Public Disclosure Authorized

Public Disclosure Authorized

31201

Trust Fund Learning and Accreditation Program Pilot Formative Evaluation

Nidhi Khattri
Zhengfang Shi
Maria Bina Palmisano
Melissa Fein

WBI Evaluation Studies
No. EG03-065

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

December 20, 2002

ACKNOWLEDGEMENTS

This formative evaluation was prepared for the Learning Board under the overall guidance of Marlaine Lockheed, Manager, Evaluation Group. The team was led by Nidhi Khattri, with support from Zhengfang Shi, Maria Bina Palmisano, Melissa Fein, and Robert Lissitz. The TLAP Project Management Team, led by Saadat Siddiqi, provided assistance in all aspects of the evaluation. The Trust Funds Four Pillars offered invaluable advice on the development of this evaluation report.

The success of TLAP Pilot is due to a concerted and persistent team effort coordinated by the Project Management Team. This evaluation acknowledges and thanks those who participated in and/or provided advice on TLAP development and who are listed by name in the following page: the Trust Fund Pillars, Project Steering Committee, Project Management Team, Consultative Group, Content Developers, Editorial/Exam Board and Evaluation Team, as well as participants who provided feedback on TLAP Pilot.

Vice President, WBI	Ms. Frannie Léautier
CLO, Learning Board	Ms. Phyllis Pomerantz
Manager, WBI Evaluation Group	Ms. Marlaine Lockheed
Task Manager	Ms. Nidhi Khattri

TLAP DEVELOPMENT TEAMS

TRUST FUND PILLARS

Paul Cadario, TQC; Brian P.Quinn, ACTTF; Ian Wright, TFS; Arif Zulfiqar, TFO

PROJECT STEERING COMMITTEE

Hartwig Schafer (Chairman), Rakesh Asthana, Eve Bosak, Richard Calkins, Thomas Duvall, Caroline Harper, Kundhavi Kadiresan, Phyllis Pomerantz, Jean-Jacques Raoul, Nigel Twose

PROJECT MANAGEMENT TEAM

Saadat Siddiqi (Team Leader), Dominick Egan, David Gray, Nidhi Khattri, Donald MacDonald, Fareeba Mahmood, Jeanette Murry, Subramanian Natarajan, Maria Bina Palmisano, Zhengfang Shi, Samuel Taylor

CONSULTATIVE GROUP

Salahuddin Babu Ahmad, Shanu Biswas, Carol Fillar Bonney, Mark Cackler, Anjali Chadha, Marjorie Dobson, Yoko Eguchi, Philip English, Janet Entwistle, Ines Garcia-Thoumi, James Garfield, Christopher Gerrard, Lucia Grenna, Hennie van Greuning, Agnieszka Grudzinska, Akhtar Hamid, Amanda Humm, Beverley Hyatt-Allen, Kundhavi Kadiresan, Venkateswarlu Kalicheti, Stephen Karam, Catherine Kleynhoff, Toby Linden, Nishi Mendis, Kathleen Mikitin, Jose Molina, Suzanne Morris, Karin Nordlander, Jaime Olazo, Gylfi Palsson, Rajashree Paralkar, Jeeva Perumalpillai-Essex, Karlene Phillips, Nancy Pinto, Nicola Renison, Jane Distelhorst Sansbury, Keiko Sato, Narendra Sharma, Larry Smucker, Gertrude Ssali, Dafna Tapiero, John Todd, Yoshiko Urakawa, Carolyn Winter, Mark Woodward

CONTENT DEVELOPERS

Marina Badrich, Pamela Bigart, Anders Bjorgung, Zafar Farouk, James Garfield, Dale Hill, Albert Howlett, Patricia Klyestruber, Yolaine E. Joseph, Irena López, Patricia Mcgowan, Isabel Mignone, Suzanne Morris, Moustapha Nadiaye, Poyyapakkam Ravi, Sajjad Shah, Velayutham Vijayaverl, Mark Woodwrad

EDITORIAL/EXAM BOARD

Marina Badrich, Pamela Bigart, Barun Parasad Chaterjee, Diana Corbin, Thomas Duvall, James Garfield, Caroline Harper, Dale Hill, Albert Howlett, Tustomu Kudo, Irina Lopez, Isabela Mignone, Kenneth Miller, Suzanne Morris, Herman Nissenbaum, Sajjad Shah, Velayutham Vijayaverl, Mark Woodward

EVALUATION TEAM

Nidhi Kahttri (Team Leader), Melissa Fein, Robert Lissitz, Maria Bina Palimisano, Zhengfang Shi

ACRONYMS

ACS	Administrative Client Support
ACTTF	Accounting Trust Funds Division
CG	Consultative Group
CTF	Consultant Trust Fund
LOR	Letter of Representation
OP/BP	Operational Policy/Bank Procedures Statement
OPCS	Operational Policy and Country Services
OVP	Operational Vice President
PMT	TLAP Project Management Team
PSC	Project Steering Committee
TF	Trust Fund
TLAP	Trust Fund Learning and Accreditation Program
TFO	Trust Fund Operations Department
TFS	Trust Fund Strategy and Donor Relations Unit
TMS	TLAP Task Management System
TOR	Terms of Reference
TQC	Trust Fund Quality Assurance and Compliance Unit
TRS	Time Recording System
TTA	Task Team Assistant
TTL	Task Team Leader
VPU	Vice Presidential Unit
WBI	World Bank Institute
WBIEG	World Bank Institute Evaluation Group
WebCT	A software system used to create online learning (e-learning) courses

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **II**

TLAP DEVELOPMENT TEAMS ..... **III**

ACRONYMS ..... **IV**

TABLE OF CONTENTS ..... **V**

EXECUTIVE SUMMARY ..... **VII**

MODULE QUALITY VII
QUALITY OF THE SELF-ASSESSMENT TEST AND THE ACCREDITATION EXAM VIII
PARTICIPANT LEARNING AND PASSING RATES IX
OVERALL QUALITY OF THE PILOT X
CONCLUSIONS XI
STRATEGIES FOR ROLLOUT XI
DIRECTIONS FOR FUTURE MONITORING AND EVALUATION OF TLAP XII

1. INTRODUCTION **1**

TLAP OBJECTIVES 1
TLAP DESCRIPTION 2
TLAP PILOT ORGANIZATION 4

2. EVALUATION OBJECTIVES AND METHODOLOGY **5**

EVALUATION OBJECTIVES 5
EVALUATION QUESTIONS 5
EVALUATION METHODOLOGY AND DATA COLLECTION 5

3. LEARNING MODULES' QUALITY AND IMPLEMENTATION **9**

CONTENT CHARACTERISTICS 9
USABILITY CHARACTERISTICS 11
OVERALL ASSESSMENT OF MODULE QUALITY 11
PARTICIPANT COMMENTS 11
E-LEARNING: ADDITIONAL COMMENTS FROM FEEDBACK SESSIONS 12
CONCLUSIONS AND RECOMMENDATIONS 13

4. ACCREDITATION EXAM QUALITY **15**

VALIDITY OF THE EXAM 15
RELIABILITY OF THE EXAM 16
ITEM QUALITY 17
MECHANISMS/ACTIONS FOR QUALITY IMPROVEMENT 19
METHOD OF TEST ASSEMBLY FOR THE ACCREDITATION EXAM IN LATER ROLLOUT 19
CONCLUSIONS AND RECOMMENDATIONS 20

5. LEARNING GAINS AND PASSING RATES **23**

LEARNING GAINS ANALYSIS 23

Passing Score Standard 24
Passing Rates 26
Conclusions and Recommendations 27

6. OVERALL QUALITY OF THE PILOT 29

Overall Quality of the Learning Modules 29
Overall Quality of the Pilot 30
The Self-Assessment Test and the Pilot Exam 31
Conclusions and Recommendations 32

7. SUMMARY AND CONCLUSIONS 35

Summary of Evaluation Findings 35
Conclusion 38
Issues and Recommendations for the Final Rollout 39

LIST OF ANNEXES 43

A. List of Modules and Learning Objectives 43
B. Pilot Schedule 43
C. End-of-Module Level-1 Checklist (Sample from Module 1) 43
D. End-of-Pilot Level-1 Questionnaire 43
E. Level-1 Survey Questions for the Self Assessment Test 43
F. Level-1 Survey Questions for the Accreditation Exam 43
G. End-of-Pilot Focus Group Questions 43
H1. Item Analysis Results of Self-Assessment Test 43
H2. Item Analysis Results of The Accreditation Exam 43
I. Selected Internal Reports and Documentation 43
A. List of Modules and Learning Objectives 45
B. Pilot Schedule 49
C. End-of-Module Level-1 Checklist (Sample from Module 1) 51
D. End-of-Pilot Level-1 Questionnaire 55
E. Level-1 Survey Questions for the Self-Assessment Test 57
F. Level-1 Survey Questions for the Accreditation Exam 59
G. End-Of-Pilot Focus Group Questions 61
H1. Item Analysis Results Of Self-Assessment Test 63
H2. Item Analysis Results Of The Accreditation Exam 65
H2. (continued) Item Analysis Results of the Accreditation Exam 66
I. Selected Internal Reports and Documentation 67

EXECUTIVE SUMMARY

In 2000 the World Bank carried out a comprehensive Trust Fund Process review, which identified a pressing need to enhance Bank staff knowledge of the management and application of trust funds (TFs) and related World Bank policies and procedures. It was recommended that the best way to guarantee an adequate level of policy and procedural knowledge would be to require that TF management be entrusted to accredited staff only. The Trust Fund Learning and Accreditation Program (TLAP) was developed in response to this recommendation. A pilot for this program geared for Task Team Leaders (TTLs) was conducted in October 2002, and the program rolled out in December, 2002.

The purpose of this formative evaluation was to assess the quality of the design and implementation of the key components of the program, to provide information to TLAP Project Management Team (PMT) for program improvement prior to rollout, and to provide information to the Learning Board regarding the progress of the program. The evaluation conducted by WBIEG was mainly based on the Pilot, which comprised a learning system of 17 online modules, a self-assessment test, and an accreditation exam based on these learning modules. The reports supporting this formative evaluation have been received by PMT, and recommendations from these reports have already been incorporated into the preparation for rollout.

This formative evaluation report covers the following aspects of the Pilot:

- Module quality and implementation of TLAP learning modules
- Quality of the self-assessment test and the accreditation exam
- Participant learning and passing rates
- Overall quality of the Pilot
- Conclusions pertaining to readiness for rollout
- Strategies for rollout
- Future directions in monitoring and evaluation of TLAP.

These aspects were evaluated by utilizing a variety of strategies including the following: self-administered surveys and checklists; collections of comments and focus-group results; self-assessment test and accreditation exam data; and meeting with the Exam Board, content experts and education specialists to set or modify the accreditation exam standards.

MODULE QUALITY

In general, the TLAP learning modules were of sufficient quality in their content, design, and implementation. On average, more than 85 percent of the respondents gave affirmative or favorable ratings on such questions as content characteristics (including *achievement of objectives, logical sequence, clarity of presentation, and relevance of information*) and usability characteristics of the e-learning mode (including *menu navigation,*

backward/forward navigation, hotlinks use, and dynamic flowchart) of the 16 modules. The main message extracted from the participants' comments indicates that the learning modules were well organized and presented, and that the e-learning approach was innovative and much better than expected.

An average of 72 percent of the respondents gave favorable ratings on the overall quality of the individual modules. This is lower than the Bank's benchmark, but may be due to the Pilot program's unique content, high stakes and e-learning mode. The ratings on usability characteristics were slightly lower than the ratings on content characteristics. There are currently no established Bank standards for e-learning usability because online learning is still a new technology and to some extent, its use is a matter of taste and habit. In addition, the ratings on these characteristics varied among different modules. These issues have been reviewed by PMT in combination with other comments from the Pilot participants.

Other than the positive feedback, the four most important comments coming from the Pilot participants were: (a) some of the information in the learning modules was incorrect; (b) some of the answers to the questions in the learning scenarios were ambiguous; (c) supplemental modalities to online learning should be provided; and (d) a level of support should be maintained for future participants. In response, PMT reviewed the related modules and questions and made appropriate revisions. To address the exclusive usage of e-learning as the learning platform, PMT has enabled the printing of the learning modules so that future participants can have hard copies to study. An email hotline has been established to provide support for future participants. In addition, PMT will watch the situation closely to see what modifications might be needed in the rollout.

QUALITY OF THE SELF-ASSESSMENT TEST AND THE ACCREDITATION EXAM

The self-assessment test and the accreditation exam meet professional standards for quality. Because the self-assessment test and the accreditation exam reflect the same learning objectives, cover the same knowledge domain, and were developed following the same rules and procedures, attributes of one apply to the other. Their quality is judged by their validity, reliability, and item quality. Both the exam and the test had evidence supporting validity: well-defined learning objectives and alignment of exam/test items with these learning objectives—participants judged that 86 percent of the exam was based on the learning modules. The reliability of the test (coefficient alpha equals 0.83) fell into the standard acceptable range.

The items on the self-assessment test and the accreditation exam were well aligned with the learning objectives and module contents, and had generally good item statistics. Potentially problematic items were reviewed and modified based on the item statistics and comments from participants. The final item yield was high—only four items (out of 144 exam/test items) were suggested to be excluded from future use. About 90 new items will be field-tested and similarly examined before being added to the exam item pool.

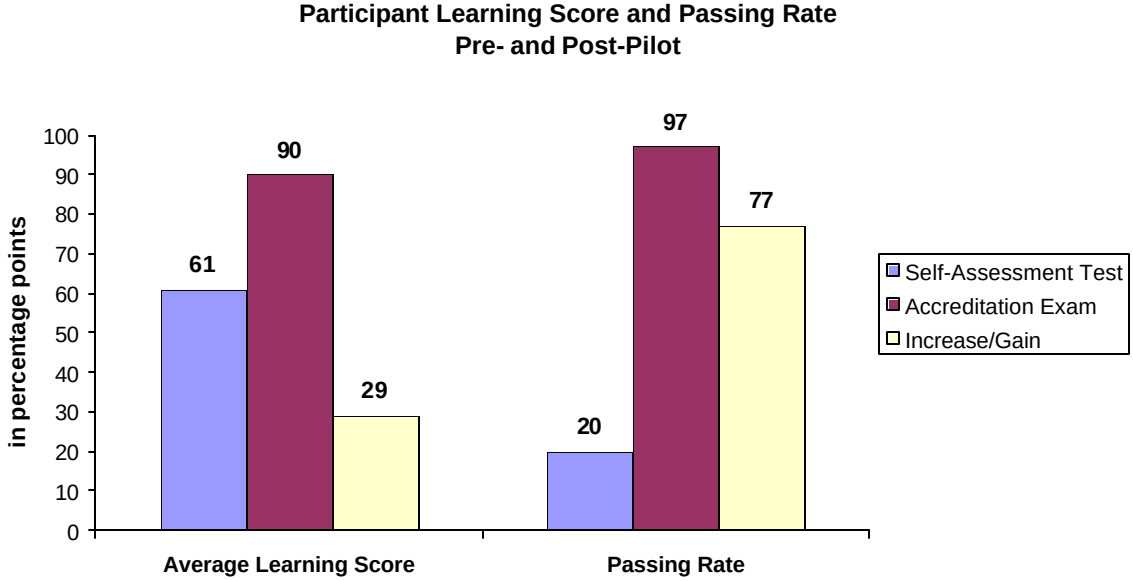
Based on this evidence and the revisions made on the items, the accreditation exams in the rollout can be assembled from the items used in the Pilot. Because of these improvements,

the accreditation exams in the rollout should have even better validity and reliability than the exam in the Pilot.

In addition, randomized test construction—an exam will be assembled by randomly selecting a set of exam items for each individual examinee—is recommended to be used in the rollout. The Pilot used fixed-form test construction—an exam consisting of the same set of pre-selected exam items was used for all the participants who were assigned to a particular form—to facilitate the collection of balanced response data for item analysis. The response data from the Pilot accreditation exam indicated that a randomly-constructed exam has comparable quality to a fixed-form exam. Moreover, it addresses the concern of security better than the fixed-form approach, allows an adequate level of fairness, and provides a convenient venue for field-testing newly developed items as well. PMT has adopted the randomization approach in the rollout.

PARTICIPANT LEARNING AND PASSING RATES

The learning modules appeared to be effective, as evidenced by high measured learning gains, high self-perceived learning gains, and a substantial increase in the passing rate after completing the learning modules. After studying the modules, the Pilot participants on average achieved a statistically significant learning gain of 29 percentage points; their passing rate increased from 20 percent on the pre-test (the self-assessment test) to 97 percent on the post-test (the accreditation exam). Participants' measured learning gains matched their self-perceived learning gains very well, providing a further evidence of the alignment of modules with the exams.



The passing score standard for the accreditation exam was initially set and later confirmed by the Exam Board at a level of 77 percent correct, equivalent to answering 37 questions correctly out of 48 exam questions. This was a policy-based decision, combined with subject matter experts' judgments on what constitutes a minimum competency to manage

TFs in the Bank. According to this standard, all but one participant (33 out of 34) passed the accreditation exam in the Pilot and will be accredited by the TFO through formal procedures.

The Pilot participants' knowledge level in TF management and their passing rates before the training varied according to gender, grade, TTL work assignment, and self-identified competency level in TF management *(expert, competent* and *novice)*. However, studying the learning modules effectively reduced these gaps for both the knowledge levels and the passing rates. Specifically:

- The self-identified *expert* group on average had the highest initial score (78 percent correct) and final score (94 percent correct), but the lowest learning gain (16 percentage points); while the self-identified *novice* group on average had the lowest initial score (49 percent correct) and final score (88 percent correct), but demonstrated the greatest learning gain (38 percentage points).
- A similar pattern was observed between non-TTLs and TTLs. The non-TTLs on average had higher initial score (72 percent correct) and final score (93 percent correct), but lower learning gain (21 percentage points), while the TTLs on average had lower initial score (56 percent correct) and final score (89 percent correct), but demonstrated higher learning gain (33 percentage points).
- No statistically significant differences in passing rates or scores by gender and grade groups were found.
- It is hard to infer from these observations how the broader population will fare in the rollout because of small sample sizes and, more importantly, the unknown representativeness of these groups in the Pilot to their respective populations. Nevertheless, a clear policy implication is that future candidates must be encouraged to study the learning modules before taking the accreditation exam.

OVERALL QUALITY OF THE PILOT

The Pilot received favorable ratings from participants on *overall achievement of objectives, overall usefulness, and overall quality* of its learning modules. The percentages of ratings of "4" or "5" on a 1-to-5 scale were 81, 87 and 78 respectively, comparable to other staff learning programs in the Bank. On the Pilot overall, a majority of the participants gave ratings of "4" or "5" on *support provided* (97 percent), *information provided* (91 percent), *quality* (84 percent), and *participant satisfaction* (88 percent). All these indicate that the whole Pilot was well-prepared and well-executed.

Participants, however, gave less favorable ratings to two other aspects of the Pilot. Fewer participants gave high ratings ("4" or "5") on *relevance of the learning modules* (68 percent) and *pacing of the Pilot* (66 percent). In addition, participants on average judged only 74 percent of the information in the learning modules was *directly applicable* to their management of TFs.

The relatively low ratings of the TLAP Pilot on relevance and applicability may not improve during the rollout. There are several reasons for this, the main one being that TTLs may not feel that they need to know all aspects of the TF management if they were only involved or interested in certain types of TFs. In other words, relevance pertaining to individual participants may not be the same as the relevance to TF management in general. TLAP is designed to equip people more broadly with the knowledge of how to manage TFs now and in the future. It is important to note that even though a TTL may not currently be using a particular TF, he/she may do so in the future.

The Pilot participants perceived the accreditation exam as relatively difficult, even though almost all of them achieved passing or higher scores. All participants used the learning modules and/or other resources during their exams, but their rating of the usefulness of the learning modules for the exam was just slightly above "helpful." In addition, participant rating on the presentation of the exam questions was only slightly better than "OK," a phenomenon that may be attributable to the limitations of the WebCT technology. These mixed perceptions and results, however, may be appropriate for such a high-stakes accreditation exam. Although the exam was not perceived easy, by studying the learning modules carefully, it is possible to pass and show significant learning gains.

CONCLUSIONS

Overall, the Pilot results indicate that the learning modules were of good quality in their content, design, and implementation; the exams were well-aligned with the learning objectives and were of good quality; and the passing score standard was reasonably set. Moreover, participants were self-motivated learners and were satisfied with quality of the Pilot and the support provided, and mostly importantly, they could pass the accreditation exam after studying the learning modules. Additionally, the comments and concerns raised by participants and other stakeholders during the Pilot have been appropriately addressed by PMT before the rollout.

The results from the formative evaluation and improvements made by PMT indicate that TLAP can now move forward with a formal rollout for accrediting the Bank's TTLs. However, rollout should not proceed without a certain amount of caution and vigilance. Areas for special attention include exam administration and online learning support. It should be noted in particular that there was a high degree of support provided in the Pilot (through email, telephone, in-class opening, review, and exam sessions), and the participants were self-selected and therefore a highly-motivated group who wanted to get accredited ahead of time. These circumstances will not be replicated in the rollout. PMT therefore has to be on the lookout for potential problems during the rollout, such as those with e-learning, content, exam, and overall program. Cautions should be taken to avoid generalizing results from the Pilot group to the full TTL population.

STRATEGIES FOR ROLLOUT

The objective of the rollout strategy is to get all approximately 1600 TTLs to take the accreditation exam before the cutoff date of June 30, 2003, and to stagger the concentration

of participants so that the system does not become overtaxed by too many participants waiting until the deadline to take the exam. The staggered rollout plan would also help assess how the program is functioning on a large scale. According to PMT, the rollout strategy encompasses the following activities:

- Communicate that accreditation is mandatory, and the deadline is June 30, 2003.
- Motivate staff to study the modules and take the exam through the use of emails, a Website (TLAP homepage), and other advertising.
- Designate TLAP Coordinators nominated in every VPU to assist with brown bag introduction sessions and "hands on sessions" for those needing more support.
- Ask VPUs to provide a schedule and plan outlining how many TTLs will take the exam and what the plan is for all to be accredited by the deadline; in addition, ask that progress be documented.
- Create an environment of competition. Progress of VPUs in accreditation will be posted on the TLAP website.
- Adopt additional strategies for Regions to facilitate introduction of the accreditation program.

DIRECTIONS FOR FUTURE MONITORING AND EVALUATION OF TLAP

Although the Pilot is of sufficient quality to proceed with rollout, continuous monitoring and evaluation activities are required to ensure that the accreditation program maintains an adequate level of quality in the rollout. Some of these evaluation activities are:

- There should be a periodic review of the learning curriculum. A mechanism should be established regarding updating the content of the modules when new concerns need to be addressed or new policies in TF management are developed.
- New items should be developed to reflect any content changes in the curriculum or to replenish the exam item pool due to the loss of items that were proved to be malfunctioning. Development of new items should follow the same rigorous procedures that were used to develop the original items.
- The exam item pool upon which the accreditation exam for the immediate rollout is based should be expanded with items that have been field-tested and reviewed by multiple subject matter experts. At the same time, module-level or stratum-level item pools should be adjusted to make the randomized exams fairer after more item difficulty data are collected.
- The exam time and passing score standard should be reviewed again, as more exam data are available to reflect the context that is different from where the Pilot was conducted and such standards were set.

- Although there are no plans to capture participant comments in the same manner as in the Pilot, some efforts should be made to get participant feedback on the exam after rollout. For example, the PMT could periodically call some TTLs and get their comments using a structured instrument. WBIEG has developed a customized Level-1 questionnaire, which will be administered to all TTLs who take the accreditation exam during the rollout. The hotline support system could be another source of information.
- And more importantly, does the TLAP reach its objectives, such as appropriate rules and policies being followed by the TTLs who have gone through the learning modules and received accreditation? An impact study is necessary after a certain period of time from the rollout in December, 2002—WBIEG will develop the design in March/April 2003, once the program logframe is finalized.

1. INTRODUCTION

1.1 In 2000 the World Bank carried out a comprehensive Trust Fund Process review, which identified a pressing need to enhance Bank staff knowledge of the management and application of trust funds (TFs) and related World Bank policies and procedures. It was recommended that the best way to guarantee an adequate level of policy and procedural knowledge would be to require that TF management be entrusted only to accredited staff. The Trust Fund Learning and Accreditation Program (TLAP) is being developed in response this recommendation, and a pilot for this program was conducted in October 2002.[1]

1.2 The Learning Board has requested the World Bank Institute Evaluation Group (WBIEG) to conduct a formative evaluation of TLAP. The purpose of the report is to provide a consolidated summary of the evaluation findings from the TLAP pilot program for the Learning Board. It combines all these specific "evaluation reports" that WBIEG had provided to the TLAP Project Management Team during different stages of the Pilot. Most conclusions and recommendations made in these formative reports have been incorporated in the preparation of the formal rollout.

1.3 The remaining part of this chapter provides an overview of the objectives and main features of TLAP, the organization of the Pilot, as well as the demographics of the participants in the Pilot. Chapter 2 explains the evaluation objectives and methodologies. Based on information collected and analytical results, Chapter 3 assesses the quality and implementation of the learning modules. Chapter 4 evaluates the quality of the self-assessment test and the accreditation exam. Chapter 5 addresses participants' learning from attending the program, and the establishment of a passing-score standard and the passing rates. Chapter 6 discusses the overall quality of the Pilot program. Conclusions and recommendations sections are included in each of these chapters. Finally, Chapter 7 summarizes overall findings from the Pilot evaluation and provides recommendations for the formal rollout in early December, 2002.

TLAP OBJECTIVES

1.4 The objectives of TLAP are to:

(a) Provide donors, Board, and Bank management the assurance that within a reasonable period of time the management of TFs will be entrusted to trained staff.

(b) Give staff the confidence that they possess the knowledge and skills necessary for proper and efficient management and execution of TFs.

[1] A full program description and details of the accreditation path can be found in two independent papers, *A Framework for the Trust Fund Learning and Accreditation Program,* and *Accreditation Exam Parameters*.

1.5 The target audience for TLAP are six main groups:

(a) Task Team Leaders (TTLs)

(b) Team Assistants working with TTLs on TFs

(c) TF Coordinators in VPUs

(d) VPU based staff with a key role in TF process

(e) Central Units-based staff with a key role in TF process

(f) Managers who approve TF-related transactions or supervise staff with a key role in TF process.

1.6 Because of the differences in background, knowledge level, and immediate needs of these diverse groups, the program development is to be split in two phases:

Phase 1: Focus on training and testing staff from groups a, b, and c above.

Phase 2: Focus on developing training and testing for the remaining groups (d, e and f).

TLAP DESCRIPTION

1.7 TLAP is the first training program in the Bank that requires accreditation through examination. Because of the high-profile and high-stakes nature of TLAP, a Bank-wide effort was initiated to develop the learning and accreditation program. Prior to the Pilot, the following bodies were established to manage or facilitate the program development:

- A Project Steering Committee (PSC) was set up to guide the program development and to make decisions on operational issues.
- A Bank-wide Consultative Group (CG), comprising two representatives from each VPU, was established to (a) provide technical input on the accreditation framework, (b) serve as a sounding board about staff concerns about accreditation, and (c) serve as a feedback loop to keep the Bank staff informed about developments in the area of TF training. In addition, both PSC and CG helped promote the program and break the resistance to the accreditation concept in the Bank.
- A Project Management Team (PMT) was put together to manage program development by coordinating the work of professional teams and consultants working on various aspects of the program.
- A team of content writers and exam developers was organized to develop the learning content and exam questions based on the skills matrix that had been developed earlier by PMT.
- An Editorial Board served as the final clearing house for all learning content and exam questions. Its work was in turn reviewed by the Trust Fund Four Pillars, which gave the final approval before the material was programmed into the learning platform.
- A technical team of learning experts and computer and web specialists simultaneously worked on the selection of a learning platform, the

development of a pedagogical template for the learning program, and the development of a training management system (TMS).

- Names of staff who served on the above groups are given in the acknowledgement section of the report.

1.8 The first version of TLAP was tested in May, 2002 through a validation workshop. An evaluation of feedback from this workshop pointed out the need for a significant revision of the learning content and exam questions, as well as the design template for the learning program. Recommendations from this workshop were incorporated in the second version of the program, which is the subject of this report, was piloted in October, 2002. The decision to release the program, currently geared toward TTLs, for Bank-wide rollout in December, 2002 was based on the results of this Pilot and the findings from the formative evaluations during the Pilot.

1.9 TLAP for TTLs comprises a learning system of 17 modules, a self-assessment test, and an accreditation exam based on these learning modules. All components of the program are accessible online, since the potential participants are located worldwide. A WebCT-based e-learning mode has been chosen for the program because of its interactive capabilities and its capacity to provide immediate feedback.

1.10 The focus of the learning modules is on areas of knowledge and competencies key to managing and executing TFs effectively and efficiently. The curriculum incorporated in the learning modules is currently geared towards TTLs, but is expected to be expanded to include the knowledge and information needs of Task Team Assistants (TTAs) and TF Coordinators. While the learning modules summarize key areas of knowledge, they also provide links to more detailed documents governing the use of TFs. The study of each module is expected to take approximately 30-40 minutes to complete. (Annex A provides a list of the learning modules and their objectives.)

1.11 The accreditation exam (the Exam) contains 48 graded questions/items. The Exam questions are closely linked to the content of the learning modules, but they do not require competency or knowledge based on the resource links provided within the modules. The first module, termed Module 0, is informational and the Exam does not contain items pertaining to this module. There are three items per module on the Exam for each of the remaining 16 modules. The Exam is designed to be completed within one hour. Examinees however, are permitted two hours to take the exam, and access to the learning modules and other materials is permitted during the Exam[2]. In addition, to help TTLs assess their knowledge level in TF management before they take the training, a voluntary self-assessment test (the Test) comparable in style and content to the accreditation exam, is also provided. Participants can use the recommendations from their self-assessment test results to guide their study of the learning modules.

[2] The Exam time allowed in the Pilot was adjusted to two and a half hours to allow participants to write their comments on the Pilot program and on the Exam.

TLAP PILOT ORGANIZATION

1.12 TLAP was piloted with 40 participants between October 2 and 21. During this period, the Pilot participants took the self-assessment test, went through the learning modules on their own, and then took the accreditation exam. Participants were advised to study at the pace of two modules per day, and were asked to provide their feedback on each of the modules they studied. Two feedback sessions, one mid-term review and one end-of-pilot focus group, were held to elicit further comments. Online navigation and hotline supports were included in the Pilot. (See Annex B for a detailed layout of the pilot process.)

1.13 The Pilot used four different forms for the accreditation exam.[3] These forms were created on the basis of 96 available items developed and rigorously revised by the content experts and educational specialists coordinated by WBIEG. Because of the small item pools in each module (each module had only 6 items ready when the Pilot was launched), the questions in these forms are not mutually exclusive. A passing score of 77 percent correct (or 37 out of 48 questions) was initially set by the TLAP Exam Board in an August meeting and later confirmed in the second meeting held in October. (See more discussion in Chapter 5.)

1.14 Forty Bank staff volunteered to participate in the Pilot, after invitations were sent to all TTLs on the Letter of Representation (LOR) list (1547 TTLs) and all CG members (53 members). Of the 40 participants, all took the self-assessment test, 36 completed the course, and 34 took the pilot version of the accreditation exam. Since the Pilot participants were all volunteers, it is not known to what extent these volunteers are representative of the entire TTL population that is expected to take the accreditation exam before July 2003. It is possible that the Pilot volunteers were more motivated and knowledgeable individuals, and therefore willing to participate in the Pilot. As shown in Table 1.1, 25 of them were TTLs. The breakdown on other demographics is provided in Chapter 5.

Table 1.1 Compositions of Pilot Participants

Job Category	Number of Participants
TTLs	25
ACS	2
Central Units	4
Networks	4
OVPs	5
Total	**40**

[3] Instead of using a fixed-form approach, the accreditation exam in the rollout will be formed by randomly drawing three items from each of the learning modules. See Chapter 5 for discussion and comparison of the randomization approach and fixed-form approach.

2. EVALUATION OBJECTIVES AND METHODOLOGY

EVALUATION OBJECTIVES

2.1 The objectives of this formative evaluation are to:

(a) Assess the quality of the design and implementation of the key components of the program. Specifically, it is to:

- Assess the quality of each of the 16 learning modules that comprise the learning program.
- Assess the quality of the accreditation exam and the self-assessment test (through the Pilot participant feedback as well as analyses of the exam and test result data).
- Assess the functioning of the technology aspects of the program (i.e., ease of navigability, access to information, speed in using the learning modules and attempting the test and the exam).

(b) Provide feedback to program designers for program improvement prior to the full rollout.

(c) Provide a report to the Learning Board on information about the progress of the program.

EVALUATION QUESTIONS

2.2 The following four questions guided the evaluation, and they will be addressed in detail in the subsequent chapters of this report. One chapter will be dedicated to each question:

(a) What is the quality of the content, design, and implementation of each of the learning modules?

(b) What is the quality of the design and implementation of the self-assessment test and the accreditation exam?

(c) What are the learning gains and passing rates of the Pilot participants?

(d) What is the overall quality of the Pilot?

EVALUATION METHODOLOGY AND DATA COLLECTION

2.3 The evaluation methodology consists of using the following strategies to answer the evaluation questions:

- Self-administered surveys and checklists.
- Collections of comments and focus group results.

- Self-assessment test and pilot accreditation exam data.
- Meeting with Exam Board, content experts, and education specialists to set or modify exam standards.

2.4 Conducted concurrently with the Pilot, a variety of data collection approaches were used in this formative evaluation, including:

- An end-of-module Level-1 checklist for each of the 16 learning modules: This end of module checklist queries participants on a variety of characteristics pertaining to the content and usability of the online learning modules.
- An end-of-pilot Level-1 questionnaire: This questionnaire solicits participants' overall rating and comments on the Pilot.
- A self-assessment test and an accreditation exam[4]: They serve as instruments for evaluating exam and item quality and learning gain, as well as being part of the Pilot.
- Level-1 surveys regarding the self-assessment test and the accreditation exam: These instruments assess participant opinions on the extent to which the exam is aligned with the learning modules, the difficulty and presentation of the items, and the usefulness of the modules and resources in the exam situation.
- An end-of-pilot focus group session: The focus group asks both general questions and specific questions about the learning modules, e-learning technology, the self-assessment test, and the accreditation exam.
- Informal observations of the pilot in-class sessions.

2.5 Table 2.1 links the evaluation questions with the data collection instruments described above.[5] Sample or copies of the end-of-module Level-1 checklist instrument, the end-of-pilot Level-1 questionnaire, Level-1 surveys regarding the self-assessment test and the accreditation exam, and questions used in the end-of-pilot focus group session are presented in Annexes C, D, E, F and G respectively.

Table 2.1 Evaluation Questions and Data Collection Instruments

Evaluation Questions	Data Collection Instrument
1. What is the quality of the content, design, and implementation of each of the learning modules?	
Achievement of objectives Logic/Sequencing of the content Relevance of the content Clarity of the of the content Design and ease-of-use of the online system	❑ End-of-module Level-1 checklist ❑ End-of-pilot Level-1 questionnaire ❑ End-of-pilot focus group ❑ Informal observations during in-class pilot sessions

[4] See further explanation in Chapter 1 about the structure of the accreditation exam and the self-assessment test.

[5] A more detailed description of the evaluation methodology was included in the *Formative Evaluation Concept Notes* listed in Annex I.

Table 2.1 (continued) Evaluation Questions and Data Collection Instruments

Evaluation Questions	Data Collection Instrument
2. What is the quality of the design and implementation of the self-assessment test and the examinations?	
Alignment with content Level of difficulty Online presentation Exam question quality	❑ End-of-test and exam questionnaires ❑ End-of-pilot focus group ❑ Item analysis of the self-assessment test and exam results
3. What are the learning gains and passing rates of the pilot participants?	❑ Results from the self-assessment test and the accreditation exam, after completing test and exam item analysis.
4. What is the overall quality of the Pilot?	
Learning Modules Overall achievement of the learning objectives Overall usefulness of the content of the modules Overall relevance of the content of the modules Overall quality of the presentation of the modules **Knowledge Gain/Applicability** Self-assessment of gain in knowledge of TF management Percent information provided in the learning modules applicable to the respondents' job as a TF manager **Quality of the Pilot** Support/facilitation provided Pacing of the pilot (e.g., overall timeframe) Information provided Pluses and minuses of the online, self-paced system Overall satisfaction with the Pilot	❑ End-of-pilot Level-1 questionnaire ❑ Focus group at the end of study session

3. LEARNING MODULES' QUALITY AND IMPLEMENTATION

3.1 The quality of the content, design, and implementation of the learning modules can be evaluated by considering their content characteristics and usability characteristics. These characteristics were rated and commented on by the Pilot participants in the end-of-module Level-1 checklist questionnaires. These end-of-module checklists were also used as diagnostic tools during the Pilot to enable the content (module) developers and the web-course designers to identify issues and their fixes.

CONTENT CHARACTERISTICS

3.2 The content characteristics included in the checklist are:

- Achievement of objectives (increased understanding)
- Logical sequencing
- Clarity of presentation
- Relevance of information

3.3 Table 3.1 on the next page provides a summary report card of content and usability related characteristics, as well as an overall quality assessment of each of the modules. This is based on the seven to 17 respondents who filled out the checklists (response rates from 18 to 43 percent). The second to fifth columns in this table represent the percentage of respondents who answered "*yes*" to the checklist content characteristics. The ratings in the column of "*Achievement of Objectives*" are the average of the ratings when there are multiple learning objectives contained in these modules.

3.4 The content characteristics section gives an overview of each module's strengths and weaknesses. As can be seen from the table, *increased understanding (achievement of objectives), logical sequence, clarity of presentation, and relevance of information* all received fairly high ratings on most of the modules. The highest average rating for a content characteristic was on *relevance of information* (96 percent), while the lowest average rating, on *clarity of presentation*, still reached 87 percent. In particular, Module 8 received 100 percent affirmative rating on all the four characteristics. Module 15, on the other hand, had the lowest ratings on three of the characteristics: *increased understanding, logical sequence, and clarity of presentation*.

Table 3.1 Module Level-1 Checklist Report Card

Module	Content Characteristics (% indicating "yes")				Usability Characteristics (% indicating "easy")				Overall Assessment (% of 4 or 5)
	Achievement of Objectives	Logical Sequence	Clarity of Presentation	Relevance of Information	Menu Navigation	Backward-Forward Navigation	Hotlinks Use	Dynamic Flowchart	
1	80	82	63	94	88	80	93	100	76
2	92	88	87	100	93	85	93	100	87
3	93	100	100	87	100	92	86	100	79
4	93	92	77	71	93	93	86	93	58
5	100	100	92	92	100	100	82	100	100
6	85	91	92	100	100	100	91	100	83
7	92	92	91	91	100	100	90	100	70
8	100	100	100	100	92	92	92	100	73
9	100	90	100	100	67	63	67	78	64
10	94	86	100	100	83	83	83	83	71
11	96	83	100	100	86	86	86	86	57
12	81	75	86	100	86	86	86	86	57
13	100	100	71	100	86	86	83	71	57
14	96	89	75	100	86	83	86	71	67
15	65	71	57	100	86	86	71	86	60
16	86	100	100	100	86	86	86	86	86
Mean	**91**	**90**	**87**	**96**	**90**	**88**	**85**	**90**	**72**

Note: The number of responses in each module ranges from 7 to 17, or response rates from 18 to 43%.

USABILITY CHARACTERISTICS

3.5 The usability characteristics of the e-learning mode included in the checklist are:

- Menu navigation
- Backwards-forwards menu navigation
- Hotlinks use
- Dynamic flowchart

3.6 The usability characteristics section in Table 3.1 represents the percentage of respondents who answered "easy" for performing the characteristics of each of the modules. If we consider 60 percent as the acceptable level for these built-in e-learning features, all these modules beat this threshold.[6] But in general, these usability ratings were not as high as the ratings on content characteristics. In particular, Module 5, 6, and 7 received the most favorable ratings in three areas (*menu navigation, backward/forward navigation,* and *dynamic flowchart*), while Module 9 received the lowest rating on almost all of the four usability characteristics.

OVERALL ASSESSMENT OF MODULE QUALITY

3.7 The last column in Table 3.1 is the percentage of respondents who gave a "4" or "5," on a scale from 1 to 5 and with "3" being the average and "5" being the best rating, on the overall quality of each of the modules. The overall rating in general was consistent with the ratings on content and usability, but the trend was not very clear. The average rating on the overall quality is 72 percent, well below the 85 percent benchmark that is commonly used for typical Level-1 ratings on questions with ***overall*** implications in the Bank. However, this should be interpreted carefully because of the unique format, the e-learning mode, and the limited sample sizes available for the analysis.

PARTICIPANT COMMENTS

3.8 Additional information was analyzed on the participant comments from the Module Level-1 checklists and end-of-pilot questionnaire. Between one (for Module 16) and 11 (for Module 4) participants provided their comments on the module checklists, and 21 (out of 34) participants submitted the end-of-pilot questionnaire with their comments.[7] These comments cover the objectives/content, instructional design, technology, and time issues as well. Overall, the Pilot was considered to be very successful by the participants. For example, participants said that the modules were "very well organized," "easy and targeted learning," "excellently presented," "for the first time I have found out where to find information about TF," and that the "learning approach was innovative."

[6] Currently we don't have a good benchmark to judge these e-learning features, partly because e-learning technology has not been used widely or evaluated in the Bank's staff learning programs.

[7] The end-of-pilot Level-1 questionnaire was administered to participants immediately before taking the accreditation exam. Thirty-four participants took the accreditation exam.

3.9 However, a few participants identified some areas for improvement in content and technology:

(a) Objectives/Content

- Need for clarification (i.e., not completely exhaustive information, need for more information, objectives hardly discussed, presentation confusing) was identified for Modules 1, 2, 3, 4, 6, 7, 10, 11, 12, 15, 16.
- Information confusing for Modules 2, 7, 16.

(b) Instructional Design:

- Scenarios and Q&A sections need review because they are too complex in Modules 1, 3, 4, 5, 7, 9, 14.
- Some repetition between examples and scenarios in Modules 1, 2, 3.
- Candidates identified the need for more information through more pop-ups in the dynamic flow charts, and definition of terms, acronyms in Modules 1, 9, 10, and 14.

(c) Technology:

- Overall, the bulk of the comments from the participants concentrate on technological aspects (e.g., back arrow, small font, too wide screen, how to navigate, slow from home).
- The dynamic flow charts were mentioned as being confusing in Modules 6, 12, 13.

(d) Time:

- Participants mentioned "the need for more time to absorb all the information provided" as a very important issue.

E-LEARNING: ADDITIONAL COMMENTS FROM FEEDBACK SESSIONS

3.10 In addition to the comments from the checklists and end-of-pilot Level-1 questionnaire, two feedback sessions also covered the topic of participants' reactions to the e-learning. Comments raised in the sessions reinforced those collected from the module checklists. Among the comments made by the few participants who joined these sessions some were positive (e.g., "the modules were excellent," or "better than expected,"), but others were more of concerns or improvement suggestions, such as:

- "It would be nice to have a book or hard copy in addition to the on-line resources."
- "A larger font would be appreciated."
- "Web access can be a problem."
- Slow home modems can cause difficulties although there were few complaints about being unable to log in.
- Simplify language of scenarios.

- The following usability issues were suggested:
  - Thumbnails.
  - More visuals, checklists, charts, etc.
  - Navigational bar for questions.
  - Improved functions related to forward-backward navigation.
  - The "next lesson button."
  - Allowing for learner observations not to be erased after moving to the next screen.

CONCLUSIONS AND RECOMMENDATIONS

3.11 In general, the TLAP learning modules were of sufficient quality in their content, design, and implementation. On average, more than 85 percent of the respondents gave affirmative or favorable ratings on content characteristics and usability characteristics of the e-learning mode for each of the 16 modules. The main message extracted from the participants' comments indicates that the learning modules were well organized and presented, and that the e-learning approach was innovative and much better than expected.

3.12 The ratings on the overall quality of the individual modules were relatively low compared to the Bank's benchmark, but this may be explained by the pilot programs' unique content, the high stakes, and the e-learning mode, as well as the limitation of the small sample sizes available for conducting the analyses. In addition, the ratings on usability characteristics were lower than the ratings on content characteristics, but there are no established Bank standards on e-learning at this time. Nevertheless, this may be explained by the fact that the usability of online learning mode is still a new technology and to some extent, a matter of taste and habit. The issue of usability needs to be watched carefully during the rollout.

3.13 Some modules may require particular attention to determine why they were rated lower than other modules. These include Module 15 on content characteristics and Module 9 on usability characteristics. Additional modules requiring attention were Modules 4, 11, 12, and 13. These modules received relatively low ratings on overall quality, although their ratings on content and usability characteristics were not the lowest.

3.14 Improvements or modifications need to be made on the online presentation of the learning modules, such as navigation functions and size of text fonts in different settings, as was indicated by the participants. Alternative learning modalities to the online e-learning should be provided (when necessary or possible), such as hardcopy/manual or classroom instructions.

3.15 Modules or sub-content areas linked to the few specific comments pertaining to ambiguities and inaccuracies should be examined carefully to determine whether these were common pitfalls or individual misunderstandings. Module pitfalls should be corrected.

3.16 One precaution: it is not known how the opinions of those who gave their ratings or comments differ from the opinions of those who did not. It is presumed that the silent majority were more or less satisfied since they were encouraged and provided with ample opportunities to contribute their feedback. It is also not known how the opinions of those who gave their ratings might differ from the opinions of the population that will be required to become accredited in TF management. This is because the participant sample size is small and the representativeness of the sample to its larger population is unknown. As future participants will be more diversified, PMT should assess the support functions provided for the content, technology, and/or the Program-related issues, assuming that adequate learning support can compensate for some of the inefficiency in the learning system established so far.

3.17 The data and interpretation discussed above were provided to PMT in October 2002. PMT had used this report to prioritize their "fixes" for the modules prior to the rollout in December, 2002.

4. ACCREDITATION EXAM QUALITY

4.1 The quality of the design and implementation of the self-assessment test and the accreditation exam can be judged by examining the following:

- Validity of the Test or the Exam
- Reliability of the Test or the Exam
- Item quality
- Mechanisms/actions for quality improvement
- Method of exam assembly

4.2 Since the accreditation exam and the self-assessment test reflected the same learning objectives, cover the same knowledge domain, and were developed following the same rules and procedures, the discussion of one applies to the other. We will not distinguish between them unless otherwise specified in the following discussions.

VALIDITY OF THE EXAM

4.3 The validity of the accreditation exam refers the extent to which the exam results provide a representative or *valid* measure of examinee's competencies in the tasks and behaviors related to TF management. The most important type of validity for a mastery test like the accreditation exam is its content validity. The accreditation exam is believed to have a high level of content validity, as well as construct validity, for the reasons discussed below.

4.4 The learning modules were based on a well-defined set of learning objectives, which allowed the knowledge domain, on which the Exam was based, to be better specified and thus more measurable. As seen in Appendix A, each module covered a narrow, well-defined range of job skills that were deemed by subject matter experts to be relevant to TF management. After careful review of these objectives and the key knowledge and competencies, it was determined that each module should have three representative items included in the Exam.

4.5 The Exam items reflected the competencies and tasks that were relevant to the learning modules. Item writers and reviewers (subject matter experts and learning specialists) followed rigorous procedures in developing and reviewing items to ensure that they adhere to the learning objectives. The whole process lasted about seven months, and many meetings were held to scrutinize both the content and quality of the items. As a result, out of 349 initially developed items, only 156 items were judged to have sufficient validity and quality to be used in the self-assessment test (48) and the accreditation exam (96).

4.6 In the Exam Level-1 Survey, participants indicated that an average of 86 percent of the Exam items were based on information presented in the learning modules. In various comments, most participants perceived that the exam items did reflect module contents.

4.7 That the participants did well on the accreditation exam after they took the training modules also provided some evidence of (instructional or construct) validity. This is discussed in the section pertaining to learning gains in Chapter 5. A significantly high correlation (r=0.66) between the self-perceived and measured learning gains further supported this validity.

RELIABILITY OF THE EXAM

4.8 Test reliability refers to the consistency of the test scores; coefficient alpha was used as the reliability measure for the Exam. A reliability level of 0.83 (coefficient alpha) was observed for the self-assessment test and an average level of 0.56 (ranging from 0.37 to 0.78) for the four Forms used for the accreditation exam.[8] Given that this is a mastery test, these numbers indicate that the Exam's reliability was acceptable.[9] Details follow.

4.9 In general, a reliability measure of 0.80 or greater is considered to be reasonably good for a norm-referenced test. For a mastery test, reliability measures can be misleading because low score variability will result in unreliable correlations. This is the case for the accreditation exam—the standard deviation of it is 6.5 percentage points while that of the self assessment test is 14.5. Of the four versions of the Pilot exam the reliability ranged between 0.37 and 0.78. As one would expect, these values were lower than the reliability measure on the self-assessment test (an equivalent pre-test). These lower reliability measures should not be of great concern.

4.10 The accreditation exam items were constructed according to the same rigorous procedures used in the construction of items in the self-assessment test. Therefore, one would expect that if the accreditation exam were given before the training modules were completed, the reliability measure would be closer to the measure seen on the self-assessment test.

4.11 Sometimes reliability is measured by comparing the correlation between the scores on two different forms of a test. The self-assessment test and the accreditation exam were designed to function as two different forms of the same test. The correlation between the scores on these two exams was 0.44 (n=32, significant at α=0.05). This provides partial evidence for reliability. One would expect a significant correlation, yet reduced by the fact

[8] The equivalence of the four fixed-form accreditation exams used in the Pilot was empirically confirmed by comparing the mean scores (ranging from 87.9 to 91.9 percentage points) of the participants who were tested on these different forms. The F statistic is 0.79 and p-value 0.51. Therefore the equivalence assumption can't be rejected.

[9] Since coefficient alpha is a lower bound of the "true" reliability based on certain assumptions, we would believe the true reliability of the exam could be even higher.

that participants had completed the learning modules prior to taking the accreditation exam.[10]

ITEM QUALITY

4.12 The item quality of the self-assessment test and the accreditation exam can be judged by examining two sources of information: (a) statistical data and (b) participants' comments. Results from both categories are presented below. Since the sample sizes were low, particularly on the accreditation exam, these results should be viewed with caution.

Statistical Data

4.13 Three most conventional statistical measures of item quality were used to assess the quality of the exam items: item difficulty, item discrimination, and distracter analysis.

4.14 Item difficulty index is measured by the proportion of examinees who got a particular item correct; the lower this number, the higher the level of difficulty. In general, items with item difficulty indexes greater than or equal to 0.80 are considered to be "easy" items. In the mastery test situation, this index may not be very relevant because item difficulty indexes tend to be high when instruction has been effective. As can be seen in Table 4.1, this is clearly the case for the accreditation exam where 89 percent of items had low difficulty levels. In the self-assessment test, the percentage was 25, which indicated a good balance between difficult and easy items in the exam if participants had not taken the learning modules.

4.15 Item discrimination index is a measure of how well an item discriminates between those who mastered the material (those with high overall test scores) and those who have not mastered the material (those with low overall test scores). An item discrimination index of 0.10 or less indicates that an item may be problematic in terms of discriminating between masters and non-masters of the material. Seventeen and 23 percent of items in the self-assessment test and the accreditation exam had such low discrimination indexes.

4.16 Distracter analysis: One method of analyzing distracters is to identify the non-functioning distracters, which are defined in this evaluation as distracters that no examinee selected among items with difficulty index lower than 0.80. There were 27 percent of the items in the self-assessment test and eight percent of the items in the accreditation exam containing such non-functioning distracters. (For detailed item analysis results, see Annex H1 on the self-assessment test and Annex H2 on the accreditation exam.)

[10] People could argue that such a correlation may not be an appropriate measure of reliability because the examinees have completed the learning modules—thus the two forms have not been taken under the same conditions.

Table 4.1 Number of Potentially Problematic Items

	With low difficulty (% of total number of items)	With low discrimination (% of total number of items)	With non-functioning distracters (% of total number of items)
self-assessment test (48 Items)	25% (12/48)	17% (8/48)	27% (13/48)
accreditation exam (96 Items)	89% (85/96)	23% (22/96)	8% (8/96)

Participants' Comments

4.17 Only a few participants provided comments. The comments, therefore, should not necessarily be construed as representative of the majority of the participants. Comments fell into general categories pertaining to *clarity*, *relevance*, and *other problems*. Table 4.2 presents the details on the number of participants who provided comments for the self-assessment test and the accreditation exam, as well as the number of comments falling into each general category. The vast majority of the participants' comments pertained to the clarity of the items. It is important to note, however, that a majority of the participants also answered these items correctly.

Table 4.2 Participants' Comments, by Exam/Test and Type of Comment

	No. of individuals who provided comments	No. of items on which comments provided*	Number of comment in type of		
			Clarity	Relevance	Other
self-assessment test (48 Items)	6	23	19	11	6
Accreditation Exam (96 Items)	20	46	32	3	11

*Note: Comments on some items fell in multiple categories, thus numbers in this column does not necessarily equal the sum of comments of different types.

4.18 The following are some typical examples of participant comments:[11]

- Example of a comment regarding *clarity*: "Ambiguous – not all TFs have Grant Agreements."
- Examples of comments regarding *relevance*: "This is not the sort of information I need to know as a TTL" or "I can't (and don't need to) remember which numbers go with which OPs and BPs. You should specify the subject of OP/BP 14.40."
- Example of a comment regarding *other problems*: "Rules just changed. New email on this [topic]."

[11] An item-by-item summary of the comments on the self-assessment test and the accreditation exams is included in Annex I.

MECHANISMS/ACTIONS FOR QUALITY IMPROVEMENT

4.19 Examining item discrimination indexes and non-functioning distracters, as well as participants' comments, provides a means of identifying potentially problematic items. Based on this information, subject matter experts and educational specialists revised the items prone to confusion and made the non-effective distracters more plausible. Subject matter experts also emphasized that there are pervasive misconceptions and common errors being made in TF management. The pervasiveness of the misconceptions is one of the motivating factors in the development of the training program and it would not be surprising for items to appear problematic (statistically) when they are actually quite suitable items for the accreditation exam.

4.20 WBIEG took three actions toward post-Pilot fixes and decisions. The first was to convene a meeting with two subject experts, a senior TFO staff and an experienced TTL who provided many of the suggestions, to review and revise the items based on the comments and the statistical data. The meeting covered the following:

- Reviewed each question on which the candidates had provided comments.
- Modified each question for clarity.
- Identified items for which fixes were needed in the modules.
- Identified items that were trivial but could stay in the Exam for now.
- Identified items that should be dropped.

4.21 Second, the Pilot statistical results were reviewed with an external expert in testing and measurement to identify further issues and make decisions regarding the quality of the self-assessment test and the accreditation exam. Third, a review meeting was held with the Exam Board to review the Exam results.[12] As result of these actions, four items were dropped and about 35 items were revised.

METHOD OF TEST ASSEMBLY FOR THE ACCREDITATION EXAM IN LATER ROLLOUT

4.22 Randomization vs. fixed-forms: In terms of test assembly WBIEG was faced with the decision of using fixed-forms versus producing a randomly-selected set of exam items for each examinee. The randomization approach was selected for future rollout of TLAP, after the following issues were carefully considered and debated.

4.23 Security issue: On the Pilot accreditation exam, WBIEG elected to use forms in order to obtain reasonably-balanced response data to run exam/item quality analysis. Although participants would sign an affidavit agreeing not to share the content of the accreditation exam, conceivably they could use the screen-print function or other means to print out a form of the Exam. This was not a concern in the Pilot because the participants were the first group of people having access to these items for the first time, but will become a major concern after rollout because of the high stakes of the accreditation requirement. Presenting each examinee with a randomly-selected set of the items reduces

[12] Final decisions regarding the passing score and testing time were also made in that meeting. Details regarding how the passing score was selected are presented in Chapter 5.

the chance that anyone would gain unfair access to a complete form of the accreditation exam.

4.24 Fairness issue: One concern of using randomization instead of fixed-forms is that an unlucky examinee might get all of the difficult items, thus making it more likely that he/she could "unfairly" fail the exam. However, a careful review of the following factors reveals that the probability that any given examinee would end up with a version of the most difficult items is very slim:

- Randomization will be done module by module.
- Within each of the 16 modules there are three strata of items covering similar objectives, and one item will be randomly selected from each of the strata.
- Each stratum contains at least than two items.
- Within a module, all the items have similar difficulty, as shown in the Pilot accreditation exam results. (see Appendix H2.)

4.25 Score range: An examinee needs to correctly answer 77 percent of the items in order to pass the Exam. The average score on the Pilot accreditation exam is 90 percentage points, and the standard deviation is 6.5 percentage points, for the sample size of 34. If the participants in the Pilot are representative of the general TTL population that will be taking the accreditation exam, a TTL with the average level of knowledge should have a more-than-95-percent chance to pass the Exam, assuming his/her performance on the Exam follows a normal distribution.

4.26 Retakes: An examinee has three chances to take the exam. As was mentioned, the probability of getting all of the difficult items on the first trial of the accreditation exam is slim. The probability of getting all of the hard items on three separate randomly generated versions of the Exam is therefore extremely low.

4.27 Ease in piloting new items: Randomization provides an incomparable means to field-test newly developed items with online testing technology. Each exam-taker can be easily presented a few randomly-selected piloting items without any knowledge of which items in their exams are graded or not. This makes the field-testing of new items more valid.

CONCLUSIONS AND RECOMMENDATIONS

4.28 The accreditation exam had supportive evidence indicating sufficient validity and reliability. Overall, item statistics showed that the items in the self-assessment test and the Pilot accreditation exam were of good quality. In addition, items were well-aligned with their respective modules. After having revised the potentially problematic items based on item statistics and participant comments, it is believed that both the underlying reliability and item quality should improve. Therefore, the accreditation exam can be assembled on these piloted items for the rollout.

4.29 After four items were dropped after revision, there are only 92 items upon which the accreditation exam in the immediate rollout will be based. This is a relatively low number for randomized exam construction, but it will be increased by adding new items which will be field-tested first in early rollout stages.

4.30 Many items in the Pilot accreditation exam were easy from a statistical point of view. This is not a big concern, given the mastery nature of the accreditation exam and the open-book testing in its administration. It is possible however, that the future accreditation examinees may not have the same level of competency as the Pilot participants, and therefore we may see an increased level of statistical difficulty for the same items.

4.31 WBIEG recommends the randomization approach be used to assemble the accreditation exam in the rollout. For each learning module, a separate item pool is established. Items within individual module pools are further divided into three content strata. The 48 items used for the accreditation exam will be formed by randomly selecting one item from each of the strata (3) of each of the modules (16).

4.32 New items need to be field-tested and added into the Exam item pool as soon as possible. After editorial and psychometric review, 90 new items are now ready for field-testing. WBIEG recommends these new items be incorporated into the accreditation exam in early stages of the rollout. It is further recommended that six new items be randomly embedded in the accreditation exam for this purpose. (TTL examinees will not be graded on these trial items.)

4.33 In order to have adequate response data for analyzing the quality of these new items soon, it is important to ensure that there are a large number of TTLs taking the Exam in early stages of the rollout. This concern needs to be reflected in the rollout strategy. WBIEG will monitor the quality of the exam item response data in the rollout on a weekly basis.

4.34 Once the response data and other comments are collected, WBIEG will use the same rigorous methodology, including editorial reviews, to examine and revise these new items. Once the modification is finalized, these items will be added to the exam item pool and be used in later stages of the rollout.

5. LEARNING GAINS AND PASSING RATES

5.1 This Chapter answers the question: "What are the learning gains and passing rates of the Pilot participants?" The learning gains analysis section provides an overview of the increase in participants' knowledge after studying the learning modules. The passing score standard section discusses how the passing score standard was established. Based on the standard, the passing rates section lists participant passing rates by different demographic categories.

LEARNING GAINS ANALYSIS

5.2 Learning gains are defined as the difference between the pre-test (the self-assessment test) and post-test (the accreditation exam) scores in percentage terms. Figure 5.1 shows the aggregated average pre-test score, post-test score and learning gain, as well as such average scores or gains for three self-identified groups of different competency levels *(novice, competent, expert).* The average pre-test score is 61 percent correct, and the average post-test score 90 percent correct. The learning gain for the whole group is 29 percentage points. The learning gains for all competency groups were also statistically significant, showing that the learning program had effectively increased participants' knowledge in TF management, as measured by the accreditation exam.

5.3 The score pattern for the self-identified competency groups in the pre-test, post-test and overall gain scores were exactly what is expected in this type of mastery situation:

- The *experts* on average had the highest initial scores (78 percent correct) and final scores (94 percent correct), but the lowest gain (16 percentage points).
- The *novices* on average had the lowest initial scores (49 percent correct) and final scores (88 percent correct), but demonstrated the greatest gains (38 percentage points).

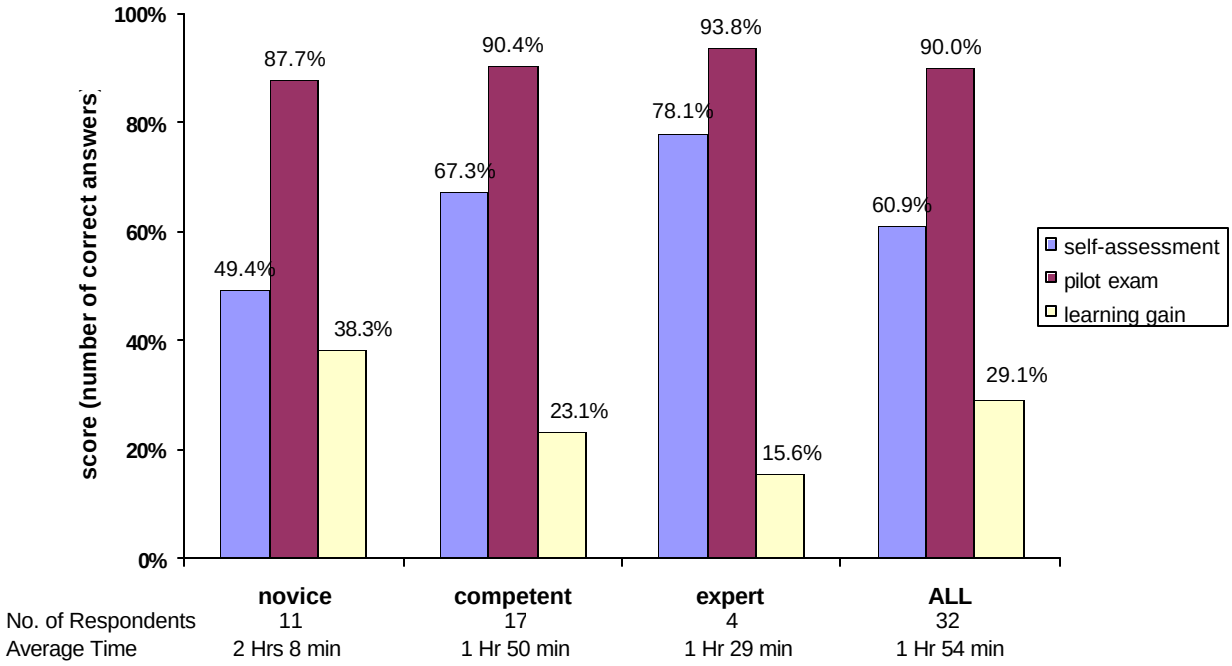
5.4 For the TTLs who were the very target group of TLAP, the average learning gain was 33 percentage points, with the average pre-test score of 56 percent correct and the average post-test score of 89 percent correct, somewhere between the average scores for the *novice* and the *competent* participants.

5.5 The *measured* average learning gain (29 percentage points) provides empirical confirmation of the *self-perceived* learning gain (27 percentage points), and the correlation between measured and self-perceived gains was 0.66.[13] The proximity of the two estimates

[13] After the four defective questions were eliminated from the accreditation exam, the adjusted average learning gain would become even closer to the self-perceived learning gain than the pre-adjusted learning gain was.

and the significant correlation support both the validity of the accreditation exam and the instructional effectiveness of the learning modules.

Figure 5.1 Pilot Participants' Learning Gain by Competency Groups



PASSING SCORE STANDARD

5.6 In order to determine the appropriate cutoff standard to categorize who has adequate mastery of TF management competencies and who does not, two score setting sessions were held with the Exam Board, the first on August 22 before the Pilot, and the second on October 30 after the Pilot.[14] Related to the score setting issue is the determination of the maximum length of time examinees are allowed to take the exam.

First Score-Setting Session

5.7 After reviewing the advantages and disadvantages of various standard-setting techniques, the Exam Board determined that a policy-based approach would be most appropriate for the accreditation exam because it would: (a) permit the use of a consistent passing score for all examination forms and (b) enable the Bank to notify examinees of their pass/fail status immediately upon completion of the examination. A policy-based passing score is one in which the standard for mastery is absolute and is established through the expert opinion of content experts and key stakeholder groups. However, to

[14] Before the Pilot, WBIEG established an Exam Board, comprising content experts from stakeholder units, including TFO, TFS, TQC, Accounting, Loans and Legal departments. Individuals representing these units reviewed the items and also participated in setting the passing score standard.

enhance the validity of the passing standard, it was decided to incorporate elements of a criterion-referenced approach, specifically the modified Angoff method.[15]

5.8 Before determining the policy-based passing score, the score-setting session participants were reminded of the ramifications of setting the passing score too high (i.e., failing examinees who in reality had satisfactorily mastered the learning modules) or too low (i.e., passing examinees who did not perform at a level of knowledge associated with basic mastery). Based on their expert opinion and the data gathered during the Angoff rating process, the Exam Board determined that a score of 77 percent correct of the exam questions constituted the basic mastery of the modules comprising the TLAP learning program.

Second Score-Setting Session

5.9 To formally establish a fair passing score on the accreditation exam and the maximum time allowed for completing the exam, the Exam Board reviewed the following information: (a) the distribution of individual scores and time used in the self-assessment test and the Pilot accreditation exam, (b) earlier score-setting procedures and previously set passing score, and (c) the basic competencies required for managing and executing TFs. Table 5.1 illustrates the score distribution of the 34 participants in the accreditation exam. This table was used in the session to show what the impact might be if the cutoff score were set at different levels.

Table 5.1 Pilot Accreditation Exam Score Distribution

Score Levels		Score Distribution		
Number of items correct	Percentage correct	Frequency	Group percent	Cumulative percent
32	67	1	2.9	2.9
38	79	1	2.9	5.9
39	81	2	5.9	11.8
40	83	2	5.9	17.6
41	85	1	2.9	20.6
42	88	3	8.8	29.4
43	90	5	14.7	44.1
44	92	7	20.6	64.7
45	94	5	14.7	79.4
46	96	4	11.8	91.2
47	98	2	5.9	97.1
48	100	1	2.9	100.0
Total		34	100.0	

[15] More information about the first score-setting session and methodology used can be found in the report, *Establishment of a Proposed Passing Score the World Bank Trust Fund Accreditation Examination*, prepared by Knapp & Associates International, Inc. Listed in Annex I and available upon request.

5.10 The Exam Board's decision was to recommend a passing score of 37 out of 48 questions (77 percent correct) and to allow two hours and 15 minutes to take the exam. (Examinees will be warned 15 minutes prior to the exam cutoff.)[16] This decision was based on the following two key observations:

- The Pilot data show no evidence to the contrary that the previously set passing score (37 out of 48 questions, or 77 percent correct) should be changed.
- Pilot participants who identified themselves as *novices* at managing TFs took an average of two hours and eight minutes to complete the exam, indicating that at least this amount of time should be allowed to take the exam.

PASSING RATES

5.11 Based on the standard (a score of 77 percent correct, or 37 out of 48 exam questions) set and confirmed in the above section, the accreditation exam passing rate for the whole group is 97 percent (33/34), only one person failed. Prior to working through the learning modules, however, their passing rate on the self-assessment test was only 20 percent.

5.12 Table 5.2 presents the passing rates of participants (as well as learning gains) in the Pilot by such demographic variables as grade, gender, competency level, and TTL assignment. Because passing rates in the accreditation exam were either 100 percent or all but one person for any of these groups, it is more interesting to look at these groups' passing rates based on their self-assessment test results. Because the sample sizes were small and the Pilot participant group were not a random sample from their population, the summaries discussed below or any inference from them should be read with great caution:

- Grade: Among these participants, Grade E and Grade G did better in both the self-assessment test and the accreditation exam.
- Gender: Male and female participants came to the training with almost the same level of knowledge in TF management, but a higher share of female participants passed the self-assessment test, and their overall gain was slightly greater.
- Competency level: The pattern in passing rates by competency level is what would be expected: the higher the competency level, the higher exam scores and the higher passing rates, but the lower learning gains.
- TTL assignment: Non-TTLs in this Pilot had better knowledge before and after the training, but lower learning gains, than TTLs. This may be because these non-TTLs were more related to TF management in their current work functions.

[16] There is ample documentation on the methodology and procedures used to determine the cutoff score and exam time in the two sessions.

Table 5.2 Passing Rates by Demographic Groups

Categories	n	Self-assessment Test		Accreditation Exam		Learning Gain [g]
		Mean Score	Passing Rate	Mean Score	Passing Rate	
Grade						
GA-GD	1	62.5%	0.0%	81.3%	100.0%	18.8
GE	4	66.7%	25.0%	92.2%	100.0%	25.5
GF	6	61.1%	0.0%	88.5%	83.3%	27.4
GG	18[a]	63.7%	35.3%	91.2%	100.0%	27.5
GH or Above	2	52.1%	0.0%	85.4%	100.0%	33.3
Gender						
Female	18	67.2%	27.8%	91.1%	100.0%	23.9
Male	14[b]	67.1%	15.4%	88.4%	92.9%	21.3
Competency Level						
Novice	11[c]	49.4%	0.0%	87.7%	90.9%	38.3
Competent	17	67.3%	23.5%	90.4%	100.0%	23.1
Expert	4	78.1%	75.0%	93.8%	100.0%	15.7
Work Assignment						
TTL	30[d]	56.2%	10.7%	88.9%	95.7%	32.7
Non-TTL	12[e]	72.0%	37.4%	93.1%	100.0%	21.1
Overall	42[f]	60.9%	20.0%	90.0%	97.1%	29.1

Note:

a. Of the 18 participants who identified themselves at the GG level, all took the accreditation exam, but one of them didn't take the self-assessment test.

b. Of the 14 male participants who answered the Exam, 13 of them took the self-assessment test.

c. Of the 11 participants who identified themselves as *novices*, all took the accreditation exam, but one of them didn't take the self-assessment test.

d. Thirty participants identified themselves as TTLs. Of these, 28 took the self-assessment test, 25 took the accreditation exam, while only 23 took both.

e. Twelve participants didn't specify themselves as TTLs. Of these, all took the self-assessment test, but only nine took the accreditation exam.

f. Forty-two people took either the self-assessment test or the accreditation exam. Of these, 40 took the self-assessment test and 34 took the accreditation exam, while only 32 took both.

g. Reported in percentage points, the learning gain is the difference between the average accreditation exam score and the self-assessment test score. For groups with the superscripts a through f, the number of participants who took the self-assessment test was different from the number of participants who took the accreditation exam, but the difference between the reported gain and the gain calculated using only matched participants was very small.

CONCLUSIONS AND RECOMMENDATIONS

5.13 Participants in the Pilot achieved high learning gains and passing rates after completing the learning modules, which indicates that the learning modules were effective learning tools to impart TF management knowledge to them and that there was good alignment between the modules and the exams. Specifically, 80 percent of the participants failed the self-assessment test, indicating that the exam was not easy—it tested knowledge that was not commonly known among people managing TFs. Ninety-seven percent of the

participants passed the accreditation exam, indicating that studying the modules did help them to gain knowledge and to pass the Exam.

5.14 Knowledge levels in TF management and passing rates before the training varied among different demographic groups. However, studying learning modules effectively reduced the gaps of both their knowledge levels and their passing rates between these demographic groups.

5.15 During and after the second score-setting session, the following issues were also identified as requiring follow-up:

- Provide the option of taking the exam in a location other than one's office (either in the library or in a designated room). This option should be available in Washington D.C. and in field offices.
- Encourage accreditation candidates to study the learning modules by showing them data on how individuals performed before and after studying the modules in the Pilot. These data could be incorporated in the information regarding TLAP like the communication strategies.
- Collect some Level-1 data after the accreditation exam (similar to data collected for the Pilot).
- Address exam technology glitches, such as those encountered by the Pilot participants: the slowness of the system response, unwarned early system-cutoff, and the problem with registration using another person's desktop.[17]

5.16 In addition, after rollout, data on learning gains and passing rates should continue to be monitored as more staff take the accreditation exam:

- Learning gains should be monitored. Areas of difficulty need to be identified and supplementary materials can be added to the modules, if necessary.
- The passing rates need to be periodically monitored, as well as the demographics pertaining to those who pass and those who fail.
- The passing score standard needs to be revisited in the context of the passing rates and the learning gains data, particularly when new items are added to exam item pools.
- The actual time examinees spend on the exam should be continually monitored. If the time limit becomes a barrier to passing for a significant number of examinees it should be reviewed. The time spent on the exam by examinees could be tracked by demographics to gain insight into any potential barriers to passing.

[17] During the Pilot Exam, for some reason, the online exam for one cohort was cut off a half-hour early without any prior notice. PMT made this up by offering an additional half-hour to those affected by the glitches. Paper-form exam for the same test was used in this additional session. In another instance, one participant used a computer that was not her own during the exam session, but no score was recorded under her name.

6. OVERALL QUALITY OF THE PILOT

6.1 The overall quality of the implementation of the Pilot can be evaluated by looking at the following: (a) overall quality of the learning modules, (b) overall quality of the Pilot, and (c) overall quality of the self-assessment test and the accreditation exam. Three Level-1 instruments were used to collect relevant feedback on these questions, i.e., one survey at the end of the self-assessment test, one survey at the end of the accreditation exam, and a stand-alone Level-1 questionnaire administered immediately before the accreditation exam. Because these questionnaires were administered in-person, their response rates were high, 95 percent (38/40), 91 percent (31/34), and 94 percent (32/34) respectively.

OVERALL QUALITY OF THE LEARNING MODULES

6.2 On overall quality indicators for the learning modules, participants' ratings were highly positive on the *overall achievement of objectives*, *overall usefulness*, and *overall quality*. Table 6.1 shows that the average ratings of the three indicators on a scale of 1-to-5 were 4.09, 4.10, and 4.09, and the percentages of ratings of "4" or "5" were 81, 87, and 78 respectively. All these numbers were very similar to the ratings received by other Bank staff training programs.

6.3 However, on *overall relevance* of the learning modules, the ratings were less positive, but above the mid-point of the scale. This above-midpoint relevance rating matches the module-level checklist relevance ratings (see Chapter 3), where on average 96 percent of the participants who provided feedback indicated that the individual modules were relevant to managing TFs.

Table 6.1 Ratings on TLAP Learning Modules

	Mean	% of 4 or 5	Std	Max	Min
Overall Quality Indicators (Scale: 1-5, with 1=very poor to 5=very good)					
Overall achievement of TLAP objectives	4.09	81	0.69	5	3
Overall usefulness	4.10	87	0.60	5	3
Overall relevance	3.87	68	0.81	5	2
Overall quality	4.09	78	0.82	5	2
Knowledge Gain and Its Applicability (Scale: in percentages)					
% info known before	51.1	N. A.	22.1	90	0
% info known now	78.1	N. A.	12.0	95	50
% info applicable to your TF management	74.3	N. A.	20.0	100	40

Note: N. A. = not applicable.

6.4 The participants themselves indicated an average of a 27 percentage point increase in their knowledge of the information presented in the learning modules. This self-

perceived learning gain is quite close to the average gain of 29 percentage points as measured by the accreditation exam, discussed in Chapter 5.

6.5 Participants reported that only 74 percent of the information was *directly applicable* to their management of TFs. TTL participants as a separate group reported that 72 percent of the information was *directly applicable* to their management of TFs, which may indicate that these non-TTLs had more contact with TF management in their work function and which may also explain why these non-TTL participants did better in the self-assessment test than TTL participants in the Pilot.

6.6 The relatively low relevance and applicability is not necessarily problematic for the following reasons:

- Lower ratings on applicability may reflect participant perceptions concerning areas that were omitted in the learning modules, but that some of them thought were important, such as the need to include information more related to Networks, WBI or other sectors of the Bank.[18]
- It may be difficult for any program to address the lack of homogeneity in participant activities. Discussions with PMT revealed that most TTLs deal with only a small set of TFs or issues related to them, and therefore it is not surprising that the Pilot participants thought only 74 percent of the information was relevant to them. However, the program is designed to equip people more broadly with the knowledge of how to manage TFs now and in the future. Even though a TTL may not currently be using a particular TF, she/he may do so in the future.
- The measure of *relevance* considered in this evaluation pertains only to the relevance as perceived by participants. It does not address the relevance of the material to problematic issues in TF management, which may be a higher priority to the Bank in terms of relevance.

OVERALL QUALITY OF THE PILOT

6.7 Participants gave highly favorable ratings on four of the five quality criteria about the Pilot: (i) *support/facilitation provided throughout the Pilot,* (ii) *information provided about the Pilot,* (iii) *overall quality of the Pilot,* and (iv) *overall satisfaction with the Pilot.* As shown in Table 6.2, the means for them were all above 4.19 on a scale of 1 to 5, with the rating on *support provided* reaching 4.69; 97 percent of the participants gave a rating of "4" or "5."

[18] This will be one of the issues to be addressed in non-mandatory learning modules developed later, according to the TLAP framework paper.

Table 6.2 Ratings on the Quality of the Pilot
(Scale 1-5, with 1=Very Poor to 5=Very good)

	Mean	% of 4 or 5	Standard Deviation	Max	Min
Support provided	4.69	97	0.54	5	3
Pacing	3.84	66	0.95	5	2
Information provided	4.31	91	0.64	5	3
Quality	4.19	84	0.69	5	3
Satisfaction	4.22	88	0.66	5	3

6.8 Ratings on *pacing of the Pilot* were slightly lower, although still above the mid-point. This indicates that some of participants may feel somewhat stressed by following the suggestion of studying two modules per day during the Pilot period. This should not be a problem when TTLs can individually pace their study in the rollout. In addition, providing hardcopies or manuals to participants might solve this problem to some extent.

THE SELF-ASSESSMENT TEST AND THE PILOT EXAM

6.9 Ratings indicate that the candidates perceived both the self-assessment test and the Pilot accreditation exam to be slightly difficult (below the midpoint level), and this perception didn't change after they studied the learning modules. (See Table 6.3.) In fact, the average scores in the self-assessment test (61 percent correct) and the accreditation exam (90 percent correct) were very different from each other; which indicates that participants didn't have an accurate estimate of their own knowledge and capabilities, although they may have been confident about passing.

6.10 The presentation of the questions/items in the self-assessment test and the accreditation exam were rated slightly above "OK." ("OK" is the mid-point level on the scale. See Annex E and F.) There was no difference between the ratings on items in the Test and on items in the Exam. This is not surprising because all these questions were developed following the same procedure and covering the same content areas, and because the presentation of them was both constrained by the same WebCT platform.

6.11 Information provided about the self-assessment test was rated slightly above "helpful," indicating that the information could be more helpful. All participants said they used the learning modules and/or related resources as references in the Pilot accreditation exam, and the modules and/or resources were rated as being slightly above "helpful" in completing the exam.

Table 6.3 Ratings in the Self-Assessment Test and the Accreditation Exam

	Mean	% of 4 or 5	Standard Deviation	Max	Min
L1 Questions in the Self-assessment Test					
Difficulty of the Test *(Scale: 1=very difficult to 5=very easy)*	2.3	0.0%	0.7	3	1
Presentation of the Test questions *(Scale:1=very poor to 5=excellent)*	3.3	30.8%	0.5	4	3
Helpfulness of information provided about the Test *(Scale: 1=not at all to 5=extremely helpful)*	3.2	30.0%	0.6	4	2
L1 Questions in Accreditation Exam					
% of the Exam based on information in modules *(Scale: in percentages)*	86.4%	N. A.	9.7%	100.0%	70.0%
Difficulty of the Exam *(Scale: 1=very difficult, to 5=very easy)*	2.4	0.0%	0.7	3	1
Presentation of the Exam questions *(Scale:1=very poor , to 5=excellent)*	3.3	38.7%	0.6	4	2
Use of learning modules during the Exam *(Scale: 1=yes, 0=no)*	1.0	N. A.	N. A.	1	0
Were learning modules helpful in completing the Exam *(Scale: 1=not at all, to 5=extremely helpful)*	3.5	58.1%	1.0	5	2

Note: N. A. = not applicable.

CONCLUSIONS AND RECOMMENDATIONS

6.12 The Pilot received favorable ratings from participants on *overall achievement of objectives, overall usefulness,* and *overall quality* of its learning modules. The percentages of ratings of "4" or "5" on an 1-to-5 scale were 81, 87, and 78 respectively, comparable to other staff learning programs in the Bank. On the Pilot overall, a majority of the participants gave ratings of "4" or "5" on *support provided* (97 percent), *information provided* (91 percent), *quality* (84 percent) and *participant satisfaction* (88 percent). All these indicate that the whole Pilot was well-prepared and well-executed.

6.13 Participants, however, were less favorable regarding two other aspects of the Pilot. Fewer participants gave high ratings ("4" or "5") on *relevance of the learning modules* (68 percent) and *pacing of the Pilot* (66 percent). In addition, participants on average judged only 74 percent of the information in the learning modules was *directly applicable* to their management of TFs.

6.14 The extraordinarily high rating on the support provided raises the concern of whether the high level of quality support can be sustained after rollout. The Pilot was well-organized and participants' study and examination were conducted in a controlled environment, where participants' problems and comments can be solved or answered immediately. The high level of support during the Pilot might explain the high levels of learning gain, passing rates, and the various ratings as well. This should be kept in mind during the rollout because it will not have the same level of support when a much larger number of participants/TTLs study the modules and take their exams simultaneously and continuously.

6.15 The accreditation exam was perceived by the Pilot participants as relatively difficult, even though almost all of them achieved passing or higher scores. All participants used the learning modules and/or other resources during their exams, but their rating of the *usefulness of the learning modules for the exam* was just slightly above "helpful". In addition, their rating on the *presentation of the exam questions* was only slightly better than "OK," a phenomenon that may be attributable to the limitations of the WebCT technology. These mixed perceptions and results, however, may be appropriate for such a high-stake accreditation exam. Although the exam was not perceived easy, by studying the learning modules carefully, it is possible to pass and show significant learning gains.

6.16 There are a few areas where further improvement can be still made, including pacing, presentation of the self-assessment test and the accreditation exam, as well as the clarity of the content and the exam questions. Among others, WBIEG has made the following recommendations to PMT:

- Review candidates' comments in the module checklists to make expert judgments regarding essential changes, including relevance and clarity.
- Review how the self-assessment test and the accreditation exam may be better presented, given the constraints of the WebCT technology.
- Review comments on the self-assessment test and the accreditation exam to identify and fix misalignment either in the learning modules or in these testing questions (already accomplished for the Test and Exam questions).

6.17 Although PMT has taken actions to compile information from different sources and fix problems in different priorities, it is possible that for some of questions, there are no easy or quick "fixes" before the rollout, given the complexity and time demand of the problems. For example, the relatively low relevance will not be an easy fix, as discussed in the first section of this Chapter. Relevance is even less of a concern because it did not appear to affect the motivation of the participants to learn the material, nor did it appear to have an effect on the participants' overall opinion of the program (although it should be noted that these participants might not be representative of the TTL population).

6.18 It is also possible that old problems resolved before the rollout could recur later. For example, when the item bank gets updated, content of learning modules changed, or functions of technology expanded, it is hard to predict how these changes will affect participants' learning and accreditation performance, as well as their ratings on the program. Therefore continued vigilance should be maintained after rollout. WBIEG has developed a set of new survey questions that will be used in combination with the self-assessment test and the accreditation exam, as well as an independent Level-1 Questionnaire that all participants who take the accreditation exam will be asked to fill out. These instruments and information collected through them will also lay groundwork for a later impact evaluation of TLAP.

7. SUMMARY AND CONCLUSIONS

7.1 This Chapter gives a concise review of the formative evaluation issues covered in the earlier chapters. The summary section recaps the evaluation findings generated by the four original evaluation questions. The conclusions section makes the recommendation that the Program can move forward with a formal rollout, but it should proceed with caution due to the high-stakes nature of the program. This is further addressed in the last section on implications and recommendations.

SUMMARY OF EVALUATION FINDINGS

Question 1. What is the quality of the content, design and implementation of each of the learning modules?

7.2 In general, the learning modules were of sufficient quality in their content, design and implementation. On average, more than 85 percent of the respondents gave affirmative or favorable ratings on such questions as content characteristics (*achievement of objectives, logical sequence, clarity of presentation, and relevance of information*) and usability characteristics of the e-learning mode (*menu navigation, backward/forward navigation, hotlinks use, and dynamic flowchart*) of the 16 modules. The main message extracted from the participants' comments indicates that the learning modules were well-organized and presented and that the e-learning approach was innovative and much better than expected.

7.3 The ratings on the overall quality of the individual modules were relatively low compared to the Bank's benchmark, but this may be explained by the Pilot program's unique content, the high stakes, and the e-learning mode, as well as the limitation of the small sample sizes available for conducting the analyses—only a few people provided their feedback. Using the Bank benchmark is therefore not applicable. The evaluation did not reveal specific reasons underlying the discrepancy between the ratings on specific characteristics and the overall ratings on the individual modules.

7.4 On average, the ratings on the e-learning usability characteristics were lower than the ratings on content characteristics. This may be explained by the fact that the usability of online learning mode is still a new technology and to some extent, a matter of taste and habit. There are no established comparable Bank benchmarks available.

7.5 The ratings on these characteristics vary among different modules. Some modules may require particular attention to determine why they received lower ratings than other modules. These include Module 15 on content characteristics and Module 9 on usability characteristics. Additional modules requiring attention include Modules 4, 11, 12, and 13. These modules received relatively low ratings on overall quality, although their respective ratings on content and usability characteristics were not the lowest.

7.6 Other than the positive feedback, the four most important comments coming from the Pilot participants were: (a) some of the information in the learning modules was incorrect; (b) some of the answers to the questions in the learning scenarios were ambiguous; (c) supplemental modalities to online learning should be provided; and (d) a level of support should be maintained for future participants. In response, PMT reviewed the related modules and questions and made appropriate revisions. To address the exclusive usage of e-learning as the learning platform, PMT has enabled the printing of the learning modules so that future participants can have hard copies to study. An email hotline has been established to provide support for future participants. In addition, PMT will watch the situation closely to see what modifications might be needed in the rollout.

Question 2. What is the quality of the design and implementation of the self-assessment test and the accreditation examination?

7.7 The quality of the self-assessment test and the accreditation exam is judged by validity, reliability, and item quality. Multiple sources of evidence support the validity of the Exam and the Test, such as well-defined learning objectives and alignment of exam items with the learning objectives. The reliability coefficient of 0.83 for the Test fell into the standard acceptable range.

7.8 The items on the self-assessment test and the accreditation exam were well aligned with the learning objectives and module contents, and had generally good item statistics. Potentially problematic items were reviewed and modified based on the item statistics and comments from participants. The final item yield was high—only 4 items (out of 144 exam/test items) were suggested to be excluded from future use. The few items considered to cover trivial content shall be replaced with newly developed items. Given the exam item pool will be expanded over time, the newly developed items need to be similarly examined before adding them to the exam item pool.

7.9 Based on this evidence and the revisions made on the items, the accreditation exam in the rollout can be assembled from the items used in the Pilot. Because of these improvements, the accreditation exams in the rollout should have even better validity and reliability than the exam in the Pilot.

7.10 The Test and Exam data, as well as the general context of the accreditation examination, provided support for the use of randomized item selection in the test construction process in the rollout. Randomization addresses the concern of security better than the fixed-form approach, allows an adequate level of fairness, and provides a convenient venue for field-testing newly developed items as well. PMT has adopted the randomization approach in the rollout.

Question 3. What are the learning gains and passing rates of the Pilot participants?

7.11 The learning modules appeared to be an effective learning tool as evidenced by high measured learning gains and high self-perceived learning gains, as well as the substantial increase in the passing rate after completing the learning modules. After studying the modules, the Pilot participants on average achieved a highly significant

learning gain of 29 percentage points. The passing rate increased from 20 percent on the pre-test (the self-assessment test) to 97 percent on the post-test (the accreditation exam). Participants' measured learning gains matched their self-perceived learning gains very well, providing further evidence of the alignment of modules with the exams.

7.12 The passing score standard for the accreditation exam was initially set and later confirmed by the Exam Board at a level of 77 percent correct, equivalent to answering 37 questions correctly out of 48 exam questions. This was a policy-based decision, combined with subject matter experts' judgments on what constitutes a minimum competency to manage TFs in the Bank. According to this standard, all but one participant (33 out of 34) passed the accreditation exam in the Pilot and will be accredited by the TFO through formal procedures.

7.13 The Pilot Participants' knowledge level in TF management and their passing rates before the training varied among such demographic groups as gender, grade, TTL work assignment, and self-identified competency level in TF management *(expert, competent* and *novice)*. However, studying learning modules effectively reduced the gaps of both the knowledge levels and the passing rates among these demographic groups. Specifically,

- The self-identified *expert* group on average had the highest initial score (78 percent correct) and final score (94 percent correct), but the lowest learning gain (16 percentage points), while the self-identified *novice* group on average had the lowest initial score (49 percent correct) and final score (88 percent correct), but demonstrated the greatest learning gain (38 percentage points).
- A similar pattern was observed between non-TTLs and TTLs. The non-TTLs on average had a higher initial score (72 percent correct) and a higher final score (93 percent correct), but a lower learning gain (21 percentage points). The TTLs on average had a lower initial score (56 percent correct) and a lower final score (89 percent correct), but demonstrated a higher learning gain (33 percentage points).
- No statistically significant differences in passing rates or scores by gender and grade groups were found.
- It is hard to infer from these observations how their broader population will fare in the rollout because of small sample sizes and more importantly, the unknown representativeness of these groups in the Pilot to their respective populations. Nevertheless, a clear policy implication is that future candidates must be encouraged to study the learning modules before taking the accreditation exam.

Question 4. What is the overall quality of the Pilot?

7.14 The Pilot received favorable ratings from participants on *overall achievement of objectives, overall usefulness, and overall quality* of its learning modules. The percentages of ratings of "4" or "5" on an 1-to-5 scale were 81, 87, and 78 respectively, comparable to other staff learning programs in the Bank. On the Pilot overall, a majority of the participants gave ratings of "4" or "5" on *support provided* (97 percent), *information*

provided (91 percent), *quality* (84 percent), and *participant satisfaction* (88 percent). All these indicate that the whole Pilot was well-prepared and well-executed.

7.15 Participants gave less favorable ratings to two other aspects of the Pilot. Fewer participants gave high ratings ("4" or "5") on *relevance of the learning modules* (68 percent) and *pacing of the Pilot* (66 percent). In addition, participants on average judged only 74 percent of the information in the learning modules was *directly applicable* to their management of TFs.

7.16 The relatively low ratings of the TLAP Pilot on relevance and applicability may not improve during the rollout. There are several reasons for this, the main one being that TTLs may not feel that they need to know all aspects of the TF management if they are only involved or interested in certain types of TFs. In other words, relevance pertaining to individual participants may not be the same as the relevance to TF management in general. TLAP is designed to equip people more broadly with the knowledge of how to manage TFs now and in the future. It is important to note that even though a TTL may not currently be using a particular TF, he/she may do so in the future.

7.17 The accreditation exam was perceived by the Pilot participants as relatively difficult, even though almost all of them achieved passing or higher scores. All participants used the learning modules and/or other resources during their exams, but their rating of the usefulness of the learning modules for the exam was just slightly above "helpful." In addition, participant rating on the presentation of the exam questions was only slightly better than "OK," a phenomenon that may be attributable to the limitations of the WebCT technology. These mixed perceptions and results, however, may be appropriate for such a high-stake accreditation exam. Although the exam was not perceived easy, by studying the learning modules carefully, it is possible to pass and show significant learning gains.

CONCLUSION

7.18 Overall, the Pilot results indicate that the learning modules were of good quality in their content, design and implementation; the exams were well aligned with the learning objectives and were of good quality; and the passing score standard was reasonably set. Moreover, the participants were self-motivated learners and were satisfied with the quality of the Pilot and the support provided, and mostly importantly, they could pass the accreditation exam after studying the learning modules.

7.19 Because of the formative nature of this evaluation and because WBIEG is also part of the PMT responsible for developing the accreditation exam, all the findings and recommendations made in the formative process have been conveyed in a timely manner to PMT and other TLAP stakeholders. These findings and suggestions have been considered by PMT for incorporation in the preparation of the formal rollout.

7.20 It should be noted that not all of the feedback received and the actions taken by PMT were evaluated by this report. In addition to what was discussed in this report, PMT itself has received comments and suggestions from different channels that may or may not be known to this evaluation team and therefore were not covered here, particularly on the

content and e-learning technology/presentation aspects. For example, the Financial Management Board and the Conflict of Interest Office provided their comments on the modules' content. Some participants provided the PMT with their comments through the support hotline. PMT adopted three "filters" that prioritize the suggested needs for treatment: factual inaccuracies—must be fixed; navigation/technical issues—to be fixed when time permits; and suggestions for improvements (nice-to-have's)—to be fixed when time permits. These feedback and actions all played an important role in preparing and improving TLAP for the rollout.

7.21 Based on the results from this evaluation report and observations or actions made by PMT, it is concluded that the Program can now move forward with a formal rollout for accrediting the Bank's TTLs. However, the preparation of the rollout and the full implementation should proceed with caution because there are still some issues or problems that need to be addressed. Some cautionary notes and recommendations for the final rollout are outlined in the next section.

ISSUES AND RECOMMENDATIONS FOR THE FINAL ROLLOUT

7.22 The section mainly discusses issues related to learning system support and the accreditation exam quality monitoring. They are organized around the following:

- Rollout strategy
- Exam administration
- Technology issues
- Support
- Generalizability of the Pilot results
- Future evaluation and monitoring needs

Rollout Strategy

7.23 During the writing of the report, PMT outlined a rollout strategy. The objective of it is to get all approximately 1600 TTLs to take the accreditation exam before June 30, 2003, and to avoid concentration of participants in the last months so that the system does not become overtaxed by too many participants waiting until the deadline to take the exam.[19] According to PMT, the rollout strategy encompasses the following activities:

- Communicate that accreditation is mandatory, and the deadline is June 30, 2003.
- Motivate staff to study the modules and take the exam through the use of emails, a Website (TLAP homepage), and other advertising.
- Designate TLAP Coordinators nominated in every VPU to assist with brown bag introduction sessions and "hands on sessions" for those needing more support.

[19] Communications from the PMT.

- Ask VPUs to provide a schedule and plan outlining how many TTLs will take the exam and what the plan is for all to be accredited by the deadline; in addition, ask that progress be documented.
- Create an environment of competition. Progress of VPUs in accreditation will be posted on the TLAP website.
- Adopt additional strategies for Regions to facilitate introduction of the accreditation program.

Exam Administration

7.24 PMT adopted a cohort approach that each month requires specific cohorts to be encouraged to work through the modules and take the accreditation exam. This will alleviate overtaxing the system by TTLs in the last few months. It also ensures that enough response data on the exam in the first few months be received for evaluating and improving quality of the exam items, as well as other exam-related issues.

7.25 A small risk of the randomized exam construction is that a given examinee might get a *slightly* easier or harder exam than another examinee, but this risk can be reduced when adjustment is made based on new item difficulty data.

7.26 A minor concern regarding piloting new items on the real exam is that an examinee might spend too much time on a piloting item and neglect items that are graded. There should be, however, no cause for complaint in this regard because the allowable exam time already took this into account.

7.27 The current exam item pool, with only 92 items, needs to be expanded soon. This will depend on how quickly response data are generated for running the item analysis and how quickly the item editorial process and technology support will be.

7.28 Because TTLs could be interrupted during their exams by various emergencies or computer/internet-connection problems during their exams, the option should be provided of taking the exam in a place other than one's office (either in the library or in a designated room). This option should be available both in Washington D.C. and in field offices. Additionally, there should be provisions for TTLs to have a pencil and paper version of the accreditation exam at proctored centers.

Technology Issues

7.29 Collection of comments and suggestions on online presentation of learning modules, resource links and navigation issues should continue. The hotline support system can be used to collect this information.

7.30 Learning modules or the Exam's online access capacity should be evaluated in the rollout. In the Pilot, there were less than 40 participants studying the modules and taking the exam at the same time. The capacity of the system to handle large numbers of participants is unknown. For example, if more than 200 TTLs taking the Exam can slow the system down, the exam time may need to be adjusted.

7.31 Emergency technology support should be provided particularly during the exam time, for example, when a computer crashes or the internet connection is down. Such support should be available at least during regular working hours, but could be a problem when a TTL takes the Exam during non-working hours.

Support

7.32 TTL participants' performance on the exam (and possibly their ratings on the program) in the rollout may depend, in part, on the extent to which the support provided mirrors the level of support provided in the Pilot. PMT has decided to provide two kinds of support after the rollout: a hotline for navigation issues and email correspondence for content-related issues. Support requests from future participants should be monitored to determine if any changes of the planned support system need to be made.

Generalizability of the Pilot results

7.33 In addition to small sample size, the Pilot participants were volunteers and may have personal characteristics that are different from the larger pool of TTLs who will be taking the accreditation exam in the rollout. Moreover, the Pilot was conducted under a controlled environment where extensive support and immediate feedback were provided to help participants walk through the learning and testing steps. Therefore it is not appropriate to generalize the results of the Pilot and opinions of this group onto the Bank-wide TTL population. A few of these issues are discussed below.

- High learning gains and passing rates: If the Pilot participants were more motivated than the average accreditation program participant, the learning gains and passing rates in the Pilot may be overstated and therefore not generalizable to the whole TTL population.
- Passing score standard: If the Pilot participants were more motivated than the average TTL accreditation program participant, there may be a higher rate of failure in the rollout. As more exam data are available, the cutoff score standard should be revisited.
- Time allowed for exam: Related to the passing score standard, when more exam data are available, the time allowed for the exam should be reviewed. In addition, time allowed should also be revisited when there is system-wide slowness or common technology problems in accessing the online exams.
- Protocol for issues related to non-accreditation: If the failure rate after rollout is higher than the failure rate of the Pilot, a protocol for handling issues related to non-accreditation needs to be developed.
- Issues that may stay in the rollout: One example is the low relevance rating of learning modules in the Pilot, which is not abnormal given different implications under these ratings.

Future Evaluation and Monitoring Needs

7.34 Although the Pilot is of sufficient quality to proceed with rollout, continuous monitoring and evaluation activities are required to ensure that the accreditation program maintains an adequate level of quality in the rollout. Some of these evaluation activities are:

- There should be a periodic review of the learning curriculum. A mechanism should be established regarding updating the content of the modules when new concerns need to be addressed or new policies in TF management are developed.
- New items should be developed to reflect any content changes in the curriculum or to replenish the exam item pool due to the loss of items that were proved to be malfunctioning. Development of new items should follow the same rigorous procedures that were used to develop the original items.
- The exam item pool upon which the accreditation exam for the immediate rollout is based should be expanded with items that have been field-tested and reviewed by multiple subject matter experts. At the same time, module-level or stratum-level item pools should be adjusted to make the randomized exams fairer after more item difficulty data are collected.
- The exam time and passing score standard should be reviewed again, as more exam data are available to reflect the context that is different from where the Pilot was conducted and such standards were set.
- Although there are no plans to capture participant comments in the same manner as in the Pilot, some efforts should be made to get participant feedback on the exam after rollout. For example, the PMT could periodically call some TTLs and get their comments using a structured instrument. WBIEG has developed a customized Level-1 questionnaire, which will be administered to all TTLs who take the accreditation exam during the rollout. The hotline support system could be another source of information.
- And more importantly, does the TLAP reach its objectives, such as appropriate rules and policies being followed by the TTLs who have gone through the learning modules and received accreditation? An impact study is necessary after a certain period of time from the rollout in December, 2002—WBIEG will develop the design in March/April 2003, once the program logframe is finalized.

LIST OF ANNEXES

A. LIST OF MODULES AND LEARNING OBJECTIVES

B. PILOT SCHEDULE

C. END-OF-MODULE LEVEL-1 CHECKLIST (SAMPLE FROM MODULE 1)

D. END-OF-PILOT LEVEL-1 QUESTIONNAIRE

E. LEVEL-1 SURVEY QUESTIONS FOR THE SELF ASSESSMENT TEST

F. LEVEL-1 SURVEY QUESTIONS FOR THE ACCREDITATION EXAM

G. END-OF-PILOT FOCUS GROUP QUESTIONS

H1. ITEM ANALYSIS RESULTS OF SELF-ASSESSMENT TEST

H2. ITEM ANALYSIS RESULTS OF THE ACCREDITATION EXAM

I. SELECTED INTERNAL REPORTS AND DOCUMENTATION

A. LIST OF MODULES AND LEARNING OBJECTIVES

Module Number	Module Name	Learning Objectives
Module 0	Introduction	The learning objective of the program is to provide training to staff to enable them to carry out TF related duties with confidence. It will educate them on (i) relevant policies and procedures applicable to TFs; and (ii) the nature and extent of the responsibilities of the Bank, as administrator of TFs, to donors and recipients; (iii) the applicability of these policies, procedures and responsibilities in common situations facing the task team leaders.
Module 1	Planning for Use of TFs	**Participants will demonstrate increased understanding of:** • The need to link TF use to VPU work programs and activities and consult recipients early in the planning cycle • The fact that TF-supported work should be supplemental and complementary to the Bank's BB-funded work • What constitutes recipient work and what constitutes Bank work • The allowable uses of Consultant TFs
Module 2	Matching funding needs with TF Programs	**Participants will demonstrate increased understanding of:** ▪ The most commonly used TFs for Recipient work and for Bank work. ▪ What encompasses Recipient work and Bank work. • The information resources available regarding TFs, including the characteristics of various types of TFs.
Module 3	Using Consultant TFs	**Participants will demonstrate increased understanding of:** • What the Consultant TF program is about • Characteristics of Bank vs. recipient work to clarify what cannot be funded under CTFs
Module 4	New TF Arrangements (ICNs and IBTFs) Requirements	**Participants will demonstrate an increased understanding of:** • The proper procedures for establishing partnership arrangements, initiating discussions on resource mobilization for a new TF, and setting terms and conditions for the TF. • The need for careful attention to policy and legal issues surrounding establishment of TFs, before discussions with donors. • The roles of various departments in clearing and approving TF documents and legal agreements

Module Number	Module Name	Learning Objectives
Module 5	Bank vs. Recipient execution	**Participants will demonstrate increased understanding of:** • The decision on whether a TF should be Bank-executed or recipient-executed
Module 6	Alignment of TOR with Approved Activities and Legal Requirements	**Participants will demonstrate increased understanding of:** • The main legal agreements for TFs financing recipient activities • The need for Consultant TORs to be consistent with the legal agreements and operational characteristics of the TF, and with Bank policy. • The need for TORs to be consistent with the objectives and approved activities for TF financing. • Whom to consult for advice on Consultant TORs
Module 7	Activating a TF	**Participants will demonstrate increased understanding of:** • The procedures and need for activating a TF • When expenditures can be charged to a TF
Module 8	Selection of Consulting Firms	**Participants will demonstrate increased understanding of:** • The method of selecting consulting firms for activities to be financed by recipient executed and Bank-executed TFs
Module 9	Selection of Individual Consultants	**Participants will demonstrate increased understanding of:** • The method of selecting individual consultants for activities to be financed by recipient executed and Bank-executed TF.
Module 10	Disbursements	**The participant will demonstrate increased understanding of:** • The disbursement and financial management provisions to be included in the Grant Agreement • Internal actions required to commence disbursements for Bank executed TFs • Information needed to approve disbursements for Bank-executed TFs • Rules for correcting erroneous recorded disbursements for Bank-executed TFs • Retroactive financing and Interim Budgets
Module 11	Monitoring a TF Using SAP	**Participants will demonstrate increased understanding of:** • The need to review expenditures and propriety of postings • The importance of paying attention to exceptional conditions, such as inactive funds and funds past expiration • TTL responsibilities, including regular monitoring of TF financial activity • The reporting documents that provide TTLs with crucial monitoring information

Module Number	Module Name	Learning Objectives
Module 12	Reporting Requirements and Disclosure Policy	**Participants will demonstrate increased understanding of:** • The need for periodic monitoring of TFs • The various reports available in SAP to assist in monitoring requirements • The Bank's disclosure policy and whom to consult when questions arise on disclosure of information
Module 13	Amendments to Legal Agreements	**Participants will demonstrate increased understanding of:** The guidelines for amending TF agreements, including: • When an agreement must be amended • Who clears proposed amendments • Who is authorized to sign amendments
Module 14	Closing a TF	**Participants will demonstrate increased understanding of:** • The fact that TFs have clear ending dates • The implications of Closing Dates and the related actions to be taken • What is involved in closing and extending a TF
Module 15	Using SAP For TFs	**Learning Objectives:** • Understand the critical information and integration points that must be created in order to allow a TF to disburse (budget structure, funds assignment, budget loading.) • Understand the major modules in SAP used to support TFs, including the key control points, and financial monitoring tools. • Understanding the significance of Letters of Representation (LORs).
Module 16	Documentation	**Participants will demonstrate increased understanding of:** • Documents to be kept on file • Where information on TFs that needs to be retained can be found • AMS 10.00: Management of Bank Group Records and Archives

B. PILOT SCHEDULE

Day	Date	Activities	Cohort-A	Cohort-B
Wednesday	**2nd October**	**Introduction Session** (introduction to program, objectives of pilot, incentives for participants, registration, navigation through WebCT and Modules 0 & 1, take self assessment test)	**Meet in Room U3-M 11.00 - 13.00**	**Meet in Room U3-M 14.00 - 17.00**
Thursday	3rd October	Complete Modules 1, 2	study at your own convenience	study at your own convenience
Friday	4th October	Complete Modules 3, 4	study at your own convenience	study at your own convenience
Saturday	5th October	Weekend		
Sunday	6th October	Weekend		
Monday	7th October	Complete Modules 5, 6	study at your own convenience	study at your own convenience
Tuesday	8th October	Complete Modules 7, 8	study at your own convenience	study at your own convenience
Wednesday	**9th October**	**Clinic and Module Review Session** (progress on study modules, questions & answers, feedback) **Optional** (on-line Navigation assistance**)**	**Meet in Room U3-M 11.00 - 12.00 Mandatory** **12.00 - 13.00 Optional**	**Meet in Room U3-M 14.00 - 15.00 Mandatory** **15.00 - 16.00 Optional**
Thursday	10th October	Complete Modules 9, 10	study at your own convenience	study at your own convenience
Friday	11th October	Complete Modules 11, 12	study at your own convenience	study at your own convenience
Saturday	12th October	Weekend		
Sunday	13th October	Weekend		
Monday	14th October	Columbus Day Holiday		
Tuesday	15th October	Complete Modules 13, 14	study at your own convenience	study at your own convenience
Wednesday	16th October	Complete Modules 15, 16	study at your own convenience	study at your own convenience
Thursday	**17th October**	**Exam Session-1** (take exam, give feedback) **Optional Session-1** (participate in focus group discussion)	**Meet in Room U3-M 10.00 - 12.00 Mandatory** **12.00 - 13.00 Optional**	**Meet in Room U3-M 14.00 - 16.00 Mandatory** **16.00 - 17.00 Optional**
Monday	**21st October**	**Exam Session-2** (only for those who are unable to attend the first session) **Optional Session-2** (participate in focus group discussion)	**Meet in Room U3-M 10.00 - 12.00 Mandatory** **12.00 - 13.00 Optional**	**Meet in Room U3-M 14.00 - 16.00 Mandatory** **16.00 - 17.00 Optional**

C. END-OF-MODULE LEVEL-1 CHECKLIST (SAMPLE FROM MODULE 1)

Trust Funds Learning Accreditation Program (TLAP)

Checklist for Module 1, "Planning for Use of Trust Funds"

Dear TLAP Participant:

The short questionnaire below asks for your opinions about **Module 1 – Planning for Use of Trust Funds** that you have just completed. The feedback you provide is very important, as it will assist the TLAP design group improve the program over the next couple of months to get it ready for the final rollout for your colleagues.

To answer, please fill the circle like this ●. If you wish to change an answer, fully erase or draw an **X** over the unwanted mark and fill the circle indicating your preferred answer. Please print or write your comments clearly. The questionnaire should take you 5-10 minutes to complete.

The information you provide is completely confidential; it will be used for TLAP program improvement purposes only.

Thank you for your collaboration!

Achievement of Objectives

1. Did the Learning **Module 1 – Planning for Use of Trust Funds** meet the objective of providing you with an increased understanding of:

(a) the need to link Trust Fund use to VPU work programs and activities and consult recipients early in the planning cycle?	? Yes ? No	If No, Please explain _____ _____ _____
(b) the fact that Trust Fund-supported work should be supplemental and complementary to the Bank's BB funded work?	? Yes ? No	If No, Please explain _____ _____ _____
(c) what constitutes recipient work and what constitutes Bank work?	? Yes ? No	If No, Please explain _____ _____ _____

(d) The allowable uses of Consultant Trust Funds?	? Yes ? No	If No, Please explain _____ _____ _____

Logic/Sequence of the Content of the Learning Module

2. Did the information in the module follow a logical sequence ?	? Yes ? No	If No, Please explain _____ _____
3. Was the information presented in the module clear and easy to understand?	? Yes ? No	If No, Please explain _____ _____

Relevance of the Content

4. Was the information in the module relevant for you for managing a Trust Fund?	? Yes ? No	If No, Please explain _____ _____

Technology

Using the options to the right, please answer each question below.

- From your work station:

	Easy	Difficult	Did not use
5. Using the menu for navigating was	?	?	?
6. Navigating backward and forward was	?	?	?
7. Using hotlinks was	?	?	?
8. Using the dynamic flowchart was	?	?	?

Below, please provide a brief explanation of the difficulties

_____

_____

9. How long did it take you to complete the module? (Please fill in)	_____ Minutes	

10. Were any of the sections of the Learning Module difficult to understand, unclear, or problematic? If so, please ✓ the sections and explain the difficulty.

		Check	Comments
a.	Scenario 1	❑	_____
b.	Expert's Observations	❑	_____
c.	Questions (Scenario 1)	❑	_____
d.	Dynamic Selection of a Trust Fund Chart	❑	_____
e.	Questions and Answers for Scenario 1	❑	_____
f.	Scenario 2	❑	_____
g.	Questions and Answers for Scenario 2	❑	_____
h.	Resources	❑	_____

Overall Assessment of the Module

	Very poor				Very good	N/A
11. What is your overall assessment of the module?	?	?	?	?	?	?

If you don't mind the TLAP program contacting you for further feedback, please provide your name below

Name

_____

Telephone

_____

Thank you for completing this checklist.

D. END-OF-PILOT LEVEL-1 QUESTIONNAIRE

Trust Funds Learning Accreditation Program (TLAP) Pilot

Level-1 Questionnaire

Dear TLAP Participant:

The short questionnaire below asks for your overall opinions about **the Trust Fund Learning Accreditation Program Pilot** that you have just completed. The feedback you provide is very important, as it will assist the TLAP design group improve the program over the next couple of months to get it ready for the final rollout for your colleagues.

To answer, fill the circle like this ●. If you wish to change an answer, fully erase or draw an **X** over the unwanted mark and fill the circle indicating your preferred answer. Please print or write your comments clearly. The questionnaire should take you 5-10 minutes to complete.

The information you provide is completely confidential; it will be seen by a small team in WBIEG only. The information will be used for TLAP program improvement purposes only.

Thank you for your collaboration!

Background Information

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

Regional Vice-Presidency				*Network* Anchors						*Other*	
?	AFR	?	LCR	?	ESSD	?	OPCS	?	DEC	?	IFC
?	EAP	?	MNA	?	FSE	?	PREM	?	HRS	?	MIGA
?	ECA	?	SAR	?	HDN	?	PSI	?	LEG	?	Other, specify:
				?	ISG			?	WBI		

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

? GA-GD ? GE ? GF ? GG ? GH or above ? Other, specify: ____

3. What is your **gender**?

? Female ? Male

Learning Modules – Content

Using the rating on the right, please answer each question below.

❑ **How would you rate the Learning Modules on:**

	Very poor				Very good	N/A
4. Overall **achievement of the TLAP objectives**?	?	?	?	?	?	?
5. Overall **usefulness** of the information presented to your management of Trust Funds?	?	?	?	?	?	?

6. Overall relevance of the information presented to your management of Trust Funds?	?	?	?	?	?	?
7. Overall **quality** of the presentation of the information?	?	?	?	?	?	?

❑ **Please provide your estimate in percentages for the questions below:**

8. What percent of the information presented in the learning modules did you already know prior to entering the learning program? _____ %

9. What percent of the information presented in the Learning Modules do you think you know well now? _____ %

10. What percent of the information is directly applicable to your management of Trust Funds? _____ %

Your Rating Of Yourself As A Trust Fund Manager

	Novice	Competent	Expert
11. How would you have rated yourself on knowledge of trust fund management and implementation **prior to joining the pilot?**	?	?	?

Quality of the Pilot

❑ **How would you rate the TLAP Pilot on:**
(Using the scale on the right, please rate each question below.)

	Very poor				Very good	N/A
12. Support/facilitation provided throughout the Pilot (i.e., in class support, hotline support)	?	?	?	?	?	?
13. Pacing of the Pilot (i.e., the schedule, the time devoted to different aspects of the pilot, the overall timeframe)	?	?	?	?	?	?
14. Information provided about the Pilot	?	?	?	?	?	?
15. Overall quality of the Pilot	?	?	?	?	?	?
16. Your overall satisfaction with the Pilot	?	?	?	?	?	?

17. Please provide any other comments you may have about the Pilot.

_____

_____

_____

Thank you for completing the questionnaire.

E. LEVEL-1 SURVEY QUESTIONS FOR THE SELF-ASSESSMENT TEST

1. In your opinion, about what percent of the Test is based on the information presented in the Trust Funds Learning Modules?
   _____ Percent.

2. How would you rate the difficulty of the Test?
   1) Very difficult
   2) Difficult
   3) Neither difficult nor easy
   4) Easy
   5) Very easy

3. How would you rate the presentation of the Test questions?
   1) Very poor
   2) Poor
   3) OK
   4) Very good
   5) Excellent

4. How would you rate the information provided about the Test?
   1) Not at all helpful
   2) Somewhat helpful
   3) Helpful
   4) Very helpful
   5) Extremely helpful
   6) Did not use

F. LEVEL-1 SURVEY QUESTIONS FOR THE ACCREDITATION EXAM

1. In your opinion, about what percent of the Exam is based on the information presented in the Trust Funds Learning Modules?
   _____ Percent.

2. How would you rate the difficulty of the Exam?
   1) Very difficult
   2) Difficult
   3) Neither difficult nor easy
   4) Easy
   5) Very easy

3. How would you rate the presentation of the Exam questions?
   1) Very poor
   2) Poor
   3) OK
   4) Very good
   5) Excellent

4. During the Exam, did you use the Learning Modules and/or related resources as a reference?
   1) Yes
   2) No

5. To what extent were the Learning Modules and/or related resources helpful to you in completing the Exam?
   1) Not at all helpful
   2) Somewhat helpful
   3) Helpful
   4) Very helpful
   5) Extremely helpful
   6) Did not use

G. End-Of-Pilot Focus Group Questions

Accreditation Program (TLAP)

Focus Group

General Questions

- What aspects of the Pilot worked best?
- What aspects were problematic?
- What are your recommendations for improving the next Pilot?

[Prompts: Facilitation, Support, Timeframe, organization, up-front general information about the program]

Questions about the Learning Program

- Which Learning Modules were especially good? Why?
- Which Learning Modules were particularly problematic? Why?
- What are your overall recommendations for improvement?

Questions about the Use of Technology

- What did you like about on-line, self-paced learning?
- What did you not like about on-line, self-paced learning?
- What are your overall recommendations for improvement?

Questions about the self-assessment test and Examination

- What aspects of the self-assessment test and the Exam worked well?
- What aspects were problematic?
- What are your overall recommendations for improvement?

[Prompts: Up-front general information about the program, amount of time, navigation, presentation]

H1. ITEM ANALYSIS RESULTS OF SELF-ASSESSMENT TEST

Item ID*	Diffi. Index[1]	Discrim. Index[2]	Distracter[3] Analysis, counts (positive PBC for non-key distracters)[5,6]					
			Key[4]	a	b	c	d	Total
48 Items, 40 Respondents								
Q1	**0.875**	0.552	d	1	3	1	**35**	40
Q2	0.625	0.240	b	14	**25**	1	0 .	40
Q3	0.675	0.511	b	8	**27**	2	3	40
Q4	0.425	0.552	d	12	8	3	**17**	40
Q5	**0.925**	0.293	d	3	0 .	0 .	**37**	40
Q6	0.500	0.313	a	**20**	3	3 (0.06)	12	38
Q7	**0.800**	0.311	b	2	**32**	5	1	40
Q8	0.775	0.350	c	4	1	**31**	4	40
Q9	0.375	0.546	d	18	5	2	**15**	40
Q10	0.375	**0.098**	c	17	4 (0.05)	**15**	4	40
Q11	0.475	0.209	a	**19**	5	5	9	38
Q12	0.575	**0.078**	c	0 .	9	**23**	6 (0.01)	38
Q13	0.725	0.331	d	3	4	4	**29**	40
Q14	0.200	0.425	a	**8**	30	0 .	2	40
Q15	**0.800**	0.218	c	4	2	**32**	2	40
Q16	0.575	0.297	c	9	7	**23**	1	40
Q17	**0.925**	0.378	b	0 .	**37**	0 .	3	40
Q18	0.575	**-0.077**	c	15	0 .	**23**	2 (0.24)	40
Q19	0.400	0.409	c	12	2 (0.14)	**16**	9	39
Q20	0.575	0.429	b	12	**23**	1	2 (0.03)	38
Q21	0.725	0.349	d	6	5	0 .	**29**	40
Q22	**0.950**	0.161	d	1	0 .	1	**38**	40
Q23	**0.975**	0.147	d	1	0 .	0 .	**39**	40
Q24	**0.800**	**-0.219**	a	**32**	1	1	6 (0.27)	40
Q25	0.350	0.307	d	7	2 (0.14)	16	**14**	39
Q26	0.325	0.465	c	16	0 .	**13**	11	40
Q27	0.650	0.355	c	5 (0.06)	2 (0.13)	**26**	7	40
Q28	0.400	0.504	c	10	6	**16**	8	40
Q29	0.600	0.409	b	8	**24**	2	4	38
Q30	0.325	0.285	a	**13**	13	5	9 (0.07)	40
Q31	0.075	**-0.007**	d	24 (0.36)	0 .	12	**3**	39
Q32	**1.000**	N/A	c	0 .	0 .	**40**	0 .	40
Q33	**0.900**	0.188	d	4	0 .	0 .	**36**	40
Q34	0.575	0.259	d	2 (0.19)	4	10	**23**	39
Q35	0.725	**-0.179**	b	3 (0.06)	**29**	4	4 (0.17)	40
Q36	0.750	**-0.591**	d	3 (0.14)	0 .	7 (0.53)	**30**	40
Q37	0.500	0.170	b	0 .	**20**	4 (0.08)	15	39
Q38	0.125	0.159	a	**5**	10	18	7 (0.11)	40
Q39	**0.800**	0.283	b	1	**32**	1	6	40
Q40	0.625	0.422	a	**25**	4	5	6	40
Q41	0.625	0.342	c	4	8	**25**	3	40
Q42	0.700	0.272	d	0 .	2	10	**28**	40
Q43	0.575	0.661	b	6	**23**	0 .	10	39
Q44	**0.925**	**0.055**	c	1	0 .	**37**	1 (0.20)	39
Q45	0.425	0.608	b	14	**17**	4	5	40
Q46	0.700	0.481	b	7	**28**	5	0 .	40
Q47	0.675	0.161	c	1 (0.04)	2	**27**	10	40
Q48	0.275	0.456	a	**11**	0 .	6	22	39
Mean	0.609	0.276						
bility Coefficient (Cronbach's Alpha)[7]:								0.83

H1.(CONTINUED) ITEM ANALYSIS RESULTS OF SELF-ASSESSMENT TEST

Note:

1. **Difficulty Index** of an item is the percentage of respondents who answered the item correctly. The higher the index, the easier the item. **Items in the test with difficulty indexes higher than 0.80 (easy items) are highlighted in bold fonts.** A high difficulty index in test probably means the respondents were already very familiar with the content area before they took the course, while a lower difficulty index in the test may imply that the content area represented was not understood by the participants.

2. **Discrimination Index,** defined here as Corrected Point-Biserial Correlation (CPBC) coefficient, is the correlation between *the score on the item* and *the score on all other items in the test*. If an individual item is in good conformity with the rest of items in the test, its CPBC should be a high positive number. In other word, if a respondent performed well in that item, it is expected that s/he would perform well in other items in the test. Low or negative CPBC coefficients could result from poor item-writing, small number of respondents, small number of items in the test, or the multi-dimensional nature of the content knowledge. **The table highlights in bold fonts the CPBC coefficients less than 0.10**, for which the training team should review the corresponding items carefully and consider revising them for future test uses.

3. **Distracters** here purport to *all* the alternative options in multiple-choice questions, *including the correct/key option*.

4. **Key** is the correct option in a multiple-choice question/item, and **marked in bold font**.

5. **Under-selected distracters** are defined as selected by no respondents that are part of items flagged for their problematic difficulty and/or discrimination indices. Good distracters, except the key, should be selected by comparable numbers of respondents. Under-selected distracters indicate either the distracters were not well prepared or the content areas behind the distracters were well understood by the respondents so that these distracters were not "distracting" at all.

6. **PBC-Positive Distracters** are distracters that have a positive point-biserial correlation (PBC) between respondents' *"scores" on this individual distracter* and their *scores on the whole test*. This concept is similar to the corrected item-total correction in Note 2 above, except that it uses the "score" on a distracter of item (rather than the score on the item) against the score of the whole test (all items). For well-behaved non-key distracters (i.e., wrong answers), the PBC should be negative. In other words, people who selected the non-key distracter should tend to score lower in the test. When the PBC becomes positive, it may mean that the distracter or the item was confusing, and should be reviewed by the training team and revised for future test uses. The table presents in parenthesis only these positive PBC for non-key distracters.

7. Cronbach's Alpha is an index of how consistently or reliably a test measures achievement. The index varies between 1 (a perfectly reliable test) and 0 (a perfectly unreliable test). Reliability in general is a matter of confidence in the dependability of a test. It usually increases when tests have more items, and when the items are of moderate difficulty and homogenous in content.

H2. ITEM ANALYSIS RESULTS OF THE ACCREDITATION EXAM

Module	Block	Item	Form				No. of	Score		Qualtiy Index	Ineffect. Disctrat.	Module Difficulty	
			A	B	C	D	Resp.	Raw	Percent			Mean	Range
Module 1	Block 1	EQ1	X		X	X	26	20	76.9%	0.564	a,c		
		EQ2	X		X	X	26	26	100.0%				
		EQ3	X		X	X	26	26	100.0%			92.0%	23.1%
	Block2	EQ4		X			8	8	100.0%				
		EQ5		X			8	7	87.5%	0.322			
		EQ6		X			8	7	87.5%	0.506			
Module 2	Block1	EQ7	X			X	16	14	87.5%	0.190			
		EQ8	X			X	16	16	100.0%				
		EQ9	X			X	16	13	81.3%	-0.078		92.9%	18.8%
	Block2	EQ10		X	X		18	17	94.4%	-0.225			
		EQ11		X	X		18	18	100.0%				
		EQ12		X	X		18	17	94.4%	0.840			
Module 3	Block1	EQ13	X		X	X	26	25	96.2%	0.840			
		EQ14	X		X	X	26	24	92.3%	0.025			
		EQ15	X		X	X	26	26	100.0%			93.9%	12.5%
	Block2	EQ16		X			8	7	87.5%	-0.332			
		EQ17		X			8	7	87.5%	-0.180			
		EQ18		X			8	8	100.0%				
Module 4	Block1	EQ19	X			X	16	16	100.0%				
		EQ20	X			X	16	15	93.8%	-0.105			
		EQ21	X			X	16	16	100.0%			93.4%	16.7%
	Block2	EQ22		X	X		18	17	94.4%	0.322			
		EQ23		X	X		18	15	83.3%	0.700			
		EQ24		X	X		18	16	88.9%	0.002			
Module 5	Block1	EQ25	X				7	6	85.7%	0.025			
		EQ26	X				7	6	85.7%	0.025			
		EQ27	X				7	7	100.0%			90.9%	14.8%
	Block2	EQ28		X	X	X	27	25	92.6%	0.893			
		EQ29		X	X	X	27	23	85.2%	0.223			
		EQ30		X	X	X	27	26	96.3%	0.506			
Module 6	Block1	EQ31	X		X		17	10	58.8%	-0.329			
		EQ32	X		X		17	14	82.4%	-0.056			
		EQ33	X		X		17	17	100.0%			90.2%	41.2%
	Block2	EQ34		X		X	17	17	100.0%				
		EQ35		X		X	17	17	100.0%				
		EQ36		X		X	17	17	100.0%				
Module 7	Block1	EQ37	X		X		17	15	88.2%	0.730			
		EQ38	X		X		17	16	94.1%	0.611			
		EQ39	X		X		17	16	94.1%	0.840		86.3%	17.6%
	Block2	EQ40		X		X	17	14	82.4%	0.126			
		EQ41		X		X	17	13	76.5%	0.300	c		
		EQ42		X		X	17	14	82.4%	0.243			
Module 8	Block1	EQ43	X		X	X	26	26	100.0%				
		EQ44	X		X	X	26	11	42.3%	0.376	b		
		EQ45	X		X	X	26	26	100.0%			75.8%	75.0%
	Block2	EQ46		X			8	2	25.0%	0.000	b,d		
		EQ47		X			8	8	100.0%				
		EQ48		X			8	7	87.5%	0.146			

H2. (CONTINUED) ITEM ANALYSIS RESULTS OF THE ACCREDITATION EXAM

Module	Block	Item	Form				No. of	Score		Qualtiy	Ineffect.	Module Difficulty	
			A	B	C	D	Resp.	Raw	Percent	Index	Disctrat.	Mean	Range
Module 9	Block1	EQ49	X		X		17	17	100.0%				
		EQ50	X		X		17	15	88.2%	0.524			
		EQ51	X		X		17	16	94.1%	0.209		93.1%	11.8%
	Block2	EQ52		X		X	17	16	94.1%	0.146			
		EQ53		X		X	17	15	88.2%	0.129			
		EQ54		X		X	17	16	94.1%	0.171			
Module 10	Block1	EQ55	X			X	16	15	93.8%	0.171			
		EQ56	X			X	16	15	93.8%	0.624			
		EQ57	X			X	16	15	93.8%	0.025		93.2%	16.7%
	Block2	EQ58		X	X		18	18	100.0%				
		EQ59		X	X		18	17	94.4%	-0.332			
		EQ60		X	X		18	15	83.3%	0.630			
Module 11	Block1	EQ61	X		X		17	11	64.7%	-0.109	a,b		
		EQ62	X		X		17	16	94.1%	0.611			
		EQ63	X		X		17	16	94.1%	0.840		86.3%	35.3%
	Block2	EQ64		X		X	17	13	76.5%	0.364	b,d		
		EQ65		X		X	17	15	88.2%	0.201			
		EQ66		X		X	17	17	100.0%	0.611			
Module 12	Block1	EQ67	X		X		17	17	100.0%	0.080			
		EQ68	X		X		17	15	88.2%	-0.143			
		EQ69	X		X		17	9	52.9%	0.372		84.3%	47.1%
	Block2	EQ70		X		X	17	16	94.1%	0.624			
		EQ71		X		X	17	12	70.6%	0.063	a		
		EQ72		X		X	17	17	100.0%				
Module 13	Block1	EQ73	X		X	X	26	26	100.0%				
		EQ74	X		X	X	26	21	80.8%	0.040			
		EQ75	X		X	X	26	23	88.5%	0.257		88.6%	37.5%
	Block2	EQ76		X			8	8	100.0%				
		EQ77		X			8	8	100.0%				
		EQ78		X			8	5	62.5%	-0.474	b		
Module 14	Block1	EQ79	X			X	16	15	93.8%	-0.471			
		EQ80	X			X	16	15	93.8%	0.025			
		EQ81	X			X	16	16	100.0%			93.3%	11.1%
	Block2	EQ82		X	X		18	17	94.4%	0.840			
		EQ83		X	X		18	16	88.9%	0.002			
		EQ84		X	X		18	16	88.9%	0.316			
Module 15	Block1	EQ85	X				7	6	85.7%	0.611			
		EQ86	X				7	7	100.0%				
		EQ87	X				7	7	100.0%			89.0%	37.0%
	Block2	EQ88		X	X	X	27	17	63.0%				
		EQ89		X	X	X	27	24	88.9%				
		EQ90		X	X	X	27	26	96.3%				
Module 16	Block1	EQ91	X				7	7	100.0%				
		EQ92	X				7	6	85.7%				
		EQ93	X				7	7	100.0%			95.8%	14.3%
	Block2	EQ94		X	X	X	27	27	100.0%				
		EQ95		X	X	X	27	25	92.6%				
		EQ96		X	X	X	27	26	96.3%				
Alpha			0.47	0.37	0.78	0.60						89.9%	75.0%
No. of Respondents			7	8	10	9							

I. SELECTED INTERNAL REPORTS AND DOCUMENTATION

(available upon request)

1. A Framework for the Trust Fund Learning and Accreditation Program
2. Establishment of a Proposed Passing Score for the World Bank Trust Fund Accreditation Exam, by Knapp & Associates International, Inc., August 22, 2002.
3. Guide for Reviewing and Modifying TLAP Exam Questions, September 2002
4. Individual Participants' Performance in the Pilot Self-Assessment Test and Accreditation Exam
5. Open Comments from Level-1 Questionnaire on TLAP Pilot
6. Participants' Comments on TLAP Accreditation Exam Questions
7. Participants' Comments on TLAP Self-Assessment Test Questions
8. TLAP Briefing on Pilot for Meeting with Four Pillars, October 31, 2002
9. TLAP Exam Parameters
10. TLAP Module Checklist Evaluation Checklist Data: Overall and Broken Down by Modules, November 5, 2002
11. TLAP Pilot Feedback Sessions—Combined Results, October 9, 2002
12. TLAP Pilot Exam Feedback Sessions, October 21, 2002
13. TLAP Pilot Item Analysis Results for the Accreditation Exam
14. TLAP Pilot Item Analysis Results for the Self-Assessment Test
15. TLAP Pilot: A Summary of Comments from Module Checklists and Level-1 Questionnaires, October 29, 2002
16. TLAP Pilot: A Summary of Level-1 Findings Draft, October 28, 2002
17. TLAP Pilot: A Summary of Pilot Findings, Reviews, and Recommendations for Rollout, October 25, 2002
18. TLAP Pilot Formative Evaluation Framework
19. TLAP Score-Setting Session, August 22 and October 31, 2002
20. TLAP Training Management System User Specification.