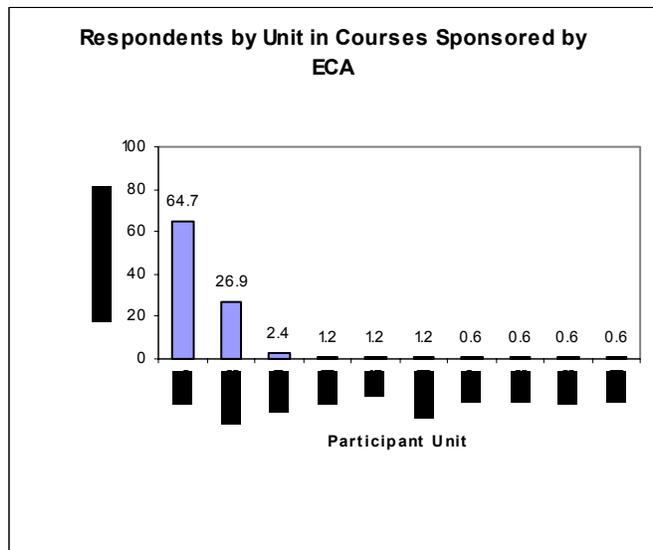
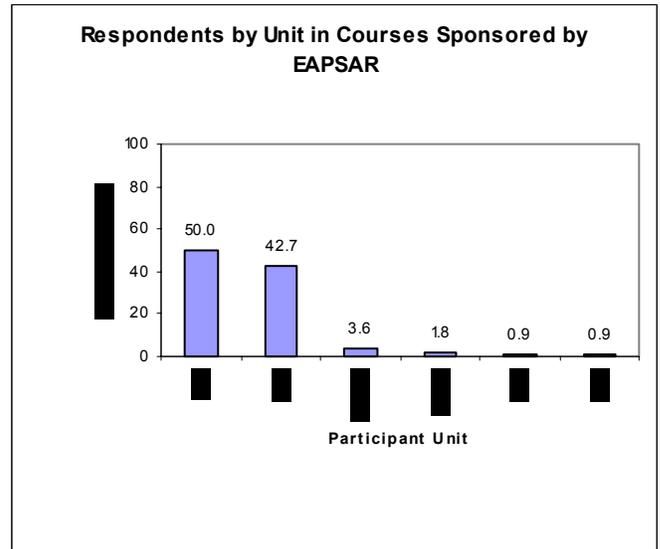
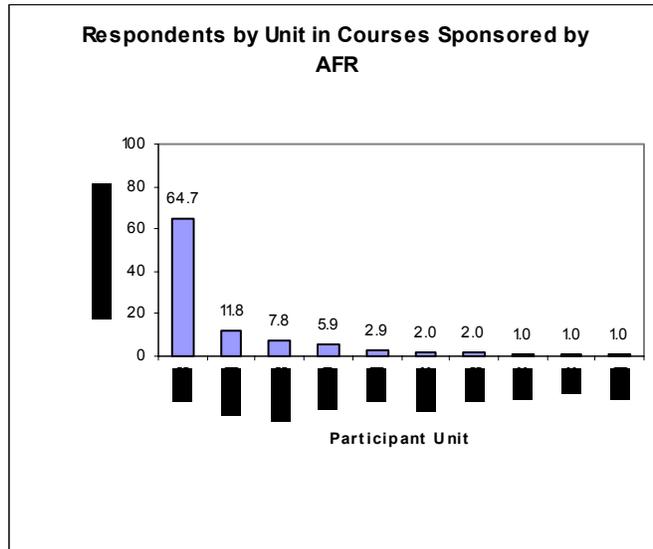
Respondents by Unit in Courses Sponsored by OPCS



Respondents by Unit in Courses Sponsored by PI

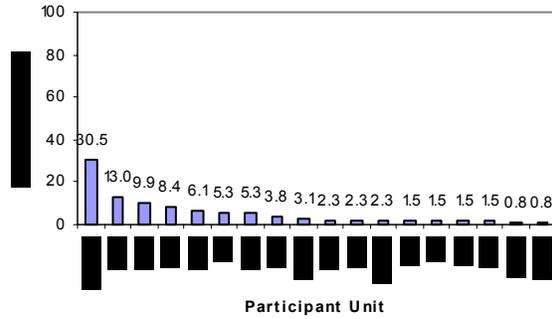


REGIONS

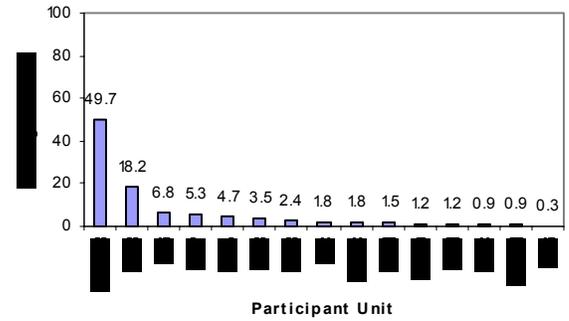


OTHERS

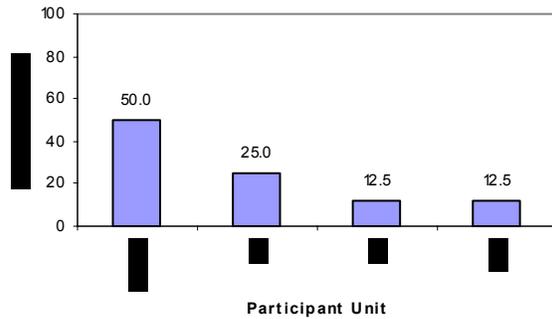
Respondents by Unit in Courses Sponsored by ACT



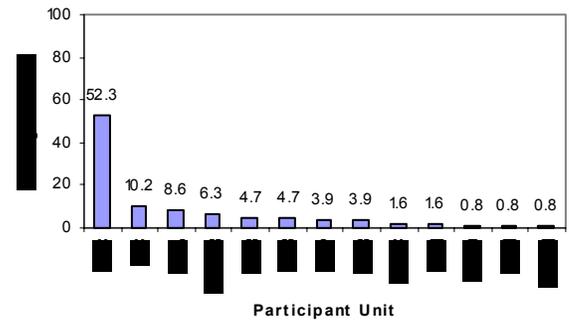
Respondents by Unit in Courses Sponsored by EXT



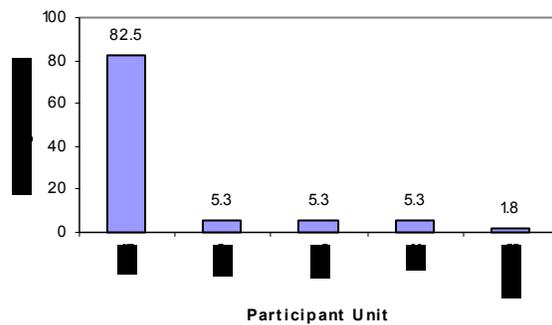
Respondents by Unit in Courses Sponsored by GSD



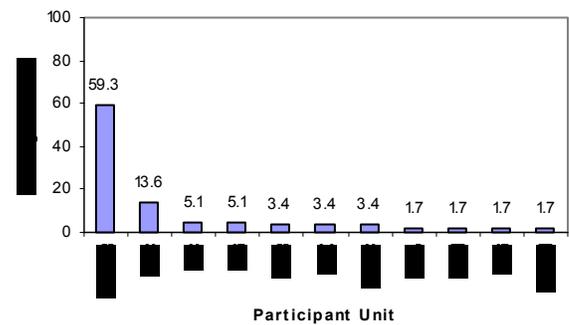
Respondents by Unit in Courses Sponsored by HRS



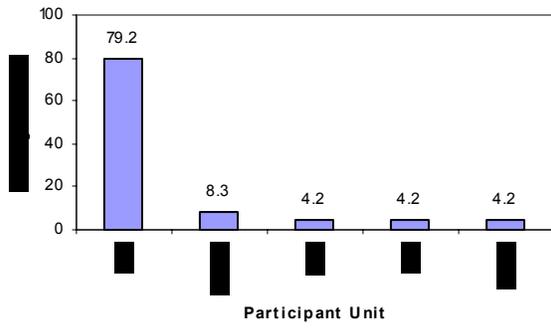
Respondents by Unit in Courses Sponsored by LEG



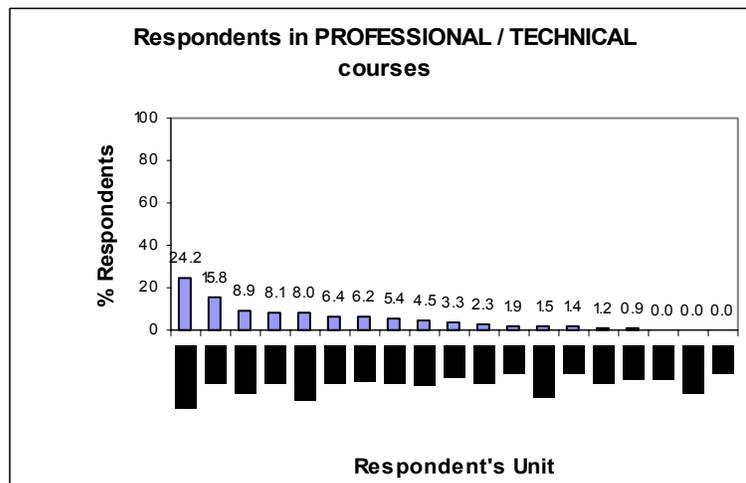
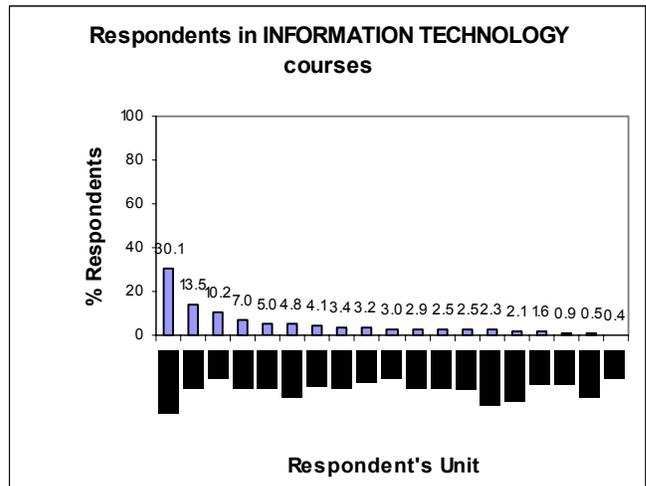
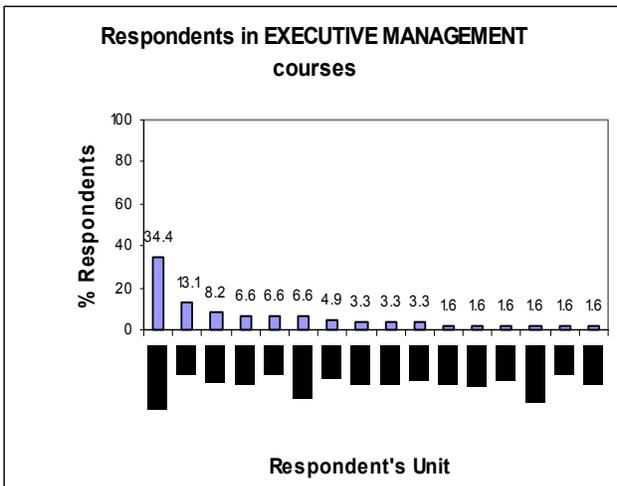
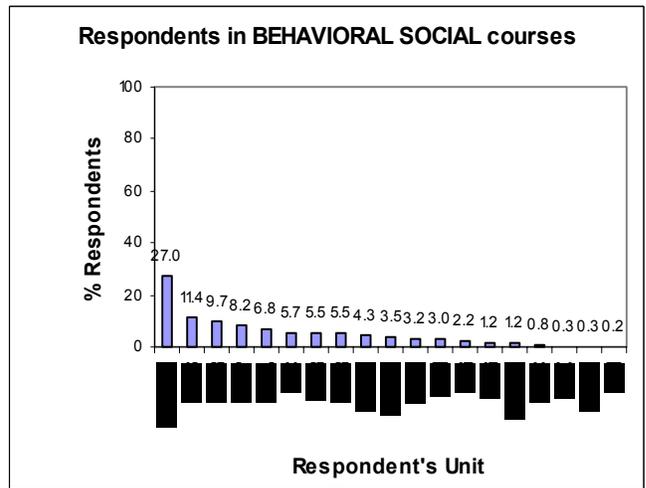
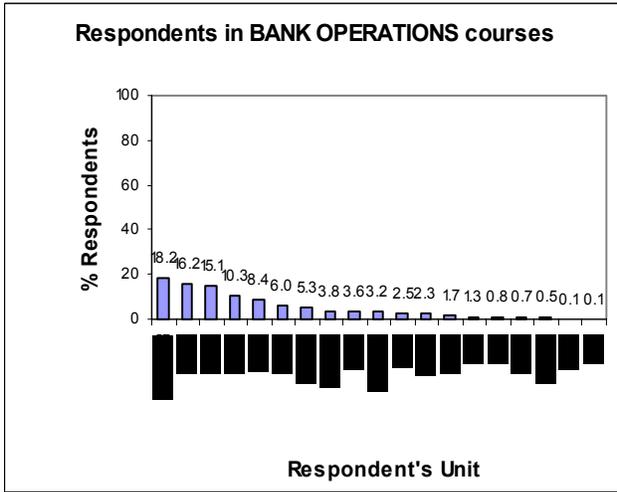
Respondents by Unit in Courses Sponsored by SEC



**Respondents by Unit in Courses Sponsored
by WBI**



ANNEX D: DISTRIBUTION OF RESPONDENTS BY UNIT IN ACTIVITY CONTENT TYPE



ANNEX E: RATINGS FOR JOINTLY SPONSORED COURSES IN FY04

Note: For Columns 1 And 2, ● = Higher Than Bank/ASTD Average, ◐ = Bank/ASTD Average, ○ = Below Bank/ASTD

Ratings of FY04, Q1-Q4, Learning Programs by Activity Sponsor (Respondent level), listed in order of overall aggregate index						
<i>Joint Sponsors</i>	Increase Knowledge And Skills ³⁸	Job Applicability ³⁹	Overall Quality ⁴⁰	Overall Usefulness*	Relevance toThe Bank's Mission *	No. of Respondents
ACT ISG	●	●	●	●	●	11
ECA OPCS	●	●	●	●	●	34
LCR LEG	●	●	●	●	●	12
EAP-SAR PI	●	◐	●	●	●	21
OED LCR	◐	●	●	●	●	37
OPCS HRS PI	◐	●	●	●	●	8
AFR PREM-DEC EAP-SAR LCR	◐	◐	●	●	●	21
ESSD EXT	◐	◐	●	●	●	8
ESSD HDN	◐	◐	●	●	●	15
EXT PI	◐	◐	●	●	●	58
MGT LEG	◐	◐	●	●	●	21
PREM-DEC PI	◐	◐	●	●	●	4
PREM-DEC WBI	◐	◐	●	●	●	23
ESSD AFR	◐	●	●	●	○	11
INF PREM-DEC	◐	○	●	●	●	37
PSD AFR	○	◐	●	●	●	9
ESSD WBI	◐	◐	○	●	●	39
INF ESSD WBI	◐	◐	●	○	●	16
HDN AFR	○	◐	○	●	●	18
OPCS HDN	◐	◐	○	○	●	20
OED AFR	◐	◐	○	○	○	11
PSD PREM-DEC WBI	◐	◐	○	○	○	6
OPCS ACT	◐	◐	○	○	○	8
AFR PREM-DEC	○	○	○	○	●	55
HDN WBI AFR	◐	○	○	○	○	23
LCR PREM-DEC	○	○	○	○	○	18
HRS PI	○	○	○	○	○	39

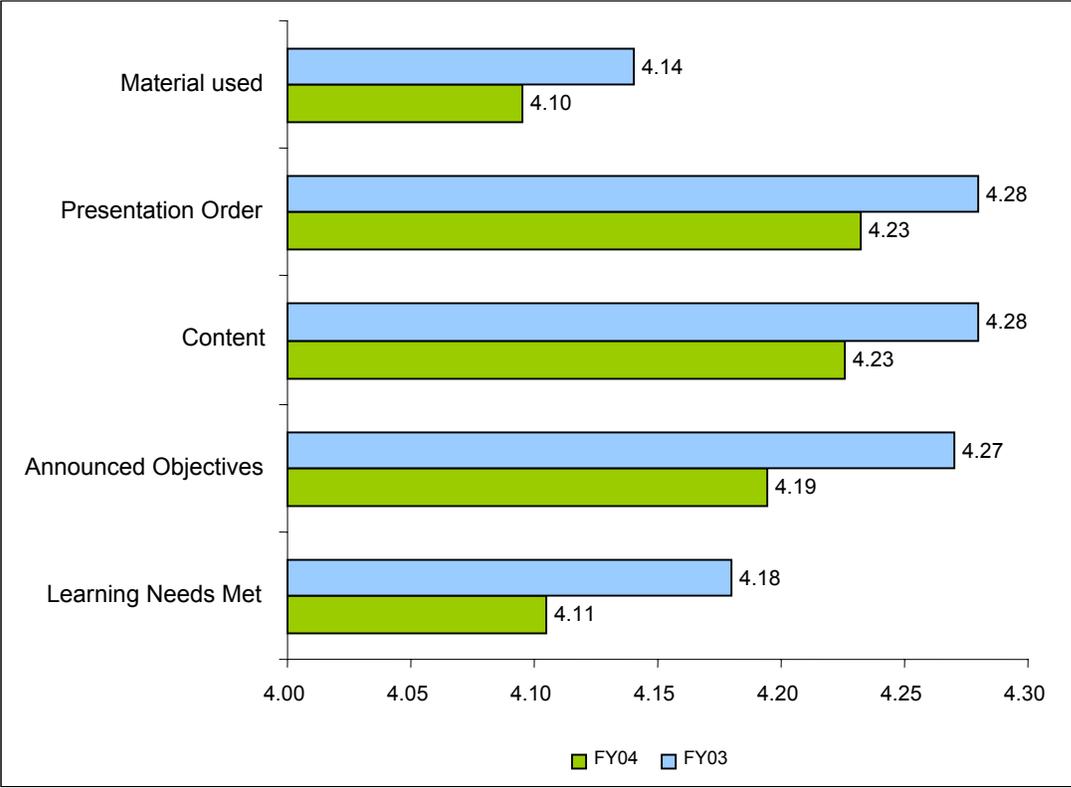
Average; For Columns 3, 4 And 5, ● = Meets Or Exceeds Bank Benchmark, ○ = Does Not Meet Bank Benchmark.

³⁸ The benchmark is the Bank average rating (4.20), statistically the same as the ASTD standard (4.21).

³⁹ The benchmark is the Bank average rating (4.18), statistically the same as the ASTD standard (4.21).

⁴⁰ The benchmark is 85% of respondents in a course gave ratings of 4 or 5.

**ANNEX F: MEAN RATINGS OF ACTIVITY CHARACTERISTICS,
FY03-FY04 (1=LOW; 5=HIGH)**



ANNEX G: VARIABLE DEFINITIONS

Variable Name	Definition	<u>Coding</u>
<i>Overall Training Quality</i>	Participant ratings of the training's applicability to job, quality, usefulness, relevance to Bank's mission; and perceived increase in knowledge and skills	Higher scores indicate higher overall training quality, coded between zero and one.
<i>Bank Operations (BO)</i>	Training content focuses on Bank operational work	1 if Bank Operations; zero otherwise
<i>Professional Technical (PT)</i>	Training content focuses on professional and technical topics (e.g. macro-economics)	1 if Professional Technical; zero otherwise
<i>Behavioral Social (Behavioral)</i>	Training content type is behavioral and social (e.g. communication etc.)	1 if Behavioral Social; zero otherwise
<i>WBdesign</i>	World Bank designed the course	1 if designed by WB; zero otherwise
<i>Total Participants (Totpart)</i>	Total participants in the course	Number of participants
<i>Course Days (ndays)</i>	Course duration measured in days	Number of days in training
<i>Training Content (Content)</i>	Participant rating of training content	Higher scores indicate higher training content quality
<i>Presentation Order (Order)</i>	Participant rating of the order in which topics were presented	Higher scores indicate better ratings of the sequence in which the material was presented
<i>Course Materials (Materials)</i>	Participant rating of the course materials	Higher scores indicate higher ratings of course material
<i>Interactive Learning (Interactive)</i>	Interactive learning was used in course	1 if interactive learning was used; zero otherwise
<i>ACS staff (Grade_AD)</i>	Participant grade level A, B, C, or D	1 if participant was in grades A-D; zero otherwise
<i>Headquarters staff (HQ)</i>	Participant based in headquarters (versus field)	1 if participant is based in headquarters; zero otherwise
<i>Participant Motivation (Motivation Work)</i>	Participants' motivation for taking the course was to enhance performance in current work load	1 if motivation was current or planned work assignment; zero otherwise
<i>Learning Plan (LearnPlan)</i>	Participants who registered the activity in their individual learning plans	1 if participant reported that activity was registered in learning plan; zero otherwise
<i>Baseline Sponsor (Sponsor)</i>	Activities in the baseline sponsor category, rated below Bank benchmarks at both the respondent and course level (INF & ESSD)	1 if activity is in baseline sponsor category; zero otherwise

ANNEX H: HLM BETWEEN-GROUP MODEL TESTING CROSS-LEVEL EFFECTS BETWEEN GRADE LEVEL AND COURSE TOPICS

Fixed Effects	Model 1	
	Gamma Coefficient	Reliability estimate
Intercept	0.8246**	.841
Sponsor	-0.0151	
<i>Participant Motivation (work)</i>	.0248**	
<i>Based in HQ</i>	.0056	
<i>Grade A –D</i>		.195
Intercept	.0092	
Professional Technical	-.0024	
Bank Operations	.0067	
Behavioral	-.0026	
Sponsor	-.0203	
<i>Learning Plan</i>	.0081**	
<i>Total Participants</i>	-.0012**	
<i>Course Days</i>	.0120**	
Designed by WB	.0083	
Interactive Learning	-.0006	
Professional Technical	-.0148	
Bank Operations	-.0110	
Behavioral	.0111	
Training Content	.3906**	.277
Presentation Order	.1908**	.299
Course Materials	.1181**	.149
Random Effect	Standard deviation	Variance
Mean quality	.0555**	.00309
Grade A-D	.0329**	
Training Content	.1605*	
Presentation Order	.1786*	
Course materials	.0920*	
R	.07055	.00498

**p<.01, *p<.05;

bold italics indicates grand mean centered;

bold indicates group mean centered

ANNEX I: KEY DETERMINANTS OF ACTIVITY QUALITY, BY SPONSOR

Sponsor	Activity Means				
	Training Content	Presentation Order	Materials Used	Total No. of Participants	Number of Days
ACS	4.18	4.27	4.18	14.36	1.91
AFR	4.33	4.17	4.17	23.83	3.33
EAP	4.63	4.50	4.63	19.75	5.00
ECA	4.11	3.78	3.50	22.89	3.00
ESS	4.14	3.83	3.79	26.14	1.26
EXT	4.11	4.16	3.94	21.70	1.25
GSD	4.00	4.00	4.00	9.00	3.00
HDN	4.32	4.17	4.00	26.47	1.00
HRS	4.33	4.00	4.11	16.56	2.33
INF	4.17	4.25	4.36	28.42	1.17
ISG	4.53	4.50	4.38	8.28	1.14
LCR	4.50	4.50	4.25	19.75	3.40
LEG	4.50	4.33	4.33	21.00	1.50
MEN	4.00	4.00	4.00	19.00	5.00
MGT	4.50	4.25	4.25	11.75	2.25
PI	4.38	4.31	4.38	13.31	3.19
PRE	4.38	4.28	4.03	27.13	1.91
SEC	3.50	3.50	3.25	17.50	1.25
WBI	3.00	3.00	3.50	15.50	1.50
OPC	4.11	4.15	4.04	18.44	2.26
ALL	4.25	4.17	4.11	20.77	1.85