



TOEFL.

ISSN 1930-9317

TOEFL iBT Research Report

TOEFLiBT-05
July 2008

*The Impact of Changes in
the TOEFL Examination on
Teaching and Learning in
Central and Eastern Europe:
Phase 2, Coping With Change*

Dianne Wall

Tania Horák

Listening.

Learning.

Leading.

The Impact of Changes in the TOEFL® Examination on Teaching and Learning in Central and Eastern Europe: Phase 2, Coping With Change

Dianne Wall and Tania Horák
Lancaster University, UK

RR-08-37



ETS is an Equal Opportunity/Affirmative Action Employer.

As part of its educational and social mission and in fulfilling the organization's non-profit Charter and Bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

Copyright © 2008 by ETS. All rights reserved.

No part of this report may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher. Violators will be prosecuted in accordance with both U.S. and international copyright laws.

E-RATER, ETS, the ETS logos, GRADUATE RECORD EXAMINATIONS, GRE, TOEFL, the TOEFL logo, TSE, and TWE are registered trademarks of Educational Testing Service (ETS). TEST OF ENGLISH AS A FOREIGN LANGUAGE, TEST OF SPOKEN ENGLISH, and TEST OF WRITTEN ENGLISH are trademarks of ETS.

COLLEGE BOARD and SAT are registered trademarks of the College Board.

Abstract

The aim of this report is to present the findings of the second phase in a longitudinal study of the impact of changes in the TOEFL[®] test on teaching and learning in test preparation classrooms. The focus of this phase was to monitor six teachers from five countries in Central and Eastern Europe as they received news about changes in the TOEFL and began thinking about how these might affect their teaching in the future. Data were gathered during the period of January to May 2005. The teachers responded to monthly tracking questions and tasks that explored their awareness of the old and new TOEFL tests, the features of their test preparation classes, their reactions to the most innovative parts of the new test, and their thoughts about the type of content and activities they would offer once the new TOEFL was operational in their countries. The report includes an analysis of the teachers' awareness, attitudes, and plans, and a discussion of the types of factors that could affect the shape and intensity of TOEFL washback in years to come.

Key words: Washback, impact, TOEFL iBT, longitudinal, Europe

The Test of English as a Foreign Language™ (TOEFL®) was developed in 1963 by the National Council on the Testing of English as a Foreign Language. The Council was formed through the cooperative effort of more than 30 public and private organizations concerned with testing the English proficiency of nonnative speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board® assumed joint responsibility for the program. In 1973, a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the Graduate Record Examinations® (GRE®) Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; GRE Board members are associated with graduate education. The test is now wholly owned and operated by ETS.

ETS administers the TOEFL program under the general direction of a policy board that was established by, and is affiliated with, the sponsoring organizations. Members of the TOEFL Board (previously the Policy Council) represent the College Board, the GRE Board, and such institutions and agencies as graduate schools of business, two-year colleges, and nonprofit educational exchange agencies.



Since its inception in 1963, the TOEFL has evolved from a paper-based test to a computer-based test and, in 2005, to an Internet-based test, TOEFL iBT. One constant throughout this evolution has been a continuing program of research related to the TOEFL test. From 1977 to 2005, nearly 100 research and technical reports on the early versions of TOEFL were published. In 1997, a monograph series that laid the groundwork for the development of TOEFL iBT was launched. With the release of TOEFL iBT, a TOEFL iBT report series has been introduced.

Currently this research is carried out in consultation with the TOEFL Committee of Examiners. Its members include representatives of the TOEFL Board and distinguished English as a second language specialists from the academic community. The Committee advises the TOEFL program about research needs and, through the research subcommittee, solicits, reviews, and approves proposals for funding and reports for publication. Members of the Committee of Examiners serve four-year terms at the invitation of the Board; the chair of the committee serves on the Board.

Current (2007-2008) members of the TOEFL Committee of Examiners are:

| | |
|-------------------------|---|
| Catherine Elder (Chair) | University of Melbourne |
| Geoffrey Brindley | Macquarie University |
| Carol A. Chapelle | Iowa State University |
| Alister Cumming | University of Toronto |
| Craig Deville | University of North Carolina at Greensboro |
| April Ginther | Purdue University |
| Bill Grabe | Northern Arizona University |
| John Hedgcock | Monterey Institute of International Studies |
| David Mendelsohn | York University |
| Pauline Rea-Dickins | University of Bristol |
| Terry Santos | Humboldt State University |
| Steven Shaw | University of Buffalo |

To obtain more information about the TOEFL programs and services, use one of the following:

E-mail: toefl@ets.org
Web site: www.ets.org/toefl

Table of Contents

| | Page |
|---|------|
| Introduction and Background | 1 |
| Changes in the TOEFL | 1 |
| The TOEFL Impact Study in Central and Eastern Europe | 2 |
| Organization of This Report | 4 |
| The Phase 2 Study..... | 5 |
| Aim | 5 |
| Underlying Theory..... | 6 |
| Participants..... | 7 |
| Data Collection | 9 |
| The Tracking Questions..... | 10 |
| Tasks 1 to 5..... | 11 |
| Computer-Mediated Communication | 18 |
| Influence of the Impact Study on the Teachers' Awareness of the New TOEFL | 19 |
| Analysis..... | 20 |
| Results of the Tracking Questions Exercise and Tasks 1 to 5 | 22 |
| The Tracking Questions (January to May 2005) | 22 |
| Question 1—Have You Learned Anything New About the New TOEFL That You Didn't Know Last Month?..... | 22 |
| Question 2—Have You Found Any New Sources of Information? | 23 |
| Question 3—Is the New Test Being Discussed in Your Institution (e.g., by Other Teachers or Management?)..... | 23 |
| Question 4—Have Students Asked About the New TOEFL?..... | 24 |
| Question 5—Are You Worried/Concerned About Anything Having to Do With the New TOEFL?..... | 24 |
| Question 6—Is There Anything Else of Interest Concerning the New TOEFL That Has Happened This Month? | 25 |
| Summary of Responses to Tracking Questions | 25 |
| Task 1 (January 2005)..... | 26 |
| Task 2 (February 2005)..... | 29 |

| | |
|---|----|
| Task 3 (March 2005)..... | 32 |
| Task 4 (April 2005)..... | 34 |
| Task 5 (May 2005)..... | 37 |
| Summary | 40 |
| Teachers' Awareness of and Reactions to the New TOEFL, Skill by Skill | 41 |
| Reading | 41 |
| Writing | 45 |
| Independent Writing | 47 |
| Integrated Writing | 48 |
| Speaking..... | 50 |
| Independent Tasks | 54 |
| Integrated Tasks | 54 |
| Summary | 56 |
| Teachers' Plans for the Future | 57 |
| Timetable for Change | 57 |
| Plans for Teaching Reading | 59 |
| Plans for Teaching Listening | 59 |
| Plans for Teaching Writing..... | 60 |
| Plans for Teaching Speaking | 62 |
| Plans for Teaching Grammar | 63 |
| Planning More Generally | 63 |
| Course Material..... | 64 |
| State of Preparedness | 64 |
| Characteristics of Communication..... | 65 |
| Types of Communication Channels..... | 66 |
| Mass Media Channels: ETS..... | 67 |
| Other Mass Media Channels..... | 68 |
| Interpersonal Channels..... | 69 |
| The Impact Study | 70 |
| Summary | 70 |
| Other Factors Facilitating or Hindering Change..... | 72 |

| | |
|--------------------------------------|-----|
| Background..... | 72 |
| Factors Within the Innovation | 73 |
| Complexity..... | 73 |
| Explicitness..... | 75 |
| Form..... | 79 |
| Factors Within the User System | 80 |
| Classroom Factors..... | 80 |
| Institutional Factors | 81 |
| Summary..... | 88 |
| Discussion..... | 88 |
| References..... | 95 |
| Notes | 99 |
| Appendix..... | 100 |

List of Tables

| | Page |
|---|------|
| Table 1. Phase 2 Participants | 10 |
| Table 2. Phase 2 Tasks: Themes by Month | 13 |
| Table 3. Teachers' Views of the Differences Between the Current TOEFL and the New TOEFL (Task 2, February 2005)..... | 30 |
| Table 4. Scores Given for Integrated Writing Task (Task 3, March 2005) | 32 |
| Table 5. Scores Given for the Speaking Samples (Task 4, April 2005) | 35 |
| Table 6. Features That Do Not Appear in Current TOEFL Teaching That Teachers Say They Will Use in the Future (Task 5, May 2005)..... | 38 |
| Table 7. Teacher's Questions About the New TOEFL..... | 76 |
| Table 8. Teacher Factors That May Influence Future TOEFL Practice | 86 |

List of Figures

| | Page |
|---|------|
| Figure 1. Henrichsen (1989) hybrid model of the diffusion/implementation process..... | 8 |
| Figure 2. Monthly pattern of data collection (January to May 2005)..... | 14 |
| Figure 3. Channels of communication..... | 67 |

Introduction and Background

This report presents the findings of Phase 2 of the TOEFL[®] Impact Study in Central and Eastern Europe, a long-term study investigating whether changes in the TOEFL test will affect classroom practice in TOEFL preparation courses. Before presenting the details of the Phase 2 study, we will review briefly what the changes in the TOEFL test are and summarize the purpose and findings of the first phase of the Impact Study. This background information will provide a useful context for the findings from Phase 2 of the study.

Changes in the TOEFL

Discussions about introducing changes in the TOEFL test began in the early 1990s, and much of the thinking behind the new test is recorded in a series of monographs, which we will refer to as *the framework documents*. These were published in the TOEFL Monograph series (TOEFL MS-16 through TOEFL MS-20). The first of these monographs (Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 2000), which set out a preliminary working framework for the whole of the new test, stated that the goals of the test development program were to design a test that

- was more reflective of communicative competence models
- included more constructed-response tasks and direct measures of writing and speaking
- included tasks that integrated the language modalities tested
- provided more information than current TOEFL scores did about international students' ability to use to use English in an academic environment (p. 3)

The remaining four monographs set out frameworks for the testing of reading (Enright, Grabe, Koda, Mosenthal, Mulcahy-Ernt, & Schedl, 2000), writing (Cumming, Kantor, Powers, Santos, & Taylor, 2000), listening (Bejar, Douglas, Jamieson, Nissan, & Turner, 2000), and speaking (Butler, Eignor, Jones, McNamara, & Suomi, 2000) skills. One of the stages in the process of test development was to introduce the TOEFL CBT (computer-based test) in 1998. This version of the TOEFL was introduced in some countries, while the paper-based test (PBT) continued to be administered in others. The CBT introduced some new features (including a computer-adaptive test for the Listening and Structure sections and the possibility for students to

word process their essays rather than producing them by hand), but the constructs underlying the CBT and the PBT were essentially the same. It was not until 2005, when what is now known as the TOEFL Internet-based test (TOEFL iBT) was launched, that much of the vision of the framework designers was realized.

The differences between the TOEFL iBT and its predecessors are clearly laid out in a table in a booklet called *TOEFL iBT at a Glance* (ETS, 2005). The most significant changes are as follows:

- All four skills are tested, including, for the first time, speaking.
- The Writing section has been expanded, from one task to two.
- There are integrated tests in which students receive input from reading and listening passages and have to produce responses in the Speaking and Writing sections.
- There is no separate test of language structure (grammar).
- Note taking is allowed throughout the test.

The original intention was to launch the TOEFL iBT worldwide in 2005, but plans changed in favor of a phased rollout. The test was introduced in the United States in September 2005 and in several other countries a short time later, and it was due to be launched in the rest of the world in 2006.¹

The TOEFL Impact Study in Central and Eastern Europe

This study (which we will refer to as the Impact Study in this report) was commissioned by Educational Testing Service (ETS), the producers of the TOEFL, in late 2002. The aim of the study was to determine whether changes in the TOEFL test would affect the type of teaching and learning taking place in institutions preparing students to take the TOEFL. This aim was in line with statements found in all of the framework documents regarding the hopes of the designers to create a positive impact (or washback) on classroom practices. Although some researchers make a distinction between *washback* (the influence a test with high-stakes outcomes has on the teaching and learning that precedes it) and *impact* (the effect of such a test on not only the classroom but on the educational system and perhaps society more generally), we use the terms interchangeably in this report, referring to the possible impact of tests on teaching, learning, and

the classroom rather than to their influence on the greater social context. See Wall (1997) for further discussion of washback and impact.

What was envisaged was a longitudinal project that would take place in several phases. Phase 1 (the baseline study) took place between January 2003 and June 2004. The main purposes of this phase were to determine what type of washback the advisers working on the development of the new TOEFL test envisaged for the future and to describe what TOEFL preparation courses looked like in a sample of institutions in the Central and Eastern European region before teachers and students knew about the launch of the new test. If there existed an accurate description of teaching and learning before the new test became operational, this could serve as a point of comparison for later phases of the project that would try to determine whether the test had influenced any changes. The report on Phase 1 has been published as Number 34 in the TOEFL Monograph series (Wall & Horák, 2006), but we offer a summary here to set a context for our report on Phase 2.

The first step in Phase 1 was to find out what type of washback had been envisaged by those who took part in the early development work. Our investigation included a review of a number of important background documents, including the framework documents mentioned in the previous section, and a survey of 10 of the original advisers to the new TOEFL. Our conclusions were as follows:

...there was a general hope that the new TOEFL would lead to a more communicative approach to teaching and that preparation classes would pay more attention to academic tasks and language, there would be more speaking, there would be integrated skills work, and some aspects would change in the teaching of other skills. (Wall & Horák, 2006, p. 17)

The second step was to describe what TOEFL preparation classrooms looked like in the region we were studying. To do this, we carried out observations and conducted interviews with teachers, students, and directors of studies at 10 institutions in six countries: Bulgaria, Croatia, Lithuania, Poland, the Slovak Republic, and Romania. The observations and interviews took place in late 2003, when there was very little awareness that a new version of TOEFL was to be introduced in the future. The teaching we saw and discussed with the teachers was, in the main, quite traditional in its approach. Most of the teachers depended heavily on materials prepared by commercial publishers and conducted classes that focused on language structure and discrete

skills. There was no integrated skills work, and although English was the medium of most classrooms, little work was done to improve students' speaking. We predicted that these teachers might find the transition to the new TOEFL, with its expanded construct and new test methods, quite challenging. (For details of the teaching of all skills areas, plus grammar and vocabulary, see Wall & Horák, 2006, pp. 32–72.)

It seemed appropriate to investigate how the teachers' awareness of and attitudes toward the new TOEFL would develop as the time of the launch drew nearer, and this became the focus of Phase 2 of our project. We did not want to take for granted that any changes we might see in classrooms after the new test was operational were due to the test itself rather than to other factors in the educational environment, so we decided to try to track what teachers learned about the test and how their thinking developed regarding the type of teaching they might do in the future. We hoped that data gathered in Phase 2 would help to determine whether there was an evidential link (Messick, 1996) between the new TOEFL and the type of teaching and learning that would take place after its introduction in these institutions in Central and Eastern Europe.

Organization of This Report

This report is divided into eight sections. In the first section, Introduction and Background, we have contextualized our study by describing the main changes in the new TOEFL test and summarizing the design and finding of our Phase 1 study. In the second section, The Phase 2 Study, we present the aim of the current investigation, the theory underlying it, a description of the teachers who participated in it, and our approach to collecting and analyzing data. In the third section, Results of the Tracking Questions Exercise and Tasks 1 to 5, and the fourth, Teachers' Awareness of and Reactions to the New TOEFL, Skill by Skill, we present the results of specific tasks we set the teachers and a summary of their awareness and reactions to the new TOEFL. In the fifth section, Teachers' Plans for the Future, we present the teachers' plans for their future preparation classes. In the sixth section, Characteristics of Communication, and the seventh, Other Factors Facilitating or Hindering Change, we present a discussion of the factors influencing the teachers' thinking during this transitional period, including the channels of communication they were using to get information about the new test (the sixth section) and important characteristics of the test itself and the educational context (the seventh section). There will be a discussion of the findings in the eighth section, Discussion, and suggestions for further

investigations to see whether the plans the teachers were forming at the end of Phase 2 were actually carried out once the test was introduced in their countries.

The Phase 2 Study

In Phase 2, we monitored how six of the Phase 1 teachers reacted as they began learning about the content and the format of the new TOEFL and tried to understand the challenges they faced as they started thinking about designing preparation courses for the future. Data were gathered between January and June 2005. This phase of the investigation was designed as a transition study that would lead on to a later phase looking at whether changes in practice had occurred after the new test became operational.

In this phase, as in Phase 1, we adopted a qualitative approach to data gathering and analysis, studying a limited number of participants in depth rather than surveying a large number of participants whose responses we would have to study out of context. A distinctive feature of the study is that we were able to work with individuals that we first met in Phase 1 (2003) and could therefore build on the knowledge that we gained of their contexts and circumstances during that baseline period. We make no claims about the generalizability of our findings as that is not the purpose of the research. Instead, we wish to claim that the attention we paid to each individual allowed details and explanations to emerge that may be more helpful than the indication of trends that is often the result of research within a more quantitative paradigm. While analyzing large numbers of questionnaires could potentially give us a broad picture of certain issues, it would not necessarily help us to understand the “why” behind our findings. We hope the points that emerge from our analysis and discussion will prove useful to ETS and to other producers of examinations with high-stakes outcomes.

Aim

The aim of Phase 2 was to investigate how six teachers in Central and Eastern Europe reacted to news that was emerging about the new TOEFL test during the period January to May 2005. None of the teachers were aware of the details of the new test when we interviewed them during Phase 1 (the last quarter of 2003). We needed to establish, first of all, how much they had learned about the test in the 12-month period since we were last in contact with them, and, next, what they understood of the new test in early 2005 and what plans they might have for conducting preparation classes once the test was operational in their countries. We predicted that

the teachers' plans would be vague at the beginning of the data collection period but that they would become more concrete as the launch date for the new test got closer.

Underlying Theory

The theoretical support for the whole of the Impact Study comes from the fields of language testing, general educational assessment, and innovation in education. For a review of the key ideas, see our report from Phase 1 (Wall & Horák, 2006). It is important to note here, however, that the main influences in Phase 2 have been the following:

- Messick's notion of the consequential aspect of validity (1996, p. 254), which encompasses the ideas of washback and impact. Messick claimed that it was essential to consider the washback of a test when evaluating its validity (p. 243). He also argued that before one could claim that washback existed, it was necessary to find an "evidential link" between "teaching or learning outcomes and the test properties thought to influence them" (p. 247).
- Chapman and Snyder's (2000) idea of the intermediate conditions that have to be met before the desired impact of a test can be achieved, the most difficult of which is getting teachers to understand what it takes to improve students' performances.
- The Henrichsen (1989) hybrid model of the diffusion/implementation process, which shows that the awareness and evaluation of those who are expected to react to an innovation (in this case, teachers reacting to the new test) are influenced by many factors, including the channels of communication that are used to transmit messages to them, the characteristics of the innovation, and other features in their educational context. (See Figure 1.)

Building on the ideas of many researchers in the field of innovation studies, Henrichsen (1989) proposed that those who are intent on introducing educational innovations must be aware of factors at three different stages of the diffusion and innovation process. It is important to understand the antecedents of the innovation to be able to decide how likely it is that the innovation will successfully take hold in the intended context. It is important to be aware of process factors in order to determine the best way to introduce the innovation and help the receivers to understand it and react appropriately to it. It is then necessary to analyze the consequences very carefully, both when the innovation has been introduced and in the future, to

see whether the effects of it are as intended and if they are sustained or change as time goes by. In this project we see the new TOEFL as an educational innovation and the intended washback/impact as part of the consequences that may occur. The Phase 1 study was designed to investigate the antecedent situation and the Phase 2 study was meant to investigate a number of factors in the process stage. In particular, we were interested in the teachers' awareness, interest, and evaluation of the new test, given what they had learned of it through various tasks we set them and by exploring the information available from ETS and other sources (factors within the innovation and characteristics of communication in the model in Figure 1). We will gather data related to the consequences in a later stage of the project, once the new test has settled in the region we are studying.

Participants

The participants were six teachers of TOEFL preparation classes whom we had interviewed and observed during Phase 1. The Phase 1 teachers had not been selected because of their own qualities but rather because they were teaching in the types of institutions we needed to make up a useful sample for our baseline investigation. As we indicated in our Phase 1 report (Wall & Horák, 2006, pp. 19–22), we initially planned to visit eight institutions in four countries in Central and Eastern Europe. We wanted to visit institutions that offered both TOEFL preparation classes and English for academic purposes (EAP) classes and that had two teachers teaching both sorts of classes. This would enable us to compare not only TOEFL and non-TOEFL teaching, but also teaching styles (Alderson & Hamp-Lyons, 1996; Watanabe, 1996). We also wished to visit institutions offering the computer-based version of TOEFL as we thought this would be more representative of teaching in the region. We used personal and professional (first- and second-hand) contacts to identify suitable institutions, asked ETS to help us with further contacts, and conducted a trawl of the Internet. We eventually put together a sample of 10 sites in six different countries—eight private language schools and two education information centers. There were six local (non-native speakers of English, local to the country they were teaching in) teachers in the sample and four American expatriate teachers. The local teachers had a high level of English language competence and had received formal training to teach in their state school system. The expatriate teachers had either only basic qualifications in teaching English as a second or foreign language or no teaching qualifications at all.

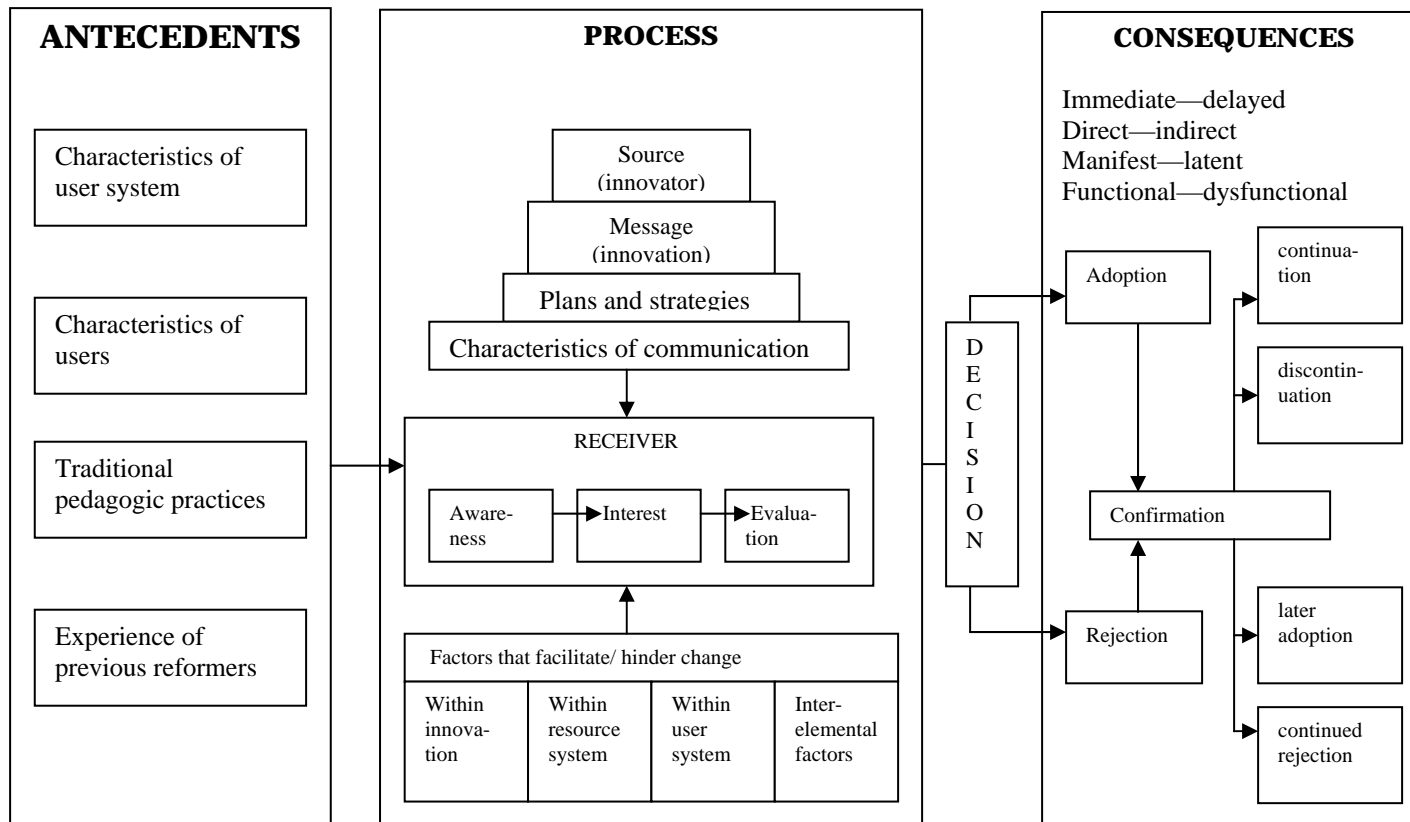


Figure 1. Henrichsen (1989) hybrid model of the diffusion/implementation process.

Note. From *Diffusion of Innovations in English Language Teaching: The ELEC Effort in Japan 1956-1968* by L. Henrichsen, 1989, New York: Greenwood Press, p. 80. Copyright 1989 by L. Henrichsen. Adapted with permission.

We invited all 10 of the Phase 1 teachers to participate in Phase 2, and 6 agreed to continue working with us. We have no reason to believe that these teachers were different from the teachers who did not decide to continue. All six participated with the full consent of their institutions, and they were paid for 25 hours of work—in all, 5 hours per month for 5 months (though some of them clearly worked more than this).

The teachers were working in five countries: Bulgaria (two teachers), Croatia, Lithuania, Poland, and the Slovak Republic. Four teachers were local and two were American expatriates. The local teachers were very proficient speakers and writers of the language. There was a wide range of teaching experience in the sample as a whole: from 2 to 24 years of English language teaching and from 2 to 9 years of TOEFL preparation teaching. Table 1 presents further details of the participants.

Table 1 also indicates the type of institution in which each teacher worked. At the beginning of Phase 2, all of the teachers were working in language schools that offered TOEFL preparation courses, but two teachers' situations changed as the investigation went on. T4 began working more in an educational support organization and became involved in activities that were broader than language teaching, and T6 had to find work in another school as her own school and the chain of schools it was affiliated with closed without warning. We decided to continue working with T6 in spite of this change as she was thoughtful and articulate and the insights she offered remained useful.

Data Collection

The main means of gathering data was via computer-mediated interviews (see the Computer-Mediated Communication section in this report for the rationale for conducting interviews via computers). Each teacher was interviewed twice a month for 5 months. The focus of the first monthly interview was the teachers' responses to a set of tracking questions, and the focus of the second interview was their responses to tasks we set them to explore their understanding of the new test and their ideas about how they might organize their teaching in the future.

Table 1***Phase 2 Participants***

| Teacher ID | Gender | Age (approx.) | Native English speaker | Years teaching in total (teaching English in brackets) | Years teaching TOEFL | Highest academic qualifications | Type of institution |
|------------|--------|---------------|------------------------|--|----------------------|---|--|
| T1 | F | 20s | No | 2 | 2 | University graduate and teacher training | Language school/national education information center |
| T2 | F | 40s | No | 16 (6) | 6 | University graduate and teacher training | Language school |
| T3 | F | 30s | No | 6 | 3 | University graduate and teacher training | Language school |
| T4 | M | 30s | Yes | 16 (7) | 2 | University graduate and masters in education (U.S.A.) | Language school, and later an educational support organization |
| T5 | F | 40s | No | 24 | 9 | University graduate and masters in teaching arts (U.S.A.) | Language school/information center/ Prometric testing center |
| T6 | F | 20s | Yes | 2 | 2 | University graduate (U.S.A.), masters in education (U.S.A.) | Language school, part of large national chain; changed to another language school during Phase 2 |

The Tracking Questions

The purpose of the tracking questions was to find out how much news the teachers and their institutions were receiving about the new TOEFL each month and how they were reacting to what they were learning. The same set of questions was sent out every 4 weeks so that we could

build up a picture of how long it took for news about the new test to reach the teachers, their managers, and their students. The tracking questions were as follows:

- Have you learned anything about the new TOEFL that you didn't know last month?
- Have you found any new sources of information?
- Is the new test being discussed in your institution?
- Have students asked about the new TOEFL?
- Are you worried/ concerned about anything having to do with the new TOEFL?
- Is there anything else of interest concerning the new TOEFL that has happened this month?

There were additional questions in the first month (January 2005) concerning the teachers' TOEFL teaching activities since Phase 1 and their general state of awareness concerning the new test at the beginning of Phase 2.

The teachers were sent the tracking questions in the first week of each month. They could respond to them immediately via e-mail, or they could simply prepare to write about them during the interview we would conduct a few days later. If they responded via e-mail, we read their responses before the interview and planned follow-up questions; if they chose not to respond until the interview, we needed to produce follow-up questions as the interview progressed.

Because the tracking questions were open-ended, the teachers could contribute whatever they felt was interesting and relevant to the research. This was bound to differ from individual to individual, as would be expected given Fullan's notion of "the subjective meaning of educational change" (Fullan, 2001, p. 32). (The results are discussed in the Results of the Tracking Questions Exercise and Tasks 1 to 5 section in this report.) We also wished to find out what all six teachers thought about particular issues and therefore set a number of specific tasks that were to be completed at monthly intervals.

Tasks 1 to 5

The teachers were sent a specific task to complete during the second part of each month. There were five tasks in all, and each was meant to explore a different aspect of either the new TOEFL test or TOEFL test preparation.

We use the term *task* because we wanted to engage the teachers in some activity—filling in a table, applying new information, writing a description, and so on—to get them to think in concrete terms about the topics we wanted their opinion on. We felt that well-constructed tasks would not only focus the teachers’ thinking but also give us evidence of how well they understood or could use certain concepts. We devised the tasks with the help of two specialist groups at Lancaster University, one of which focused on teacher expertise and the other on language testing and assessment. We met with these groups early in the study and asked them to brainstorm a list of features of preparation classes for tests with high-stakes outcomes—features that distinguished these classes from nontest preparation ones—and to indicate which features they believed were relevant to TOEFL preparation classes. We needed this information in order to have a point of comparison for the teachers’ responses to what eventually became Task 1 (see the Task 1 section in this report).

Next, we asked the groups to review a list of tasks typical of communicative language classroom teaching and to indicate whether the tasks would be suitable for use in a TOEFL preparation classroom. This information was used in the construction of Task 5 (see the Task 5 section in this report). We also asked for the groups’ feedback on our plans for Phase 2, as well as their ideas for tasks that could elicit the kinds of insights we hoped to obtain during this phase.

Once we developed the initial idea for each task, we piloted it on teachers from the two research groups who were chosen for having some experience and insights into teaching test preparation classes.

The tasks were typically in three parts and could thus be described as *compound tasks*. The first part required the teachers to reflect on and write about a specific aspect of that month’s theme (e.g., describe your TOEFL preparation classes). The second part asked them to draw a comparison with another aspect of the same theme (e.g., compare your TOEFL classes with your other advanced level English classes) or apply the information from Part 1 in a specific way (e.g., give a score to a sample of writing). The final part involved some synthesis or discussion of the information from the first two parts. We encouraged the teachers to complete each part before previewing the next one to capture their “naïve” impressions about a topic before exposing them to new input and asking for their reactions.

The fact that we only had a 5-month period for collecting data made it difficult to probe the teachers’ views on all aspects of the new TOEFL. We chose to prioritize those aspects that

differed most between the current test and the new test, namely, the integrated tasks and the testing of speaking. The themes that we dealt with are listed in Table 2.

Table 2

Phase 2 Tasks: Themes by Month

| Focus of each task | |
|-------------------------|--|
| Task 1 January 2005 | The nature of TOEFL classes |
| Task 2 February 2005 | Teachers' awareness of the TOEFL, both the current version and the new version |
| Task 3 March 2005 | Teachers' reaction to the integrated writing tasks |
| Task 4 April 2005 | Teachers' reaction to the speaking test |
| Task 5 May 2005 | Possible content and methodology of future TOEFL preparation classes |

The monthly pattern of data collection is illustrated in Figure 2.

Task 1. The aim of this task was to establish what the teachers considered the distinctive features of a TOEFL class to be. Part 1 asked them to list the key features of their current TOEFL preparation classes. Part 2 asked them to indicate whether these features appeared in other test preparation classes they were teaching or had taught in the past and whether there were any features that appeared in other test preparation classes that did not appear in their TOEFL classes. Part 3 asked them to compare features of their TOEFL classes with the features of other advanced level classes they were teaching (including EAP classes). What we hoped to determine was whether the TOEFL preparation courses were in any way narrower than the preparation courses for other tests or the advanced courses students took for reasons other than test preparation. (For more on narrowing the curriculum, see Madaus, 1988, p. 85.) It was important to see whether the teachers felt the TOEFL restricted their teaching in any way and whether they would feel the same way in the future as they prepared their students for the new TOEFL. The

task itself is not included in this report, due to space limitations, but the results are discussed in the Task 1 (January 2005) section below.

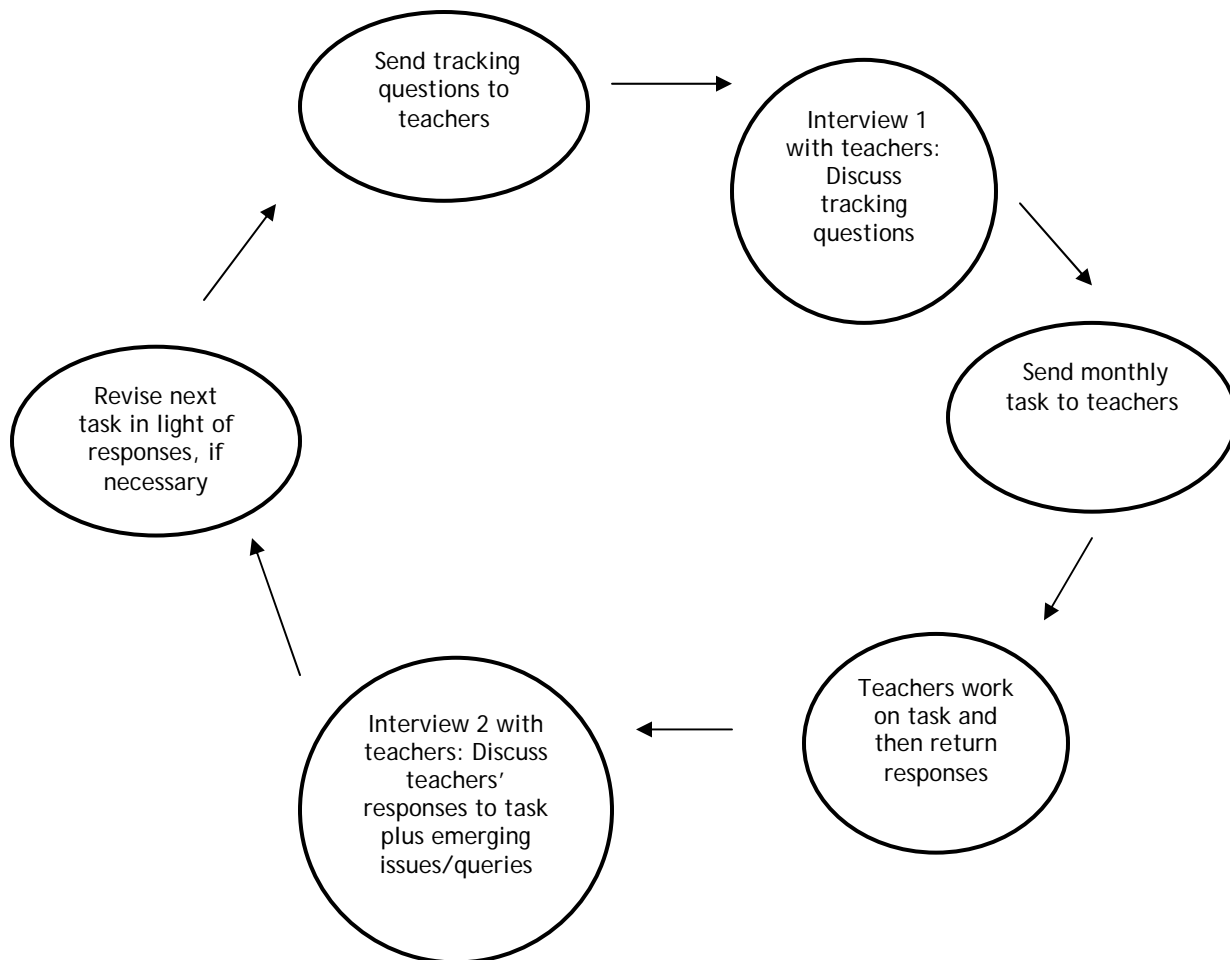


Figure 2. Monthly pattern of data collection (January to May 2005).

Task 2. The purpose of this task was to determine the teachers' level of awareness regarding both the current and new versions of the TOEFL. We wanted to compare their impressions of both tests, but we first needed to establish how much they knew about them and whether their understanding was accurate. In Part 1 of the task, we asked them to fill in a table with key information about the current TOEFL (the TOEFL CBT or the TOEFL PBT, depending on which they were teaching at the time)—for example, how much time was allowed for each section, how many items there were, and what abilities were being tested. The teachers did the same in Part 2 of the task for the new TOEFL. In Part 3 of the task, we asked them to compare

the two sets of information and to indicate what they thought the main differences were and whether they had been aware of these differences before they did the task. We also asked them to indicate their sources of information for both descriptions. Finally, we asked for their initial reactions to the new test and their first thoughts about the possible implications for changes in their future courses. The results for Task 2 are discussed in the Task 2 (February 2005) section below.

Task 3. The purpose of this task was to introduce the integrated writing task in the version of the online practice test then available on the TOEFL Web site. As noted earlier, we wanted to probe the teachers' reactions to the integrated writing task as this was one of the most innovative features of the new TOEFL. Part 1 of our third task asked the teachers to respond to the writing task themselves, as if they were TOEFL candidates. They were to read a short passage, listen to a short lecture on the same topic, and then respond in writing to instructions requiring a comparison of the information they had received from both sources. In Part 2, the teachers looked at a response to the integrated writing task written by a non-native English-speaking student at Lancaster University. The teachers were asked to indicate the type of feedback they would give this student about his writing. This task was meant to give us further insights (beyond those established in Phase 1) into the teachers' views of the type of writing needed for the current TOEFL.

In Part 3, the teachers were asked to score the student writing using the integrated writing task rubric (scoring criteria) for the new TOEFL, which we provided for them. They were asked to underline those parts of the rubric that helped them to decide which score to give. We elicited details of problems they encountered when deciding on a score and asked them whether they would feel confident working with the criteria in their future TOEFL classes. We also asked about the types of support they thought would be helpful if they did not feel confident. The final question asked teachers for their opinions of the integrated writing task and the associated criteria.

We asked for ETS's assistance to find out whether the teachers used the writing rubric correctly. One of the TOEFL experts involved in the development of the integrated writing tasks agreed to score the same piece of writing as the teachers and to give the official view of how the student would have fared on the new TOEFL. We used the expert's score as a measure against which we could first of all assess the performance ourselves (we, too, needed to learn about new

TOEFL standards) and then decide whether the teachers had scored it appropriately. The results of Task 3 are discussed in the Task 3 (March 2005) section below.

Task 4. Task 4 focused on the new Speaking section, another of the innovative features in the new TOEFL. In Part 1 of this task, the teachers were asked to study the new speaking rubric (scoring criteria) for both the independent and integrated speaking tasks. In Part 2, they listened to six speaking samples available in the online practice test and used the rubric to score three of them (Samples 2, 4, and 6). We would have preferred that they scored all six samples, but time constraints prevented this. As with the writing task the previous month, the teachers were asked to explain what, if any, problems they had encountered in scoring; how confident they felt in using the rubric; what support they would find useful in the future if they were not confident now; and what their general opinions were regarding both the independent and integrated speaking tasks and the associated rubrics. In addition, we asked them how they would explain to their students, in language the students would understand, how speaking would be judged on the new test. This was a way of investigating whether the teachers really understood what the task demands were. Lastly, we asked the teachers whether this change in the test had any implications for their teaching in the future.

We again requested assistance from ETS, and an expert from the speaking test development team scored the samples for us and provided a rationale for the scores. We used these comments to decide whether the teachers understood the standards being set by the new TOEFL. The results of Task 4 are discussed in the Task 4 (April 2005) below.

Task 5. Task 5 investigated the teachers' ideas about the types of activities they might offer in their preparation classes for the new TOEFL. We had tried to find out about their plans at various other points during the data collection period but received little from them in the way of details (probably, as will be seen in later sections, because of their preference to wait for test preparation coursebooks before deciding what to do). This task was meant to encourage them to think in concrete terms.

Part 1 of the fifth task focused mainly on the content of classes, and the teachers were asked to react to a list of possible features of a TOEFL class. The list was based on our Phase 1 classroom observation schedule, and was divided into six sections: Listening, Reading, Writing, Speaking, Structure, and Vocabulary. These represented sections in the current and new versions of the TOEFL, and an additional section for vocabulary since this figured so prominently in the

current version (though not in a separate test). Each section was further divided into Subskills Practiced, Exercise Types, and Classroom Activities. The teachers were asked to indicate whether the items listed featured in their advanced level English classes (including EAP classes). In Part 2, they were to indicate whether the same items featured in their current TOEFL preparation classes. The aim of these two parts was to elicit information parallel to that gathered in Task 1, but using a structured technique rather than relying on spontaneously offered information. The teachers were also asked to explain why something might appear in a TOEFL preparation class but not in an advanced level English class and vice-versa.

Part 2 also asked the teachers to consider the methodology of their current classes. They reviewed a list of task types that were typical of communicative language classrooms and indicated whether they used these activities in either their advanced level English language classes (including EAP classes) or their current TOEFL preparation classes. We chose to concentrate on communicative task types because a survey we conducted of experts who had advised on the design of the new test indicated that TOEFL teaching would be more communicative in the future (see Wall & Horák, 2006). The communicative task types were taken from a taxonomy drawn up by Samuda, Johnson, and Ridgway (2000). We included a glossary with the list to ensure that all the teachers understood what each task type entailed. We again asked the teachers to explain why they used some task types in advanced level English classes but not in current TOEFL preparation classes and vice-versa. This helped us to see the range of teaching methods each teacher normally utilized, thus contributing to our understanding of their beliefs and abilities (see the Teacher Factors section in this report for factors that could contribute to the impact of the new TOEFL).

Part 3 of the fifth task, which asked the teachers to describe what they had in mind for their future TOEFL classes, was divided into two: Part 3A and Part 3B. For Part 3A, we presented the teachers with the same list of features that they reviewed in Part 1 of the task, and they were then asked if they would include these in future TOEFL preparation classes. The response options they were given were

I definitely will use this / I might use this / I definitely will not use this.

We gave them the “I might use this” option so that we could be more confident that the features they said they would definitely use or definitely not use were true responses. The teachers were also asked to give their reasons for each response. For Part 3B, they were then

given the list of communicative task types from Part 2 of the fifth task and asked whether they would use any of these in future TOEFL preparation classes.

Computer-Mediated Communication

We decided at the beginning of Phase 2 that we wanted to interact with the teachers rather than just send them tasks and collect their responses. This interaction was desirable so that we could deal with any queries the teachers had about the tasks, ask them to clarify what they had written, and explore issues that were important to them as individuals. Face-to-face interviews were not feasible, given the longitudinal nature of our study and the fact that the participants were working in five different countries. The most attractive alternative we could find was to conduct interviews via computer-mediated communication (CMC).

The use of CMC for language development purposes is well-documented (see, amongst others Burnett, 2003; Greenfield, 2003; Kung, 2004; Reese, 2002; Torii-Williams, 2004), but its use as a research tool is less well known. Beauvois (1997; cited in Kung, 2004, p. 164) has referred to CMC as “conversations in slow motion.” The fact that our interaction proceeded more slowly than in a face-to-face interview proved to be an advantage, for, as Burnett puts it, “summarizing and waiting moves exploit the medium’s potential to generate wide-ranging responses” (2003, p. 259). Each of the participants in the interview had the text of the whole interaction on screen and could peruse this and reflect on it while waiting for the next segment/ response/ question to appear. Another advantage, according to Herring, is that “the availability of a persistent textual record of the conversation renders the interaction cognitively manageable” (1999, p. 2). A third advantage was that a complete transcript of the interview was available to us, the research team, immediately after the interview. This gave us the possibility to discuss issues and queries before the next cycle of data collection began.

We used a combination of asynchronous (e-mail) and synchronous (MSN Messenger) communication, depending on what was most convenient for each teacher. We preferred MSN for our twice monthly interactions due to the slight time advantage it had over e-mail, allowing us to gather more data in the time slot agreed with the teachers. It was necessary to use e-mail in some circumstances, however. For example, one of the teachers moved to a new city halfway through the study and no longer had access to the Internet at home or reliable access at her institution. Using MSN at a local Internet café proved problematic on most occasions. It was

therefore necessary to use e-mail within an agreed time frame, producing contiguous messages and replies.

There are possible disadvantages to using CMC, such as the lack of paralinguistic clues (Burnett, 2003, p. 248), conversation threads getting “out of sync” (Jepson, 2005), and “topic decay” (Herring, 1999, p. 10). These problems are more likely to occur, however, when there are several participants joining in the same conversation rather than when, as in our case, there was just a researcher and a single teacher. We experienced occasional misunderstandings but these were usually dealt with quickly. If they were not caught during the interaction itself they were usually identified in our review of the data afterwards and cleared up in the next interview.

One of the most important issues to consider when using CMC is whether the participants feel comfortable expressing themselves when the researcher is far away and perhaps faceless. “Such willingness requires trust, which can be difficult to build up without an existing prior relationship between the collaborators” (cited in Clyde & Klobas, 2000, p. 284 ; Jarvenpaa, Knoll, & Leidner, 1998). The fact that the teachers knew us from our visits to their institutions in Phase 1 probably contributed considerably to the level and quality of the interaction. Another challenge when interviewing, via whatever medium, is assessing the participants’ candor (Mann & Stewart, 2000, p. 212). Again, this was made easier by the personal contact we had already had with all six teachers. We were familiar with them as individuals—their style of discourse, their background, and their experience—and we believe we would have realized if they were not expressing themselves frankly.

The technology sometimes let us down during data collection. All of the participants (including the research team) had technical problems at least once during the study. We are confident, however, that this did not detract from the quantity or quality of the data collected, as all delays were compensated for by a willingness on the part of the participants to reschedule or make alternative arrangements of some kind.

Influence of the Impact Study on the Teachers’ Awareness of the New TOEFL

One important aspect of our study, which needed careful consideration right from the beginning, was the role we would be playing in raising the teachers’ awareness of the new TOEFL test when this was itself an aspect of the situation under investigation. We accept that our study could not claim to be “naturalistic” (Lincoln & Guba, 1985), observing exclusively “from the outside in” without affecting the process being studied, but we made great efforts to

reduce the level of intervention. We could not impart any privileged information since we had none ourselves, and we provided only as much other information as needed to enable the teachers to carry out our tasks. Most of the materials we provided were in the public domain. The teachers could have tried out sample materials on the TOEFL Web site even if we had not asked them to, and they might well have done this, given that they told us in the beginning that this Web site was their main source of information. We have to acknowledge, however, that taking part in the study might have raised their awareness faster than what was already happening:

T3: . . . this project helped me understand what the new TOEFL would be like. It helped greatly.

Interviewer: That's good. Do you think you would have eventually found out what you now know?

T3: Maybe, yes, but it would have taken a lot of time and effort (T3: 127.142).²

We must also acknowledge the possibility that the teachers thought longer and harder about the changes in the test because they were taking part in a research project (the Hawthorne Effect— see Cohen, Manion, & Morrison, 2000, p. 156). We cannot, however, verify that such an effect existed as ours was not an experimental design with a limited number of variables. What would be interesting to investigate in the future is whether the teachers who took part in this phase of the investigation will offer TOEFL classes that will be noticeably different from those of the teachers who only participated in Phase 1.

Analysis

There were at least five data collection opportunities for each teacher every month: the tracking questions, all three parts of the monthly task, and the post-task interview. In some cases, there were more—for instance, if it was necessary to have more interviews to clear up things that were not understood. A separate file was created for each set of data that reached us in prose, whether it was a section of a task (e.g., Part 1 of Task 1) or an interview. There were 161 prose files, totalling 131,917 words. These files were coded using the facilities provided in Atlas-ti (Scientific Software Development, 2000). The tabulated data from Task 5 (Parts 1, 2, 3A, and 3B) were recast into four new tables that combined the responses of all the participants. These

tables were not amenable to coding using Atlas-ti, and so we dealt with them manually. (See the Task 5 (May 2005) section in this report for details.)

The codes we used to analyze the data were mainly based on the Henrichsen model (1989; see Figure 1), as adapted by Wall (1999). The details of our coding system are explained in Wall & Horák (2006, pp. 29–30), but, briefly, most of the codes fit under headings that correspond to the main sections of the Henrichsen model. We used 157 codes, 84 of which were also used during the analysis of data in Phase 1. There were a number of other Phase 1 codes that were not relevant to Phase 2 as they had to do with the views of students and directors of studies. The new codes represented new themes emerging in this stage of the investigation—for instance, the role of management and the teachers’ relationship with management within the structure of the institutions. (See the appendix for the complete list of codes.)

All of the data loaded into Atlas-ti were coded independently by both researchers. The results were then combined and the analysis proceeded on the basis of a joint view of the data. The data were divided up in various ways: month by month to gain the sense of progress across the group, and person by person to gain an individual perspective. According to Fullan (2001), participants in the process of change experience innovation in their own way, depending on their own combination of circumstances. We wanted to explore not only the general sense of awareness and understanding of the test amongst the teachers, but also the individual teachers’ perceptions, preparation, and planning. If some teachers proved to be better prepared for change than others, we wanted to have insights into why this should be the case. We also wanted to see if some teachers did not seem to be as well-prepared as they should have been.

It is often the case in qualitative investigations that the data are so rich and complex that the researchers have to be selective in what they can explore to conform to time and space restrictions. We have found ourselves in this situation in this study, having to focus our analyses in a way that we feel is most timely and informative given our need to move forward with data collection for Phase 3. In the pages that follow, we present the main results from our tracking question exercise and the five monthly tasks (the Results of the Tracking Questions Exercise and Tasks 1 to 5 section in this report). We then synthesize information from all the communications we had with the teachers, to present our views of the teachers’ understanding of the various skills tested in the TOEFL (the Teachers’ Awareness of and Reactions to the New TOEFL, Skill by Skill section in this report), their early thoughts about future teaching (the Teachers’ Plans for the

Future section in this report), and factors from the Henrichsen (1989) model that could influence the teachers' planning and delivery of preparation courses once the new TOEFL is established in their teaching contexts.

Results of the Tracking Questions Exercise and Tasks 1 to 5

The Tracking Questions (January to May 2005)

We sent the teachers the same set of questions 5 months in a row because we wanted to track whether and how their awareness of the new TOEFL grew and how this awareness affected their own plans and activities within their institutions. We summarize the results for each question as follows.

Question 1—Have You Learned Anything New About the New TOEFL That You Didn't Know Last Month?

The teachers' responses generally suggested that they were not learning very much about the new test apart from what they learned by carrying out our tasks. We know from their answers to Question 2 (below) that they checked the ETS Web site and other sources on occasion but their knowledge of the new test's content and format did not seem to grow as much as we would have expected during this period. We are not sure whether this was because little new information was being posted on the Web site or because the teachers were not processing information as efficiently as they could have. Even information as important as the announcement of the phased rollout (see the Changes in the TOEFL section in this report) seemed to go unnoticed by some of them. One teacher reported that she had found out about the rollout as early as February but others reported it in March (this is also when we also learned about it), in April and as late as in May. It is possible that two of the teachers were still unaware of the delay even at the end of our contact with them in May (T4 and T6).

It is also worth noting that T2 picked up several pieces of wrong and/or irrelevant information during the 5 months of data collection. This may be because she often discussed the new test with her students, who had received information from their peers in other teaching situations. Some of this information may have been incorrect to begin with, or some of it may have been correct but was distorted during the process of transfer from student to student to teacher.

Question 2—Have You Found Any New Sources of Information?

We began asking this question in March, after the teachers had gone to the TOEFL Web site to learn about the new test for our Task 2. It is clear that the teachers did not use many other sources of information during the data collection period. The additional sources they mentioned were the ETS tour of the new TOEFL (one had a CD version of this and one succeeded after several attempts to access the online version), a seminar that was held in one of the institutions in May, and some information that arrived from the Prometric organization (a U.S.-based provider of testing and assessment services) in May. T4 claimed that no one in his network had any information about the test (T4: 123.15). We know from other questions that some of the teachers contacted coursebook publishers' representatives for news. T2 was especially active in this endeavour, and she also tried to get information from the Fulbright office in her city (the Fulbright Program sponsors exchanges between the United States and other countries in the world)—but to no avail (T2: 97.91).

Most of the teachers continued looking for information even after they learned of the delay in the test launch date. T3 stopped looking in March, however, feeling no pressure to learn more and no longer having the responsibility for selecting materials for her institution (T3: 73.88).

Question 3—Is the New Test Being Discussed in Your Institution (e.g., by Other Teachers or Management?)

The test was being discussed at some level in four of the six institutions, but much of the discussion seemed quite general and revolved around the questions of how to deal with the teaching of speaking (e.g., T2: 1.81, T2: 69.43, T4: 3.49, and T4: 37.12) and how to get hold of commercial test preparation materials (T1: 347, T1: 40.23–55, T3: 2.57, T3: 36.20, and T3: 42.16). T2's institution made an early decision to expand their course to include the teaching of speaking in TOEFL classes (T2: 126.161). T5 reported elsewhere in the study that she talked with her director of studies about the test, but they did not discuss what future courses would look like as this decision was T5's alone (T5: 124.68). T6 reported early on that the management within her institution was not discussing the test at all (T6: 5.98). It became apparent in March that the managers must have had other priorities as the institution closed its doors, reportedly because of bankruptcy.

Question 4—Have Students Asked About the New TOEFL?

Five of the teachers reported that students either did not yet know about the new test or were not interested in finding out about it because they planned to take the current TOEFL instead. T2 was the only teacher who regularly talked to her students about the new test and who reported the questions they put to her (e.g., T2: 1.100, T2: 69.61, and T2: 97.59). Many of these had to do with the testing of listening and speaking. The students were concerned about the possibility of having to listen to non-native English speakers in the Listening section, even though they believed that this would accurately reflect their target-language use situation. It is not clear where they got the impression that there would be non-native speakers in the Listening section, as this was not mentioned in any of the TOEFL information we had access to. The students were particularly concerned about the Speaking section. T2 was worried because she did not always know how to answer their questions (T2: 121.128).

Question 5—Are You Worried/Concerned About Anything Having to Do With the New TOEFL?

The teachers reported a variety of concerns during the 5-month period. The first, which appeared in January and continued throughout, had to do with the testing and teaching of speaking. They had questions about whether speaking was really necessary in the target language use situation (e.g., T4: 3.62 and T4: 37.24), what the criteria were for scoring (T6: 5.159—this was cleared up in Task 4 in April), what weighting would be given to pronunciation as opposed to other features of speaking (T4: 119.67 and T6: 113.113), how to prepare students for speaking (T3: 2.69), and whether new equipment was needed to teach speaking (T1: 6.57 and T1: 40.184).

An equally important area of concern was the nonavailability of information and materials. T2 worried that there were not enough free sample tests available online (T2: 69.66). T6 worried about whether her students would even be able to get online: The poor technical provision in her country made it difficult for students to access whatever materials there were (T6: 5.130, T6: 39.6, and T6: 76.166). There were two queries about the standard of the test: how it was decided which standard of performance was adequate for university study (T4: 117.43), and whether the standard for the new test would be the same as that of the current test (T5: 93.229). Finally, there were questions from T4 about the construct underlying the new test: in particular, whether what being tested was language ability, level of education, or the candidates' ability to guess what the scorer was looking for in the case of the writing test (T4: 92.118). T4,

who was a native speaker of English, was annoyed that the response he had produced for the integrated writing task in Task 3 had received a lower score from the automated rating system than he thought it should have.

Question 6—Is There Anything Else of Interest Concerning the New TOEFL That Has Happened This Month?

There was some overlap between Question 6 and Question 5, with issues such as the nonavailability of sample materials, the standard for the test, and the construct being tested (this time the question had to do with whether the TOEFL exam also tested computer skills). Two teachers mentioned that their institutions were not able to send them to training workshops because it was too expensive (T2: 41.154, T5: 38.51, and T5: 44.172), and one of these mentioned how disappointed she had felt when she learned that the discussion list on the TOEFL Web site was for students rather than for teachers (T5: 44.184). Two teachers commented on the delay in the test launch date, with both concluding from this that the test was not yet ready (T3: 98.61, T5: 75.89). One thought the delay was connected to a lack of test preparation materials; the second felt that ETS might have known early on that the test was not going to be ready on time but had announced it anyway for marketing reasons.

Summary of Responses to Tracking Questions

We know from Task 2 (February) that the teachers were not well-informed about the new test at the beginning of this study, and we know that their first exposure to the details of the integrated writing task and the speaking tasks was in Tasks 3 and 4 (March and April). The teachers knew that they needed to learn about the test to design new preparation courses, but their responses to the tracking questions suggested that they did not learn a great deal more over the 5 months we were collecting data. Most of the teachers depended on the TOEFL Web site as their main source of information, but it is not clear that they accessed it regularly or looked at it thoroughly since they reported the delay of the launch date at very different times. They raised a number of points when we asked them about their worries, but their major concerns were how to prepare their students for the Speaking section and when they would be able to obtain commercial preparation coursebooks so that they could begin detailed planning of their new classes.

The purpose of the tracking questions was to show not only what the teachers' concerns were but also when they developed. The most important observation we have here is that the

worries about obtaining commercial materials emerged early on, at the beginning of the study, and were still there 5 months later. Other questions—for instance, regarding the format of the integrated tasks and the Speaking section and the criteria that would be used to score them—appeared early on but generally faded once the teachers had gone through the tasks we set them to familiarize them with these sections. It is important to ask whether they would have gained this information so “early on” (the quotation marks indicate irony since the original launch data was, after all, only a few months away) if the teachers had not participated in this study. Although the TOEFL Web site now includes details such as the criteria for scoring, not all of this information was available to teachers at the time we were collecting our data.

The tracking questions were not the only source of information about the teachers’ awareness and reactions to the new test. Tasks 1 to 5 also yielded a great deal of information, as will be reported below.

Task 1 (January 2005)

As detailed in the Methodology section of this report, the purpose of this task was to find out what teachers thought was distinctive about TOEFL preparation classes as opposed to other test preparation classes and other advanced English or EAP classes. We wanted to discover what the teachers did in TOEFL classes that they did not do in other types of classes, and we hoped to find out what they did in other classes that they did not do in TOEFL classes.

We found that there were certain features appearing in TOEFL preparation courses that might or might not appear in other test-preparation courses but that definitely did not appear in advanced English or EAP courses. These were mainly familiarization activities, aiming to acquaint the students with

- the general structure of the test
- the test preparation coursebook
- test-related software (either ETS-produced or accompanying the coursebook)
- the instructions for every section of the test
- the types of passages used in the Reading and Listening sections
- the subskills tested in the Reading and Listening sections

- the types of questions used in the Reading, Listening and Structure sections
- the topics used in the Writing section
- different levels of performance in the Writing section
- the criteria used to score writing
- the timing for all sections
- the number of scores available for all sections
- ways of analyzing questions so as to narrow down possible answers.

The teachers also mentioned getting students to perform under test-like conditions, giving practice tests on computers (where these were available) and asking the students directly about their strengths and weaknesses. T4 mentioned learning how to learn (T4: 13.07), but this feature was not mentioned by any other teachers. (T4 was not happy doing courses that were test-oriented. All of the features he listed for his TOEFL courses were also listed for his other courses.) One feature that was not mentioned by any teacher was giving the students practice in typing. This may have been because students could practice this at home (indeed, some teachers asked the students to do their writing outside class hours) or because students preparing for the current TOEFL could still do the writing test by hand.

The teachers were asked about features in their other test preparation classes that did not appear in their TOEFL classes. They mentioned several activities to prepare students for the demands of other tests (especially T2; see T2: 17.250), but the activity they mentioned most often was practicing speaking (T1: 16.70, T2: 17.237, and T5: 20.91). This suggests that more speaking may appear in TOEFL classrooms after the introduction of the new test. Other activities that were listed that could conceivably appear in future TOEFL classes include note taking (T6: 21.128) and more analytical reading (T5: 20.93).

We learned in Phase 1 that T1's approach to TOEFL teaching included what she called tips and tricks for answering questions. When we asked her what kinds of tricks she was referring to, she mentioned what seemed to us to be sensible strategies for dealing with language learning in any situation. In Phase 2, however, she definitely seemed to be referring to skills that were useful only on the TOEFL:

In TOEFL courses, the emphasis is on learning tricks and strategies to score high, and every task that is done has reference to a task on the actual test ... In other exam preparation courses, there are more tasks that broaden the test-taker's general knowledge of English, which can then be used to solve the tasks on the test. These tasks do not necessarily directly correspond to the type of tasks on the test (e.g., learning to describe a person, reading texts to learn new vocabulary, practicing passive and not, e.g., "what happens on the test in case of a passive sentence"). (T1: 16.60)

It is important to monitor this teacher after the new test is introduced to see whether she still feels that TOEFL preparation has to be so different from preparation for other tests.

Finally, the teachers were asked about features appearing in their other advanced English or EAP classes that did not appear in TOEFL classes. Speaking was the skill mentioned most often, by five of the teachers (T1: 22.48, T2: 23.354, T3: 24.60, T5: 26.95, and T6: 27.118). Other activities that did not appear in current TOEFL classes but could conceivably appear in the future were using information from listening for follow-up tasks (T5: 26.91), using different text types for reading (T2: 23.246 and T5: 26.94), and practicing different kinds of writing (T3: 24.61 and T5: 26.100), including summary and argumentative writing. One form of interaction that was mentioned by several teachers was group work (T3: 24.62, T5: 26.89, and T6: 27.112). T6 said that in advanced English classes she had more opportunities for group work, kinaesthetic and tactile learning activities and "creative loafing," which she defined as "giving students time to practice conversation and express their own opinions" (T6: 27.112). She said that this kind of activity was not effective for preparing students for the Test of Spoken English™ (TSE®), an optional test that could be taken with the current TOEFL. It is interesting that this teacher was one of two who later questioned whether the Speaking section of the new TOEFL would be useful (T6: 113.146). Her concern was that the tasks were too brief to assess any meaningful speaking.

To sum up the results from Task 1, it is clear that most of the teachers (but not T4) saw TOEFL courses as being different to other test preparation courses and courses of advanced English or EAP. (In later tasks, T4 modified this view.) The main difference was that TOEFL courses mainly offered content and activities that mimicked the current TOEFL. There was less variety in input texts, question types, and writing output and there was little to no attention paid to speaking.

These results match what we saw in Phase 1, where the lessons were very much bound to the pattern of the test, as mediated by the authors of commercial preparation coursebooks.

Task 2 (February 2005)

In Task 2, the teachers were asked to list features of the current TOEFL and the new TOEFL and to say what the differences were between them. They were also asked whether they had been aware of the new features before doing the task, how they reacted to them, and what the implications would be for their future test preparation classes.

Table 3 presents a summary of the teachers' views of the differences between the two versions of TOEFL. T2 mentioned the greatest number of differences (14), but she also indicated that she had not been aware of about two thirds of these before doing the task. It is important to note that T2's original list of differences (before we summarized it for this report) resembled a parroting of the TOEFL Web site (e.g., "Thanks to the new approach, the Next Generation TOEFL test offers a more realistic measure of how well the individuals can communicate in academic settings" [T2: 59.149]). We were not sure whether she understood what she was writing about and later contact with her reinforced this suspicion. The other teachers also reported differences that they had not known about before. The only feature that was mentioned by all six teachers was the testing of speaking, and this feature is one that they all had been aware of previously. Most of the teachers recorded that there would be integrated skills tasks and no separate test of grammar, but they indicated that they had not known about these before. Several differences were mentioned by only one or two teachers, and some of these were things they had only just learned about. The overall impression was that the teachers had not learned much about the new test during the period between our Baseline Study interviews (autumn 2003) and the beginning of Phase 2, 15 to 17 months later.

Four of the teachers reacted positively to the changes they now knew about. (T2: 59.164, T3: 60.21, T5: 62.67, and T6: 63.108). T1 was mainly positive (saying that the test would be more authentic, reliable, and interesting) but the inclusion of a glossary in the Reading section and note taking in the Listening section led her to believe that the new test would also be more difficult (T1: 58.30). T4 was mostly positive but he was negative about the testing of speaking:

. . . because a student doesn't produce spoken English well doesn't mean they cannot be competitive or productive at the university level (T4: 61.55)

Table 3

*Teachers' Views of the Differences Between the Current TOEFL and the New TOEFL
(Task 2, February 2005)*

| Differences | T1 | T2 | T3 | T4 | T5 | T6 |
|--|----------------|----------------|----------------|----------------|----------------|----------------|
| New name—TOEFL iBT | | ✓ ^a | | | | |
| Launch date—September 2005 | | | | | | |
| General focus is communication | | ✓ | | | | |
| Closer to “product-based education” | | | | ✓ ^a | | |
| Different format overall | | ✓ | | | | |
| Different timing for sections, but same timing overall | | ✓ | | | | |
| Results reported online within 15 days | | ✓ ^a | | | | |
| Question types | | | | | | |
| More question types | | | | ✓ | | |
| Analytical questions | | | | ✓ ^a | | |
| Gives more opportunity to show comprehension skills | | | | ✓ ^a | | |
| Exercises in creativity and structural ability | | | | ✓ ^a | | |
| Reading | | | | | | |
| Length of section different | | | ✓ | ✓ ^a | | |
| Fewer passages | | | ✓ | | | ✓ ^a |
| Longer passages | | ✓ | | | | |
| Note taking allowed | | ✓ ^a | | | ✓ ^a | |
| New question types | ✓ ^a | | | | | ✓ ^a |
| Glossary included | ✓ ^a | | | | ✓ ^a | ✓ ^a |
| Listening | | | | | | |
| Different tasks | | | ✓ ^a | | | ✓ ^a |
| Different varieties of English | ✓ ^a | | ✓ ^a | | | |
| Different number of passages | ✓ ^a | ✓ ^a | | | | ✓ ^a |
| More than 2 speakers in longer conversations | | | | | | ✓ ^a |
| Note taking allowed | ✓ | | ✓ ^a | | ✓ ^a | ✓ ^a |

(Table continues)

Table 3 (continued)

| Differences | T1 | T2 | T3 | T4 | T5 | T6 |
|---|----|----------------|----------------|----------------|----------------|----------------|
| Writing | | | | | | |
| Writing will be tested (as opposed to PBT version, where students took TWE [®]) | | | | | | ✓ |
| Two writing tasks (independent and integrated) | ✓ | ✓ ^a | | | | |
| New rating scale 0–5, not 0–6 | | ✓ ^a | | | | |
| Speaking | | | | | | |
| Speaking will be tested | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Integrated tasks | | | | | | |
| Integrated tasks will be included | ✓ | ✓ ^a | ✓ ^a | | ✓ ^a | ✓ ^a |
| Note taking allowed | | ✓ ^a | | | ✓ ^a | |
| Structure/grammar | | | | | | |
| No separate structure test | | ✓ ^a | ✓ ^a | ✓ ^a | ✓ ^a | ✓ |
| Less specific grammar work | | | | ✓ ^a | | |

Note. ✓ = The feature is present.

^aThe teacher was not aware of this feature before doing Task 2 in February.

The teachers' ideas about how the differences might affect their teaching were mostly quite general—for instance, “Organization of lessons and classroom tasks will have to be changed” (T1: 58.37) and “We will pay less attention to grammar and work on the skills of reading, writing, listening, and speaking” (T3: 60.52). There were, however, a few specific ideas beginning to emerge—for instance, the need to practice summarizing and paraphrasing (T1: 58.56) and the potential use of self-assessment (T2: 59.157). Some of these ideas will be discussed in later sections of this report.

In summary, Task 2 provided us with our first opportunity to see how teachers processed and reacted to information about the new TOEFL. It was interesting to see what they focused on and what did not catch their attention. The testing of speaking and the introduction of integrated tasks were the features that stood out most for them. Other features were mentioned by only one or two teachers. They generally seemed to understand what they had read, but at this stage they had not yet been introduced to the details of input, expected output or criteria for judging

performances. Their reactions to the differences they knew about were mainly positive. Their thoughts about how the differences might affect their teaching in the future were not well-developed, as might be expected when so much of what they were considering was still quite new to them.

Task 3 (March 2005)

In Task 3, the teachers were asked to study the criteria for scoring the new integrated writing task and then score a sample of writing produced by one of our international students. The teachers also provided information relating to:

- why they chose their score
- the problems they encountered while scoring
- how confident they felt using the scoring criteria
- the kind of support they needed in the future
- their opinion of the integrated task itself, rather than of their own ability to cope with it

Table 4 shows the range of scores given by the teachers and by a TOEFL expert from the writing test development team.

Table 4
Scores Given for Integrated Writing Task (Task 3, March 2005)

| Score given | Who gave the score? |
|-------------|---------------------|
| 3 | T3, T4 |
| 2/3 | T1 |
| 2 | T2, T5, T6 |
| 1 | TOEFL expert |

Even the harshest teachers gave a higher score than the TOEFL expert, and half the teachers gave a score that was two levels higher on a 0 to 5 scale (T1: 83.14, T3: 85.8, and T4: 86.6). The expert gave a 1 because the student’s writing contained “little or no coherent information from the lecture,” a phrase that comes directly from the scoring rubric. An analysis of the teachers’ reasons for selecting these scores showed that all of them were aware of this

particular problem, but none of them gave the score that corresponded to it on the scale. The teachers mentioned other problems in the writing sample, including organization, grammar, vocabulary, and punctuation, and they generally seemed quite critical, but for some reason they did not select the score that was most appropriate.

When asked about the problems they had encountered when scoring, three of the teachers stated or implied that they felt the language in the writing sample was better than the content (T1: 83.19, T2: 84.14, and T6: 88.12). Two of these (T1: 83.20 and T2: 84.13) had trouble choosing their score because of the student's uneven performance. T6, however, said that she "saw the appropriate score immediately"; she gave a 2 rather than the expert's 1 (T6: 88.13).

All six teachers said that they felt confident using the criteria, and two commented favorably on particular features: one (T5: 87.23) on the descriptor for 0, which she felt would discourage plagiarism; and the other (T6: 88.22) on the importance given to content in the scale.

The teachers wanted support in the future, however. Most mentioned that they wanted writing samples that were scored by experts. Several requested annotations to help them to understand the experts' reasoning. Two teachers wanted more guidance on what to do in cases where performance was uneven. T2 requested

...more sample integrated tasks in which the students' responses are good concerning grammar, structure, and spelling, but they fail to connect the points made in the lecture and the reading. (T2: 84.34; see also T1: 83.33)

T5, who was one of the more reflective teachers, said that she would also like

... to read what was behind the discussion for introducing this part in the TOEFL. What were the primary objectives and what experimental testing has been done? (T5: 87.30)

This comment is important as it relates to a common finding in the literature on educational impact and test washback: that the teachers who are expected to adopt new approaches often do not understand the reasoning underlying the changes. This prevents them from implementing the new approach in the way that was intended.

Finally, all six teachers had positive opinions of the integrated writing task. Several reacted favorably to the scoring criteria. T4, who often seemed reserved when asked about aspects of the new test, was complimentary when it came to the integrated writing component:

If the point of the test is . . . measuring education levels as well as language comprehension and use, then that is done. (T4: 86.26)

T2's comments suggested a real possibility of beneficial washback:

. . . the integrated writing task will help students improve their ability to connect ideas, compare different points and make them think analytically, which is essential in an academic environment. They shouldn't just copy ideas from different sources . . . but should also connect ideas from different sources—like lectures and articles. (T2: 53.139)

In summary, the results of Task 3 showed that the teachers understood how the integrated writing task functioned and generally felt confident about using the scoring criteria. They all gave a higher score than was appropriate, however, which suggests that they needed more guidance on interpreting the descriptors. They requested support in the form of writing samples that were scored and annotated, which they could use to familiarize themselves with the scoring standards and could also use with their students. Their attitudes towards the integrated writing task were positive.

Task 4 (April 2005)

In Task 4, the teachers scored three speaking samples from the TOEFL online practice test and gave reasons for the scores they had chosen. They also provided information about

- the problems they encountered while scoring
- how confident they felt using the scoring criteria
- how they would explain the criteria to their future students, in words that the students would understand
- the sorts of changes they envisaged for their future courses
- the sorts of materials and resources they would need
- the techniques and activities they would use
- the kind of support they needed in the future
- their opinion of the speaking component

Table 5 presents the scores that the teachers gave to the three speaking samples, along with the scores that were given by the ETS expert.

Table 5
Scores Given for the Speaking Samples (Task 4, April 2005)

| Speaking tasks | | | | | |
|------------------|--------------------|--------------------------------|---------------------------|---------------------------|--------------------------------|
| Independent task | | | Integrated tasks | | |
| Rater | Score ^a | Confidence rating ^b | Task 1 score ^a | Task 2 score ^a | Confidence rating ^b |
| T1 | 3 | 2 | 1 | 2 | 2 |
| T2 | 2 | 3 | 1 | ^c | 3 |
| T3 | 2 | 2.5 | 1 | 2 | 2.5 |
| T4 | 4 | 1 | 3 | 3 | 1 |
| T5 | 2 | 1 | 1 | 2 | 3 |
| T6 | 2 | 1.75 | 2 | 3 | 2.8 |
| TOEFL expert | 3 | n/a | 2 | 2 | n/a |

Note. n/a = not applicable.

^a Scale of 0–4. ^b Scale of 1 (not at all confident) to 3 (very confident). ^c Technical problems—unable to score this performance.

The teachers gave a range of scores for each task (scores on a 4-point scale from 2 to 4 for the first task, from 1 to 3 on the second, and between 2 and 3 on the third), showing that they did not share the same understanding of the speaking rubric (scoring criteria). They not only disagreed with one another but with the ETS expert. For two of the samples, only one teacher gave the same score as the expert; in the third sample only three of them did. This is perhaps not surprising though, given that they had only just been exposed to the rubrics (two sets—one for independent tasks and one for integrated tasks) and they were trying out their understanding for the first time. Technical problems prevented us from viewing the specific criteria the teachers focused on when they scored each sample; however, we know from comments they made elsewhere in the task that they focused on different features. The weighting that should be given to pronunciation seemed particularly confusing to them. It was clear that they needed more

exposure to and practice with the criteria if they were to pass them on to their own students in the future.

The main problem the teachers mentioned was how affected they were by hearing a narrator commenting on each speaking sample before they heard the sample itself (T1: 108.09, T4: 111.10, T5: 112.10, and T6: 113.11). The only speaking samples we had access to at the time were from the TOEFL online practice test and this commentary was an integral part of the offering. What is interesting here, though, is that the teachers still differed in the scores they gave, even after hearing the official commentary. This again shows a need for further guidance and more opportunities to practice scoring.

Table 5 also presents the degree of confidence the teachers said they had in their ability to use the speaking rubric. The important point to note here is that the confidence levels ranged from 1 to 3 on a 3-point scale for both independent and integrated speaking tasks. The factors that affected confidence levels included lack of practice, lack of understanding of the standard, not knowing how to deal with uneven profiles (when candidates were stronger in some areas than others), and not knowing what weight to give to accent (as opposed to pronunciation). Time limitations have prevented us from making a full analysis of the teachers' attempts to explain the scoring criteria in language that their students would understand, but our general impression was that the teachers understood the concepts involved even if the scores they gave showed that they did not yet know how to apply the concepts.

The teachers were asked how they would deal with speaking in their future courses. What is notable here is that one of the teachers (T3) said she had not given much thought to this aspect of teaching. We know from other parts of the study that this teacher was waiting for commercial materials to appear and guide her (T3: 36.20, 42.16, T3: 70.13, T3: 91.202, and T3: 98.92). Several teachers mentioned that they would like to obtain software (or even a tape recorder) so that they could put their students in a test-like environment and get them to record their own voices (T3: 7.52, T5: 4.45, T5: 112.170, and T6: 113.79). Four teachers mentioned that they would teach note taking in the future (T1: 114.141, T2: 126.199, T4: 111.57, and T5: 93.162). Several other techniques were mentioned but none by more than one teacher.

The support the teachers wanted included further explanation of the speaking rubric (including the issue of weighting individual aspects of speaking), speaking samples, and lists of possible topics. Even the most experienced and articulate local teacher felt that the teaching of

pronunciation would be challenging. One of the native-speaking English speakers, who seemed in most parts of the study to be very confident in her own opinions, admitted that she was on unsure ground when it came to evaluating speaking:

I feel I would need a lot of preparation to score these tasks, as, as I mentioned before, if it hadn't been for the narrator telling me beforehand how the tests were scored and what was wrong with them, I wouldn't have necessarily given them those scores or seen those faults. Even though I had the scoring rubric in front of me, I didn't agree with all the flaws or strengths that the narrator said the pieces had. I would also need training in how much weight to give each criterion, like content, and how well it answers the question, pronunciation/accent, and ease of speaking. (T6: 113.110)

In summary, the teachers appeared to be very interested in the Speaking section, but their first attempts to judge speaking samples showed that they interpreted the speaking rubric in different ways and their scores were generally not in accord with the TOEFL expert's. Their degree of confidence varied, as did the amount of thinking they had already put into how to teach speaking in the future. The types of support they thought they needed were similar to the types they mentioned for integrated writing—annotated samples of student performances.

Task 5 (May 2005)

Task 5 was a complex activity that first asked the teachers to indicate whether they were focusing in their current teaching on the subskills, exercise types, and classroom activities listed in the Phase 1 observation schedule.³ *Current teaching* refers to both their TOEFL preparation classes and their advanced English or EAP classes. They were then asked whether they planned to focus on these aspects in their future TOEFL classes. We hoped to find out what kinds of changes the teachers thought the new TOEFL might cause them to make in the future. We also asked the teachers to make the same decisions for a set of communicative task types, so that we could build a picture of what each teacher might be capable of teaching given the right conditions. If teachers did not use certain task types now, even in their relatively unconstrained, nonpreparation classes, then it would seem illogical to expect them to use them within future test-preparation classes.

Table 6 presents the subskills, exercise types, and classroom activities that three or more teachers said they did not use in their current TOEFL classes but would begin paying attention to in the future TOEFL teaching.

Table 6

Features That Do Not Appear in Current TOEFL Teaching That Teachers Say They Will Use in the Future (Task 5, May 2005)

| Relevant skill | Feature (subskill, exercise type, or classroom activity) |
|----------------|---|
| Listening | Note taking while listening |
| | Summarizing what has been heard |
| Reading | Linking information from written text to a listening passage on the same topic |
| | Note taking while reading |
| Writing | Summarizing information from different sources (e.g., a listening passage and reading text) |
| | Organizing ideas from listening and/or reading before writing |
| | Writing on topics from ETS pool of possible topics |
| | Writing essay based on a listening passage |
| | Writing essay based on reading text |
| Speaking | Writing essay based on both a listening passage and reading text |
| | Familiarization with ETS TOEFL speaking rubric |

This list suggests that teachers took seriously what they had learned about integrated tasks, as most of the features listed relate to note taking and making connections between different types of input material (listening and reading). What is surprising is that so little appears under speaking. Most of the teachers (all but T1) claimed to be using speaking-related content, exercises, and classroom activities in their current TOEFL classes. This finding does not correspond to our observations in Phase 1. Unfortunately, we were not able to probe this area further as this was the final data collection exercise under the agreement we made with the teachers at the beginning of Phase 2.

Another curious result was that in quite a few of the cases where teachers indicated that they would be introducing features that they did not use in their current TOEFL classes they also indicated that they were not using the features in their advanced English or EAP classes, either. Why might they feel that the new TOEFL required or encouraged content, exercise types, and classroom activities that did not appear in *any* of their current classes, especially when these features were the sort that are generally thought of as being communicative, authentic, realistic, useful, and so on?

In the final part of Task 5, we asked the teachers to look at a list of communicative task types and indicate which ones they used in their current teaching and which they might use in their future classes. We wanted to find out whether the teachers would use different task types if they were preparing students for the new TOEFL. The task types that three or more teachers said they did not use now but might use in the future were

- consensus
- describe and draw/describe and arrange
- information gap
- information transfer
- jigsaw
- jumbled story
- developing narrative
- pair or small group work
- problem-solving
- scenario
- sequencing
- survey

There are two odd findings here. The first is that many of these activities involved student-to-student interaction, which would not, at first glance anyway, seem to be the most obvious way to develop the sort of speaking that the new TOEFL speaking tasks now requires.

The second is that, as was the case above, most of the teachers (especially T2, but not T4) said they would use task types that not only did not appear in their current TOEFL courses but also did not appear in their normal advanced English or EAP courses. Why should their future TOEFL classes contain more communicative task types than their current, nonrestricted, advanced classes? We were surprised by the responses that indicated that there was a possibility that some of these task types would be used in the future, mainly because it did not fit the pattern that we saw in Phase 1 and most of Phase 2, in which teachers conducted classes that contained very few activities that were not directly and obviously related to the test. There is, of course, the possibility that the teachers were reacting positively to these tasks because they thought this was the right thing to do or that they unintentionally became enthusiastic as part of a Hawthorne effect (see the Influence of the Impact Study on the Teachers' Awareness of the New TOEFL section in this report). We were unable to investigate this issue any further since we had reached the end of the data collection period that we had agreed with the teachers. It would be useful in future phases of the research to investigate whether the teachers who said they might use these task types actually include them in their teaching.

Summary

By the end of the data collection period (May 2005), the general picture was that most of the teachers were significantly more aware of the new TOEFL test than they had been in January. They seemed reasonably confident that they understood the layout and content of the new TOEFL. They seemed clear about the main differences between the current test and the new one, the order that the different sections would appear in the new test, and the question types that would be used. T5's response in this exchange could be seen as representative of the whole group:

Interviewer: Which—if any—sections of the new test do you feel you don't understand well enough yet?

T5: I cannot say that I do not understand any section in particular. I rather think I do.
(T5: 129.82; see also T1: 120.19)

What the teachers had not demonstrated, however, was full understanding or control of the criteria that would be used in judging writing and speaking.

Teachers' Awareness of and Reactions to the New TOEFL, Skill by Skill

It has now been established that the teachers' general awareness of the new TOEFL grew from January to May 2005. It is important, however, to analyze what they knew and did not know about specific sections of the test. We will review these sections one by one, presenting first what we understand the important changes in each section to be and then presenting the teachers' awareness and reactions. There will be some overlap in this discussion with what has already been presented, but there will also be new ideas that emerged from the interviews we held with the teachers at least twice a month for 5 months.

Reading

The main changes to the Reading section are that passages on the new TOEFL are longer than before (700 words on average, as opposed to 250–350 on the TOEFL CBT), there are several new item types (sentence simplification and categorizing information to fill in a chart or complete a summary), and there is a glossary that students can refer to (via mouse clicks) to find the meaning of selected words in the passages.

This was the section of the new test that was least commented on by the teachers. This is probably not surprising, given that it seems to resemble the current TOEFL more than any other section and given that we did not devise any tasks to explore the section in detail. Several teachers noticed that passages are longer (e.g., T2: 59.69 and T6: 68.87) and two noticed that there are fewer passages (T3: 60.19 and T6: 63.39). There was little mention of text types or topics. There was some confusion about the amount of time that would be allowed for answering questions (T2: 41.39 and T6: 68.87). Item types were generally perceived as being similar to those in the current TOEFL, although some mention was made of summarizing, paraphrasing, table completion, and inserting text (T1: 58.56, T6: 63.40, and T4: 99.179). T1 felt that the new item types would be “more interesting” (T1: 58.55).

The area where there seemed to be the greatest difference of opinion was over the subskills that were being tested. Two teachers felt that the new test was assessing the same skills as the old test (T2: 65.10 and T3: 66.16), while two others felt that it would be testing more analytical skills (T5: 62.70 and T6: 57.34). T5 had the impression from early on (February) that in the new test

. . . reading resembles GRE[®]/GMAT/SAT[®] . . .(it) does not simply test questions about the text, questions written from top to bottom, but requires synthesizing, comparison, selection—higher order skills! (T5: 44.152)

T6 felt that if the students had fewer passages to read but the same amount of time, and if they were given what she perceived to be more analytical item types, then the reading would be “more reflective of what goes on at an American university” (T6: 68.2). She welcomed this change and also commented that “multiple-choice testing alone is a flawed method for testing a student’s ability to read and comprehend” (T6: 63.19).

T4 felt that at least two other skills were being tested in the Reading section. The first was writing ability.

Some questions asked about topics of paragraphs and led into understanding of introductions, conclusions, and topic sentences. In being competent with these writing structures, one can answer the questions easier. (T4: 49.75)

T4 did not seem to regard this either positively or negatively. He did, however, react somewhat negatively to the idea of having to choose which fragment of text fitted into a larger text:

I could choose two or three most of the time, depending on how I thought the author wanted to write. (T4: 99.181)

At least one other teacher made assumptions about what was being tested. T5 thought study skills such as dictionary use and note taking were being assessed (T5: 56.24). Our understanding is that these are auxiliary skills that help students to maximize their potential in the Reading section rather than skills that are themselves being tested. T5 may have got the idea that dictionary use is being assessed because a glossary function is included in the test. It will be interesting to observe her classes in the future to see whether she spends class time on teaching dictionary skills.

Other teachers did not see the glossary as representing a skill that needed to be mastered, but they welcomed its inclusion for a variety of reasons. T1 felt that if the new tests included a glossary it must be because the reading tasks were going to be more difficult (T1: 64.80);

however, she also recognized that by giving the student this type of support the test might be assessing

. . . something else in these sections, something that can be tested even though the test-taker is given the glossary or the opportunity to write notes. (T1: 64.87)

T6 said that while she believed that students should be asked to work out the meaning of words from context, this could “put undue stress on a student with high English skills to know words that many native speakers would not know” (T6: 68.19). T5 thought that the glossary, along with the review facility and the possibility of taking notes, gave the students “a lot of support and relative freedom in arriving at certain conclusions about what the text is informing us about” (T5: 56.7). She amplified this later, saying that the new test

. . . acknowledged that people have different styles of retrieving information. . . some need a glossary, some work out meaning from context, some analyze mentally, some draw schemes in notes, etc. (T5: 67.2)

The remaining comments made about reading were in the form of queries. One teacher was not sure how the Reading section would be scored, specifically whether students would be penalized for wrong answers (T6: 130.186). Another questioned the effect of asking students to read from the screen (T2: 53.13).

It is interesting that most of the comments relating to reading focused on the Reading section itself rather than on the reading that students would have to do for the integrated writing and speaking tasks.

Listening

The main changes to the Listening section are that only two types of input are used (conversations and lectures) rather than the three types used in the TOEFL CBT. The short dialogues in Section A of the TOEFL CBT no longer appear and the conversations and lectures are generally longer. The test is no longer computer-adaptive. In addition, note taking is now allowed. Indications were made in one of the TOEFL leaflets (ETS, 2005b) that English accents other than those from the United States might feature in some passages. A replay feature of certain questions was also mentioned (ETS, 2005c). Neither of these was mentioned in the samples available on the TOEFL Web site during the data collection period, however.⁴

As was the case with the Reading section, the teachers did not perceive that the Listening had changed significantly from the current version of TOEFL, although the reduction from three to two types of input was noted in response to our prompts in Task 2 (T1: 52.19, T2: 53.19, T3: 54.18, and T6: 57.18). Two of the teachers picked up on the fact that there might be more than two participants in the conversation input (T3: 53.22 and T6: 63.47), and T6 viewed this favorably. She did not feel that students needed different skills in order to listen to a multi-participant conversation, but they did need better control of basic skills:

. . . you need these skills to a higher degree, as there is more vocabulary to comprehend, more inferences and conclusions to make, and most likely more details to remember. If students can do this with dexterity in a foreign language, it shows that they have achieved fluency in the skills being tested. (T6: 68.279)

The other teachers either did not notice this or did not feel it was worthy of comment.

Three teachers mentioned that comprehension of the speakers' meaning and attitude was to be tested (T2: 53.30, T3: 54.33, and T6: 57.38, and T6: 68.97), which is indeed stated in promotional material (ETS, 2005b). This is somewhat confusing, however, since the current TOEFL frequently includes questions asking what the speaker really means (T4: 66.116). In what ways the new items differ from the current items, to what degree, and how prominent this aspect is intended to be, are unclear. However, the teachers will have processed this feature as being new because it was listed on the Web site as being one of the distinguishing features of the new Listening section.

The issue of accents was commented on by only one teacher (T2), but she understood that the input passages might feature non-native speakers of English rather than native speakers speaking a non-U.S. variety of English. She reported that her students were concerned about this, although they recognized that they would be expected to discuss academic matters with other non-native speakers of English in the target language situation (the setting they were aiming to study in; T2: 97.67). T2 thought it was suitable for there to be such input, "as long as the speakers are fluent in English" (T2: 97.80), though in a later interview she admitted that "it might also cause them problems to complete the task" (T2: 109.112). It is not known where she got the idea of non-native input from.

The new feature that attracted the most attention was note taking. This was unanimously welcomed, as it would make the listening tasks more authentic (T1: 120.128), “alleviate the stress of having to remember everything” (T6: 68.178), and help students to organize their responses better (T2: 58.135). Several teachers felt the current test was unfair, in that students had to listen to detailed input before seeing the questions they were expected to answer, therefore having to depend on their memory (T3: 66.114, T2: 29.8, and T4: 49.38). T4 was particularly negative about the current test:

. . . They are not allowed to take notes. Also these are first-time conversations with unfamiliar topics and unfamiliar people. Who doesn't ask clarifying question in those situations? Responding to general questions about themes and topics I can understand. Detailed questions especially in multiple-choice format, no. (T4: 49.44)

As was the case with reading, most of the teachers' remarks about listening had to do with the separate Listening section rather than with the listening skills the students would need to respond to in the integrated writing and speaking tasks.

Writing

The main difference in the Writing section of the new TOEFL is that there are now two writing tasks instead of one. One task is independent and similar to the current TOEFL writing task, and one is integrated—that is, based on input from a reading on an academic topic and a listening passage on the same topic but from a different perspective. The independent task asks students to state a preference or give an opinion about a specific issue; the integrated task asks them to write on a particular aspect of the relationship between the reading and the listening passages. The two tasks are judged by separate rubrics (rating scales), both of which have descriptors for Levels 0 to 5 (rather than from 0 to 6, as in the current TOEFL). The rubric for the independent task is similar to the one used for the current writing task, although it contains more detail and specifically requires adequate detail from the students. The rubric for the integrated task concentrates on the adequacy and accuracy of the ideas taken from the reading and listening passages, though attention is also paid to organization and language errors. Students are required to type their responses rather than being given a choice between typing and writing by hand.

The teachers generally reacted positively to the news that the Writing section would be expanded. T6 was pleased because she thought the Writing section was “the most important part of the test to start with” (T6: 63.58) and that expanding it would stress its importance. Two teachers mentioned the importance of the writing skill in the university setting, during “training” (T3: 66.66) and even as part of the admissions process:

Most top universities evaluate applications very heavily on essays. Ivy League applications are *mostly* essay. They equate written communication with education. (T4: 92.149)

Two teachers felt that having two tasks gave students a better chance of demonstrating their skills, especially since the tasks required different types of writing (T1: 52.53 and T3: 59.61).

There was also a positive reception for the new writing rubric (scoring criteria), with several teachers saying that the new rubric was more specific, more precise, and more student-friendly than the writing rubric for the TOEFL CBT (T2: 90.149, T3: 91.88, and T5: 75.48). T5 said that she now understood how good students had to be to get the best score (*ibid.*). T4 stated that the rubrics were “concise enough to differentiate,” but he qualified his statement by saying that they were useful if there were “multiple judges” (T4: 92.14). T4 was the only teacher to ask what level of writing would be adequate for university admission. His view was that the people who score the test “should provide teachers with that standard” (T4: 117.61).

Curiously, there was some confusion about whether the writing test would be scored by human raters or by an automated writing analysis tool (such as *e-rater*[®]). T6 felt that having human raters was the only way that tasks could be successfully assessed, as writing was “an art, not a science, so it’s not as simple as choosing the correct answer between a, b, c, and d” (T6: 101.39). T4 felt that human raters would be more tolerant than an automated scoring system (T4: 92.203). On the other hand, T2 felt that such a system could be used for the integrated writing task (“. . . if it just tracks some of the main topics and ideas”) but not for the independent task:

. . . the independent writing task is related more to the individual understanding of the topic and individual ideas and perceptions. So there cannot be a fixed model.
(T2: 97.209)

T2 was quite happy with the idea of automated scoring for the integrated task, “if there is suitable software, which I think has already been developed for the purposes of the new TOEFL” (T2: 97.196). It is likely that the idea that the new writing test might be scored in such a way stemmed from the use of automated rating in the online practice test that the teachers were exposed to as part of our Task 4.

Finally, there was also some confusion about how writing scores of 0–6 would be converted into section scores of 1–30 (T2: 97.18).

Most of the preceding discussion has concerned the teachers’ reactions to the writing test in general. Their reactions to the independent task and the integrated task will now be discussed separately.

Independent Writing

Most of the teachers felt that the independent task was similar to the current TOEFL writing task, in which the students have to defend a point of view (e.g., T3: 66.60). However, T5 initially saw the task as somewhat different from the current one and criticized it for not being sufficiently academic. She gained this impression from looking at a writing sample (source unknown):

Interviewer: Can you tell me more about. . . why you don’t feel it’s a task of academic writing?

T5: Because the sample answer does not appear to be ideally structured. . . I don’t think there were more than three paragraphs, which is less than people write now in 30 minutes for an essay, and less means not enough points analyzed. Also, the first paragraph ends in a question. I read recommendations that, for academic writing, questions are not appropriate. Best state a thesis. (T5: 44.128)

T5 later suggested that the Writing section had become less structured and concluded that this was because not everyone who was taking the TOEFL was doing so in order to gain entry to academia (T5: 44.148). She questioned why it was necessary for students to do an independent writing task if they were also going to have to do an integrated task. She felt that the students would not be demonstrating anything different in the independent task that they would not also be demonstrating in the integrated task (T5: 56.18). This was an early reaction, however, and her

confusion seems to have cleared up later. It is worth noting that this teacher was not a novice teacher, but, rather, a very experienced teacher of general English and EAP at the university level. She, nonetheless, had trouble understanding the different task requirements.

As noted earlier, the teachers were generally positive about the scoring rubrics, but two teachers had reservations. One felt that the rubric for independent writing was more general than that for integrated writing (T6: 88.5), while another expressed a worry about subjectivity:

. . . because some people might like the ideas that the writer has given, (while) other people might not like them and think that they are not relevant. (T3: 91.137)

Integrated Writing

Although the teachers occasionally confused the terms *independent* and *integrated*, most of them understood the difference between the two tasks fairly early on. Several reported in February that the aim of the integrated task was to judge the students' ability to comprehend the key points in input material and reproduce this in a well-organized manner and using clear and accurate English (e.g., T2: 53.85 and T3: 54.70). Two teachers mentioned that students would need to compare information (T3: 54.18 and T5: 56.53). Task 3 (March) focused specifically on the integrated writing test so the teachers had the opportunity to try the task themselves and to score the writing of a student. We expected them to be fully aware of task demands after that point.

The teachers' attitudes towards integrated writing seemed mainly positive. T6 was particularly enthusiastic about the addition of this task to the Writing section, saying that it was her favorite (T6: 130.203). She was impressed by its "authenticity to real life in a university classroom" and stated that

. . . writing about academic subjects pursued in an academic environment is, after all, what the TOEFL is supposed to assess the test-taker's ability to do. (T6: 63.69)

She had a clear picture in her mind of what type of writing her students should do to practice for the test:

I think (the five-paragraph essay) format would be more than suitable for the integrated writing task, since that is the type of assignment the five-paragraph essay was designed for in the first place. (T6: 68.201)

T2 appreciated the addition of the new task because it asked for different writing abilities. She stated that although she and her students were confused about the task requirements at first, . . . we all liked the tasks since they help them develop new skills—connecting ideas from different sources. (T2: 126.34)

She was careful to distinguish between making connections between ideas from different sources and merely copying ideas from different sources. T5 was also interested in the matter of plagiarism, perhaps due to her experience as an EAP teacher and being aware of this problem at university level (T5: 124.89).

T5 gave a lot of thought to the integrated task and was concerned about what else it might test in addition to the ability to understand outside input. She felt that an element of individual interpretation was present as well and that the scoring rubrics should explicitly state such interpretation was valued (T5: 8.91).

T4 felt that the task showed a good balance between “comprehension, language use, language production, as well as measuring organized writing skills” (T4: 86.24). He also felt that the task measured “education levels” (T4: 86.26).

The teachers differed in their opinion of the relative difficulty of the independent and integrated tasks. T5 felt the integrated task would be more challenging:

It takes longer to process all the notes and shape your opinion in a structured way, than just brainstorming the issue and writing an essay. (T5: 93.192)

T1 believed that the independent task would be harder since the students had to generate their own ideas as well as find the language to express them in. This seemed more taxing than using concepts and language provided by the input material (T1: 108.1).

Most teachers liked the scoring rubrics for the integrated task. T6 stated that they were clear in relation to content, organization, and language use (T6: 88.23), though later she stated that “more details in the rubrics would help even more” (T6: 119.210). T5 felt confident using the criteria and stated that the descriptor for the score of 0 would be “especially informative . . . of how a source material should be used and discourage them from plagiarism” (T5: 87.24). All of the other teachers reported that they felt confident using the criteria, although T1 felt that she

would want to add half-scores if she used them when scoring her own students' work in the classroom (T1: 83.26).

It is ironic, then, that none of the teachers scored the sample of writing that we gave them in the same way as the TOEFL expert (see the Task 3 (March 2005) section in this report for details). We did not tell the teachers about the expert's score because we did not want to influence them in their understanding of the test any more than was necessary. It will be important to track how their standards for scoring this task develop in the future.

Only one of the teachers expressed negative views about the integrated task, but these were based on a perception of how scoring was carried out rather than on a view of the task itself or the scoring criteria. As part of Task 3, T4 had written his own response to the instructions given in the practice test and had received a 4 from the automated rating system rather than the 5 he expected. He felt that the program did not approve of how he presented certain information:

The program disagreed with me on points I raised and how I compared them. I'm not bitter, but using it as an example. It said I should've mentioned this point ahead of that point and so on—to which I actually disagree! (T4: 92.184)

He felt that a human judge would have given him a higher score based “more on context than on a single point” (T4: 92.204).

It will be interesting to see how this teacher's standards develop in the future, especially given that as a native speaker of English with experience teaching in U.S. universities, he was possibly the teacher who felt most confident about judging writing in an authentic way.⁵

Speaking

The TSE exam has been available to TOEFL candidates since 1979, but they took it only if they wanted to or if they were required to by the particular educational institutions they were hoping to enter. In the new TOEFL, the speaking test is compulsory. The test consists of six tasks: two independent tasks and four integrated tasks (two with listening and reading input, and two with only listening input). Both the independent and the integrated tasks are scored against the same criteria (general description, delivery, language use, and topic development) but the scoring rubric for the integrated tasks includes more detail, especially in the area of topic development.

The teachers mentioned the speaking test relatively frequently and gave the impression that this was the part of the new TOEFL they were most interested in. This seemed natural, given that it was a totally new entity for them. There was some worry that the students would find this test particularly challenging:

Effective preparation for the speaking part will probably be crucial since this could turn out to be the most difficult part or the part students will feel most anxious about. (T2: 6.120)

Nevertheless, most of the teachers were positive about the addition of this new skill to the TOEFL (T1: 52.4, T2: 58.24, T5: 62.12, and T6: 57.66). T5 felt that speaking was “one of the main language use competencies” and T6 felt that it was in this section of the test that students would “truly demonstrate their language proficiency” (T6: 57.66). T1’s reaction was more complex. She felt that others considered the speaking skill to be

. . . the most automatic, the most spontaneous, so sometimes it seems to be the best way to see how proficient someone is. (T1: 114.101)

She distanced herself from this view, however, saying that she did not consider any aspect of language to be more important than any other.

T4 was very much against the inclusion of speaking in the new test, calling it an “irrelevancy” (T4: 3.62) and an “unnecessary hurdle” (T4: 37.24). His main argument was that students do not need to speak in order to study and learn. He also felt certain that if students did well on the other parts of the TOEFL, they would have the basic skills necessary to speak as well. There was, for him, no difference between writing and speaking, apart from there being a need to worry about pronunciation for the latter (T4: 55.669). He felt that students could “pick up” pronunciation later, “being immersed in the language” (T4: 3.69).

Even though the other teachers were generally positive about speaking, they voiced some concerns about specific aspects of the test. T6, who was initially quite enthusiastic about speaking, was disappointed when she learned what the tasks looked like. She felt they were “too brief and too unspecific” (T6: 113.130). She was especially concerned about the independent tasks, which did not, in her mind, yield enough information to allow her to judge how students

would perform at an English-speaking university (T6: 113.138). She compared the TOEFL speaking tasks with those of another major English as a foreign language (EFL) examination:

I would like to see the Speaking section done as it is in the Certificate of Advanced English exam, where test-takers speak together and with a live scorer rather than to a computer. (T6: 119.223)

T5 was concerned about the time limits placed on the students in all the tasks. She thought that “speaking within a time frame” would be “a bit awkward” for some students:

. . . some will not round up; some will stop short before time. Filling out the time is challenging. (T5: 56.74)

She was planning to advise her students to use a timer so that they would be more conscious of how long 40 seconds was (T5: 118.74). T6 believed that the speaking samples she listened to in Task 4 deserved higher scores than the narrator said they should get, because the speakers were working under serious time pressure:

I don’t think even some native speakers would be able to formulate a response in 20 seconds or so that would live up to the rubrics’ ideas of a good talk. (T6: 119.34)

Finally, there were some concerns about the mechanical side of the Speaking section. T5, having read the description of the task on the TOEFL Web site, noted that “very few people feel comfortable talking to a microphone” (T5: 93.245). T1 got the impression early on that the test would be telephone-mediated (T1: 6.19). This may have been due to information she had read about PhonePass technology being used during the development process.

There were a number of comments about the scoring rubrics for writing. One teacher was pleased with their explicitness:

. . . students have to be aware that they need to be almost fluent in English if they want to pass the exam, and they should not have any illusions and they should not underestimate the high standards of the new TOEFL. (T2: 97.142)

Other teachers also made positive comments, but with qualifications such as “some bits not clear” (T5: 112.41) and “more detail would be a help, though I don’t think the rubrics are useless now” (T6: 119.210). T6 was trying to overcome a natural tendency to “make a

mathematical formula (i.e., a percentage-based system) out of something that can't be limited to this method" (T6: 119.208).

There were a number of queries about weighting. T5 found it difficult to believe that there was equal weighting for all the criteria (T5: 118.20) and seemed particularly interested in how students' pronunciation affected the score they got for delivery (T5: 118.25). Two other teachers thought pronunciation to be quite important (T2: 53.132, T2: 65.159, T6: 113.13, and T6: 113.40), while in reality it is only one of several features that are judged within the delivery section of the rubrics, itself only one of four criteria used to score speaking. The teachers seemed to have merged their own opinions as to what should be weighted most heavily with what was actually specified. It may have been perplexing for them to listen to the speaking samples on the TOEFL Web site and to be faced with very different pronunciation issues from those they were used to when working with their own students:

I suppose that my students will notice the different pronunciation of the speakers in the sample tasks in the new TOEFL, and it might give them a wrong impression that the pronunciation is not that important as a feature that is scored. (T2: 115.164)

T6 was interested in the importance of accent as opposed to pronunciation (T6: 119.79). She was particularly concerned about how close a student's accent had to be to an American standard to be accepted:

Think of the word *renaissance*, for example. In the U.S., we put the accent on the first syllable. In Britain, from what I've heard, the accent is on the second. If an American heard a Brit say that, he would think it was mispronounced and vice-versa. But was it? Or was it simply accent and generally accepted speech patterns? (T6: 119.68)

There were also queries about the notion of topic development. T1 wondered whether this criterion should have the same weighting in the integrated task as it did in the independent task, given that the students received information to work with in the former. She said that her impression was that the weighting should be different in the two sorts of tasks. She hastened to emphasize that this was an impression rather than an opinion as she needed more exposure to and experience with the tasks (T1: 114.70). T6 was quite firm in her belief that for the integrated

tasks “content and how well the question is answered should be 80% of the final score” (T6: 119.104).

Independent Tasks

There were fewer comments about independent speaking than there were about integrated speaking. We have already mentioned T6’s disappointment that the tasks did not require extensive speaking (T6: 113.149). T3 felt that it could be difficult for students to speak if they did not know much about a topic or had never thought about it (T3: 116.183). This did not mean that she did not agree with the idea of including such tasks, however. She felt, in fact, that this made the speaking test more authentic. While she went along with the idea of students being given lists of topics to prepare for the writing test (e.g., the list of TOEFL CBT topics on the TOEFL Web site), she did not think this was appropriate for speaking:

. . . for the speaking, it would be inappropriate for the students to think about the topics in advance. When they go to university or college, they will need to be able to speak without having the chance to prepare for that in advance, while for the writing students are always given a chance to write the essays or papers at home. (T3: 116. 210)

Integrated Tasks

There was initially some confusion about the input for these tasks (T1: 52.40, T2: 53.19, and T6: 57.18), but this disappeared by the time the teachers completed our Task 4 (April), which asked them to pay special attention to how the speaking test functioned.

There were also varying understandings of what the integrated tasks were testing. The ability to summarize was mentioned by most of the teachers in response to our Task 2, but there was little consensus otherwise. T1 suggested only “topic development, delivery, and language use,” which, while not incorrect, did not show a detailed understanding of the criteria (T1: 52.19). The terms *clearly* and *fluently* were used repeatedly by T2 (T2: 53.119, T2: 53.125, T2: 53.127, and T2: 53.153), and also, though less so, by T3 (T3: 54.46). Fluency was suggested by T5 (T5: 56.67), which was interesting as this term does not actually appear in the scoring rubric. *Fluidity* appears in Bands 3 and 4, and terms such as *choppy rhythm*, *fragmented*, and *telegraphic* are used elsewhere. These terms seem to have been subsumed under a concept of fluency, whether the test designers intended this or not.

The teachers seemed more informed in their understanding of the criteria by Task 4 (April). One part of the task asked them what they would say to students who asked how they would be assessed on the Speaking section. One of the teachers was able to summarize the main messages from the scoring rubric:

Independent Tasks? You should speak clearly, without much hesitation and pauses. You should try to answer the question in details and with as much variety of vocabulary and grammar as possible. In order to do that, you should make a plan of what to say in advance.

Integrated Tasks? You should answer the question in details having in mind the information you read and/or listened to. Do not make long pauses; plan in advance what you are going to say. Use a variety of speech patterns and grammatical structures. (T3: 110.41)

We have already noted that there was a range of responses when the teachers were asked about their confidence in using the scoring rubrics. The responses ranged from 1 to nearly 3 on a scale from 1 to 3. T4, who placed himself at 1, said that he could give a judgement but he did not know the standard (T4: 111.28). T3, who placed herself close to 3, said,

The scoring standard is pretty clear but it is my first attempt at doing it. I am not saying that I am absolutely confident, because I have not done it many times. It will take some times and practice before I can say I am very confident. (T3: 110.27)

In fact, only one of the teachers (T1) scored the speaking samples in Task 4 in line with the TOEFL expert, and she matched the official judgement only two times out of three. Three of the teachers were more severe than the expert, one (T4) scored considerably higher, and one scored higher at times and lower at times. This suggests that T3 was correct in saying that it would take some time and further practice before the teachers understood the criteria and could use it correctly (T3: 85.25). This matches the sentiment expressed by T6 near the end of the data-collection period.

I wasn't sure what kinds of questions would be asked and what the time constraints would be. Now I know the basic information and can tell my students. Since I have the

rubrics, I have more to go on as far as scoring goes, though I by no means feel completely comfortable with it. (T6: 119.192)

Summary

The teachers generally seemed able to talk about the surface features of the new TOEFL, but they rarely mentioned the constructs lying underneath the new question types and formats. They seemed to understand information that fit their personal representation of the purpose and structure of the current test, but they showed limited evidence of awareness of what was being tested beyond basic subskills such as scanning, ability to paraphrase, and so on. There were instances in the data, however, of questioning in this area, even if no concrete conclusions were reached. T4, for example, asked several times whether it was language ability that was being tested in the new TOEFL or intelligence or education (T4: 61.21). T5 reflected on the role of interpretation in the integrated tasks, as opposed to comprehension (T5: 8.77). On the whole, though, there was not much deep thinking taking place during the data collection period, perhaps because the teachers had not yet had time to reflect on what they were learning. This is what T3 seemed to be indicating when she said the following:

I understand all the bits, but I need to understand all of them in more depth. . . It is one thing to know all about the sections (and another) to teach them after that. I will need to do a few tests myself, see all the problems that a student might face, and think of a solution. (T3: 127.157)

It was also the case, though, that some teachers were waiting for test preparation materials to appear before they thought too hard about what they would be teaching (e.g., T1: 40.181, T2: 121.54).

This point will be discussed further in the Factors Within the Innovation section in this report, under the heading Form.

It was clear that we had as many interpretations of the nature of the new test as we had informants. All six teachers based their descriptions on the same body of information (the TOEFL Web site), but they took away different interpretations, such as different versions of the importance of pronunciation or the role of note taking. Let us recall T2, who told us:

My only concern is that students have to be aware that they need to be almost fluent in English if they want to pass the exam and they should not have any illusions and they should not underestimate the high standards of the new TOEFL. (T2: 97.142)

Neither the importance of fluency nor whether the test is, overall, more difficult than the present test were discussed on the Web site. This was the teacher's own personal construct. It is important to monitor whether these constructs change as the launch date for the new test grows nearer, and whether they cause the teachers to teach in ways that match the vision of the advisers on the earliest versions of the new TOEFL.

Teachers' Plans for the Future

The aim of Phase 2 was to investigate how the teachers were coping with change. The first step in this process was to analyze what the teachers knew about the new TOEFL test and how they reacted to what they understood. The next step was to analyze what they said about how the changes in the test might affect their TOEFL preparation courses in the future.

We realized that what we were exploring was only the teachers' preliminary thinking and that some of the ideas they told us about might not actually work out in practice. We felt, however, that any plans they mentioned would give us further insights into their understanding of the nature and requirements of the new test. This was important, as understanding the nature of an educational innovation is the first link in a chain of events that leads to changes in the classroom (Chapman & Snyder, 2000). If the teachers' preliminary plans made sense in the light of the changes they knew about, then there was some hope that the test might have a positive impact on their teaching. This is what Messick (1996) referred to as an *evidential link*. The true impact, of course, will be determined not only by the teachers' plans at this early stage but also by other factors in their teaching context. These will be explored in the next two sections, titled Characteristics of Communication and Other Factors Facilitating or Hindering Change.

Timetable for Change

We began our data collection in January 2005, 9 months before the new TOEFL was due to be launched worldwide. We tried to establish when the teachers envisaged offering courses to prepare students for the new test. Some seemed not to have thought about this yet, while others responded that they would probably start offering new courses in May or June. This meant that

they would have to begin their planning fairly soon. We expected that they would already be searching for information about the new test when our study began, but in fact there was little TOEFL-related development work going on in the institutions at that time. Only one institution had really begun planning, and this took the form of searching for commercial materials (from which they would later take ideas and extracts to design their own in-house book) and wondering about whether new equipment was needed for the teaching of speaking (T1: 6.57 and T1: 40.184). Two other institutions were also discussing how to teach speaking. One still did not know whether the Speaking section would be compulsory (T4: 3.45 and 37.09); the other was expanding the teaching of this skill (T2: 126.161). A native speaker of English had just been invited to join the staff of the latter institution. Two other teachers in the sample were beginning to ask themselves how they might teach speaking, but they were not yet discussing this with their colleagues.

The teachers reported little progress in planning in February and March. Most of them had consulted the TOEFL Web site to do our Task 2 in February, but they did not report learning much that was new about the test during the 2 months that followed. Several reported that they would begin planning when they got hold of new materials, and one teacher reported that if her institution could not get hold of new materials, it might postpone offering new courses until materials were available. The institutions that were actively looking for materials (not all of them were) were still searching in May, when our data collection came to an end. Four of the teachers reported hearing about the test launch delay by May (T1, T2, T3, and T5). The pressure was off then, and several institutions put their planning on hold (T1: 71.39, T3: 73.88, and T5: 75.90).

There was quite a bit of uncertainty about when the launch might actually take place. T3 thought it would be November 1, 2005, in her country (T3: 59.46). T1 thought that it might be in 2006:

T1: Well, this date is mentioned as the earliest one possible but not *the* date. And, as I said, I spoke to my director, and he expressed doubts about it and said that it will probably be an even later date.

Interviewer: Were you thinking of later in 2006 or even later?

T1: Even later is also possible. We are used to the fact that in our country a lot of things are introduced, implemented, or started much later than in the rest of the world. (T1: 96.69)

T5 thought there might be delay of two years (T5: 75.90). A delay was not necessarily seen in negative terms, however. T3, for example, believed that her institution would have more time to plan, and it would be able to design a better course as a result (T3: 98.87).

Plans for Teaching Reading

We mentioned in the Reading section in this report that the Reading section of the TOEFL was the section that teachers saw as changing the least. It is not surprising then that they had little to say concerning changes in how they might teach this skill in the future. Only three teachers mentioned reading when they spoke about planning. T4, for example, said only that he would get his group to practice new question types (T4: 61.32). T6 said that she would focus on in-depth reading rather than simply searching for the answers to the questions. She thought this was appropriate since the new item types were “more analytical than simple multiple choice” (T6: 63.93 and T6: 68.35). T2 referred to the type of reading students would have to do for the integrated tasks rather than the reading in the Reading section. She believed that it would not be difficult because “students can read the passage and take notes, and it appears twice on the screen”⁶ (T2: 90.72).

Plans for Teaching Listening

The change that was most commonly mentioned when teachers discussed the Listening section was that students would now be allowed to take notes. All six teachers mentioned this, even if they did not go into detail about how they intended to incorporate note taking into their classes (e.g., T2: 90.95 and T3: 127.111). T5 brought up the topic more than the other teachers, perhaps because she had experience teaching EAP classes at university level and was more familiar with some of the issues involved. She was planning a note taking project, where students would listen to the same lecture three times and take better notes each time (T5: 118.137). She had gone as far as searching for materials on the Internet by May (T5: 124.48).

A second change the teachers mentioned was that the input passages would generally be longer. T1, T3, and T6 said they would focus on longer lectures in the future (T1: 58.37, T3:

59.73, and T6: 63.33), and T3 and T6 would search for these on the Internet. T5 felt that students would have to build up their listening stamina:

. . . listeners have to be in the habit of listening to longer presentations in English, not just in the habit of exchanging conversation remarks. People get tired after long concentration on the information in a foreign language. Even with note taking, it is more tiring than in one's native language. (T5: 118.193)

T4 said that he would help students to deal with the new items types, but he did not mention in what ways (T4: 61.32).

The only other comments made about listening had to do with the listening needed for the integrated tasks. T2 thought listening to lectures would be “a bit tricky” (T2: 90.74). She noted that students would need to compare information from different sources:

Their previous practice was focused listening—answering questions. . . but not connecting ideas from both listening and writing, and listening and speaking. So I think this is what they have to practice more. (T2: 115.118)

T2 was also planning to direct their attention to useful phrases such as *in fact* and *new findings* [sic] to help them to identify contradictory claims in the passages (T2: 90.76).

Plans for Teaching Writing

There was little discussion of plans for preparing students for the independent writing task. As we believe was the case with reading, the lack of discussion could be due to the teachers' perception that this part of the test seemed to have changed very little. T1 mentioned that she would probably teach as she did for the current test, but she had questions about whether there would still be three topic types (T1: 89.189). T2 reported that she would continue to give this kind of writing task as homework (T2: 90.248). This would allow her to devote her limited class time to the integrated writing task.

The teachers mentioned a number of ideas for preparing students for the integrated writing task. T1 said that she would teach them

. . . which information is worth noting down, how to summarize properly, what is important to be mentioned as key points, how to paraphrase, how to use most effectively the reading passage that is visible all the time. (T1: 89.166)

T2 and T3 said that they would concentrate on getting students to compare the different inputs to this task, while T5 would warn students not to plagiarize (T2: 90.22, T3: 91.69, and T5: 93.135).

Most of teachers' ideas had to do with the content of their classes. They did not go into detail about what they would actually do in the classroom. T6, who was very interested in teaching writing skills, offered the only concrete ideas about methodology:

I would have the students pick an academic topic of interest to them, choose a reading portion, and deliver a lecture so that the other students can write about it and answer the question. (T6: 113.95)

She elaborated on this idea as time went on and produced a similar idea for preparing students for the integrated speaking tasks. She recognized, however, that her vision might not be very practical in the future:

This is just an idea, of course, conceived in the ivory tower, and it may or may not work in the trenches, but I'm willing to give it a shot. (T6: 76.198)

T4 mentioned that he would definitely not ask students to write in class, but his courses during this period were general courses with an element of test preparation and not TOEFL-dedicated. He said that writing took up "conversation and important language use/instruction time" and that he would only spend time on writing if the course was "a purely NGT⁷ course" (T4: 128.200).

Several of the teachers mentioned the importance of students being familiar with the scoring rubrics for writing (T3: 90.238, T4: 92.92, T5: 93.99, and T6: 94.139). Peer review was mooted as a possible activity by T5 and T6 (T5: 112.75 and T6: 94.90). Two teachers mentioned that they would first need support materials in the form of scored samples so that they could understand the standard of the writing test themselves (T4: 92.93 and T6: 88.29). Only then, they felt, could they pass it on to their students.

Plans for Teaching Speaking

All six teachers had begun thinking about teaching speaking skills by the time the data-gathering period was finished. Most were planning to alter the balance of skills they taught so that there would be more time for developing speaking. Many of their ideas were still quite general. These included finding a way of helping students to overcome their fear of the Speaking section (T2: 126.89) and provoking students to speak in class (T3: 66.174). There were several references to getting students to think about timing (T1: 108.95, T2: 114.150, and T5: 118.130). Other general ideas included getting students to avoid personal bad habits when speaking, getting them to record themselves so that they could analyze their own performance (T1: 114.150 and T6: 113.89), and getting them to practice mutual evaluation (T1: 108.100). There were several references to technology being used, including microphones, headphones and recording software. The most concrete activities were again put forward by T6. She proposed the same sort of activity she had talked about for integrated writing (students choosing reading passages and delivering lectures to each other), but she also talked about recycling material developed for the current TOEFL:

I would also take old TOEFL materials, like the longer conversations and short talks, and put a reading together with them and ask the students to record each other answering questions related to the topic. (T6: 113.99)

T4 also mentioned some specific activities to develop speaking skills, but these did not represent a change for him since he was already concentrating on speaking in class, via debate and presentations (as seen in the Phase 1 findings). As mentioned in the Writing section above, his classes were not generally as test-oriented as those of the other teachers. His focus on speaking was not because it was now a feature of the TOEFL (as was the case for teachers like T1 and T2 [T1: 16.70 and T2: 115.226]), but because he had long felt it important for his students to develop this skill. This is ironic, however, as it was T4 who questioned most strongly the decision to include speaking in the new TOEFL.

It is clear that we have a more detailed picture of how the teachers were planning to deal with the writing and speaking components. This was to be expected because these skills were the focus of two of the tasks we set (Tasks 3 and 4) and they were novel and probably more thought-

provoking. The distance from current to projected future practice was also potentially greater. This can be seen clearly with T1, who first described her current classes:

In my current TOEFL classes, I do not include activities for developing speaking skills because the students who are going to take the current TOEFL know there is no speaking included and they do not want to have speaking practice. (T1: 126.158)

T1 then mentioned at least eight different ideas for dealing with speaking in the future. Among these was the notion of reducing class size to only three students (T1: 108.100).

Plans for Teaching Grammar

One question that teachers had differing views on was whether grammar should be taught in the future, given that there was no longer going to be a separate grammar section on the TOEFL. Some teachers seemed to be planning their new courses so that these would map onto the sections of the new test, much as they organized their current TOEFL classes. This would mean downgrading the role of grammar. T2 felt that if the students did not have the required level of grammar knowledge they should study elsewhere before commencing a TOEFL course (T2: 59.54). T6 said that she would be doing much more speaking in her classes and much less grammar, which indicated a significant shift in her course content (T6: 63.22). T3, on the other hand, felt that it was important to continue giving attention to language form:

We need to reinforce their knowledge, because even if there is no grammar section in the test they still need grammar for some of the other sections. (T3: 127.247)

T3 said that she would still include some explicit teaching and some revision where needed (T3: 127.233), although her main focus would have to be on the four skills (T3: 60.52). T5 stated that grammar, in the form of revision of typical mistakes, “will have to remain as part of preparation for writing” (T5: 129.195).

Planning More Generally

Thinking more generally, several teachers said that they would have to change the content of their courses as there was now a lot more to cover (e.g., T1: 89.166). T1 also suggested that groups should be smaller, although as we will see in our discussion of who makes the decisions in the institutions (see the Other Factors Facilitating or Hindering Change section in

this report), financial considerations are as likely to dictate such matters as pedagogical considerations. All six teachers mentioned that the main change would be the incorporation of integrated skills work in the curriculum (for example, T2: 59.40, T3: 91.59, and T6: 94.90), although, as we have seen, only T6 gave details of how this was to be achieved (e.g., T6: 68.183, T6: 76.186, and T6: 113. 89).

Few of the teachers queried whether the changes they thought they had to make would be beneficial. T6 was an exception. She mentioned that note taking was quite new to the students and she wondered whether it would help or hinder them (T6: 63.43).

Course Material

Another area that most of the teachers discussed was the role of the coursebook as they decided on the content of their future teaching. Several indicated that coursebooks would shape their courses, as generally happens at present (see Phase 1 findings; T1: 125.207, T2: 71.39, T3: 98.97, and T3: 127.79).

Some researchers have looked critically at the influence that commercial test preparation books can exert on teachers. Andrews (1994), for example, found that “those teachers making extensive use of their own material are very much in the minority” (p. 78), and he asked, “How far do the responses of the teachers (to his questionnaire) reflect thinking that has been conditioned by the textbooks they used?” (p. 79). Hilke and Wadden (1997) claimed that the 10 TOEFL preparation books they analyzed varied considerably in their accuracy in representing the test. This suggests that the teachers in our sample may be basing their plans on less than ideal foundations.

The teachers also mentioned using other sources of materials, however, including the Internet (the ETS Web site and others—T2: 59.40, T2: 109.69, and T3: 91.178). They searched for or were planning to search for sample tasks and mini lectures (T2: 59.72), and scored writing samples that included the rationale for the scores given (T3: 91.120 and T4: 92.93).

State of Preparedness

How prepared were the teachers? Preparedness may be assessed in terms of ability or confidence to handle the task in hand. It is difficult to evaluate these teachers’ preparedness for several reasons. The first has to do with the delay in the test launch date. It seems that the teachers took a pragmatic stance and decided not to spend time preparing materials when they

were not sure when they would be needed. The second reason has to do with the teachers' assumption, based on their previous experience, that commercial materials would eventually become available. Two teachers said that they would be able to produce materials themselves if commercial materials did not arrive on time (T3: 116.159 and T5: 129.110), and T6 seemed quite excited about the prospect of trying something new and creative (T6: 63.72 and T6: 94.240). Others, however, were less confident (T1: 120.169 and T2: 64.28) and most, whether confident or not, seemed to prefer to wait for commercial materials to appear before beginning their planning. As we saw in Phase 1, the teachers tended to accept without question that these coursebooks were appropriate for their purposes. T4 was the only exception, but we have already seen that his aim in teaching was different from those of the other teachers.

To summarize this section, changes in the TOEFL seemed to affect the teachers' plans for their future test preparation classes. The changes that they discussed seemed to be mainly related to content. What we cannot know from this data, both because it was from such an early stage in the transition period and because it is self-report data rather than observational data, is whether there will be any significant changes in the teachers' methodology. Previous washback studies would suggest that this will not happen (e.g., Andrews, 1994; Cheng, 1999; Wall, 2005). It is important to continue tracking the teachers as the date of the test launch draws nearer, to see whether there will be any changes in the way they manage their classes.

Characteristics of Communication

It is important in any impact study to analyze the channels of communication through which messages flow from the source of the innovation (the originators, in this case ETS) to the receivers (in this study, the teachers of TOEFL preparation courses). This is illustrated clearly in the Process section of the Henrichsen (1989) model (see Figure 1). If the channels of communication are not well chosen and the message being transmitted is either not clear to begin with or it gets distorted in the process of transmission, then there is little chance that the receivers will gain the awareness that they need to be able to react as the originators of the innovation intended. We saw in Phase 1 that the experts advising on the design of the new TOEFL hoped the test would have positive washback, or impact, on classroom teaching. Was the ETS communication about the test clear enough so that the teachers could see what to do to make their teaching more effective in the future? Were other messages being transmitted through other channels that either facilitated or hindered the teachers' understanding of the test and how they

needed to react to it? We discuss in this section the different channels of communication being used by the teachers during the period January to May 2005, and we comment on their effectiveness.

Types of Communication Channels

Rogers and Shoemaker (1971) classified communication channels as either “mass media” or “interpersonal” (p. 24). The channels found in this study are represented in Figure 3. The mass media channels included official ETS channels (the TOEFL Web site and other TOEFL products), non-ETS Web sites, agencies of various sorts, and commercial preparation materials. The interpersonal channels included the teachers’ managers, their colleagues, and their students. In the case of interpersonal channels, the nature of the social relationship will influence both whether the message is transmitted and the effect it has on the receiver (Rogers & Shoemaker, p. 24). This will be commented on in the Other Factors Facilitating or Hindering Change section in this report. A further channel of communication was this impact study. We have already commented on the influence our tracking questions and tasks would inevitably have on the teachers’ awareness of the new test (see the Methodology section in this report), but we will include further comments in the section titled The Impact Study.

What is not illustrated in Figure 3 is the interaction between the various communication channels. Whereas ETS, as the primary source of information about the new TOEFL, might wish its messages about the new test to flow through the official channels and reach the teachers directly, it is also possible for secondary sources to take up the messages and transmit them through their own channels. The messages could in fact pass between several intermediaries before reaching the teachers, with possibilities of distortion accumulating along the way. We reported an example of this in the Question 1—Have You Learned Anything New About the New TOEFL That You Didn’t Know Last Month? section in this report, where T2 may have picked up wrong information from talking to her students, who had received the information from their friends in other institutions. It is difficult to know where along the line of transmission the original information might have become distorted.

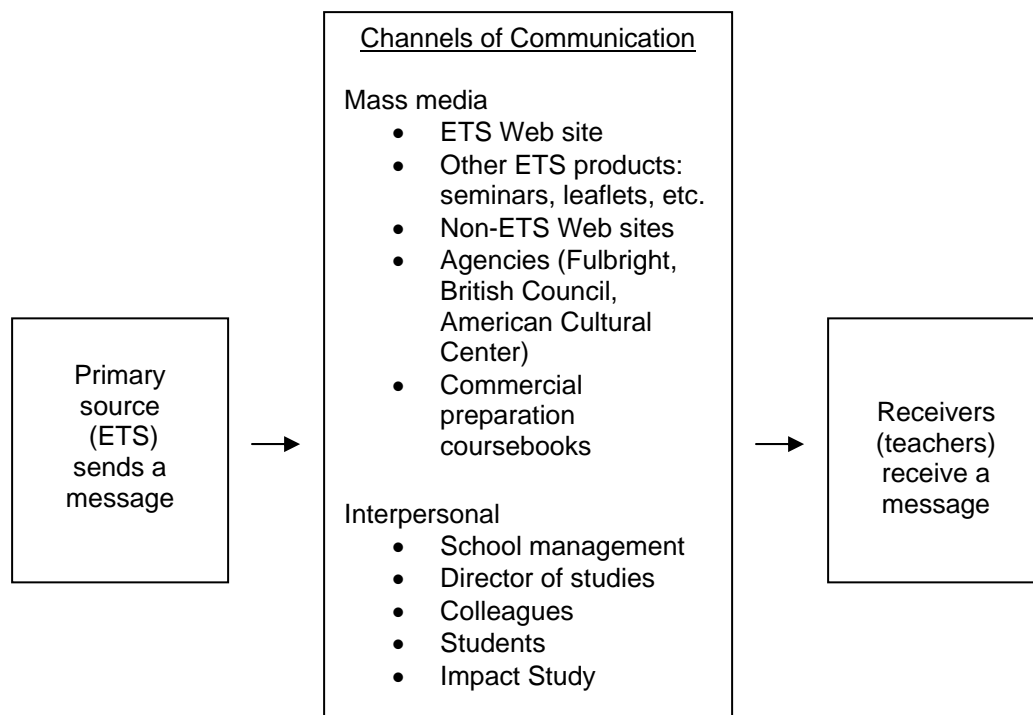


Figure 3. Channels of communication.

Mass Media Channels: ETS

The teachers made use of several mass media channels during the data-collection period, but it was the ETS Web site that they referred to most, both when trying to keep up to date with new TOEFL developments and when trying to complete the tasks we set them. Two teachers also had an ETS promotional CD called, *The Next Generation TOEFL Test Introductory Tour—Communicate With Confidence* (T1: 40.136, T5: 38.4, and T5: 44.27). T2 tried to access the online version of this tour, but owing to technical difficulties she was only partially successful (T2: 115.27). Some teachers knew about the ETS workshops available for teachers, though only T3 was able to attend one (T3: 127.195). None of the teachers mentioned seeing any ETS promotional leaflets such as *The Next Generation TOEFL Test—TOEFL iBT Timeline*,⁸ *Make the Connection*’ (ETS, 2005a), or *TOEFL iBT at a Glance* (ETS, 2005b), or using the downloadable versions of these materials. The officially endorsed ETS iBT TOEFL study guides (Beaumont, 2005; Solorzano, 2005) were not available during the data collection period.

Another potential source of communication between ETS and teachers was the discussion list available on the *TOEFL Practice Online* page of the ETS Web site

(<http://toeflpractice.ets.org/>). T5 mentioned that she had visited this list but was disappointed by the contents, since it seemed to be aimed at, and was indeed being used solely by, TOEFL candidates (T5: 44.180). As a result of this we asked other teachers how they would react to the idea of a discussion list for teachers. T4 thought this would be a good place for teachers to “post lesson ideas and ask questions” (T4: 123.195). T6 saw further possibilities:

I’m doing my first Internet course and it is going smoothly, so I believe I could have as much benefit from that as a standard seminar, and perhaps more if the seminar included ongoing discussion via discussions boards. (T6: 94.268)

She felt though that a short focused workshop with other teachers should still be valuable (T6: 94.260). T5 indicated that she would be as happy with an online discussion as with a face-to-face support session:

Face to face, you always get more information. Internet is good, too. I find it more conducive [sic] to reflection. (T5: 93.128)

Other Mass Media Channels

The teachers also used non-ETS Web sites as sources of information, as we saw in Phase 1 (T1: 64.169, T1: 71.117, T2: 41.14, T2: 47.96, T6: 21.31, and T6: 51.95). These sites were specifically aimed at teachers and students who were involved in high-stakes language test preparation. They resembled published coursebooks to some extent, offering information about TOEFL, exercises mirroring TOEFL test questions, and full practice tests. The teachers did not accept all the materials uncritically. T3, for example, complained about a Web site that offered writing samples supposedly scored in the way TOEFL scorers would score them:

I must admit, some of them are absurd and they give a very high score. . .

For example, there was an essay with the score 5 but it had awful grammar, inappropriate use of vocabulary, and inappropriate structure—no conclusion. I sometimes use that essay to show the students that they can be better than that. (T3: 30.88)

There was little information about the new TOEFL on these sites during the data collection period, but it was clear that the teachers would consult them in the future to supplement information they received from official sources.

Another mass media channel that figured prominently during Phase 1 and was clearly going to be important before the launch of the new test was commercial preparation coursebooks. Half the teachers turned to their coursebooks rather than the ETS Web site when they were asked to describe the current TOEFL (Task 2 in February). T2 contacted the local representative of one of the main English language teaching (ELT) publishers when trying to get information about the new test. She had told us in Phase 1 that the representative worked closely with her institution and was generally regarded as a good source of information. She reported this time though that he could not help her with descriptions of the test and even seemed unaware of the impending changes (T2: 121.78). We did our own search of publishers in order to find out what materials were available for the new test. We expected to find many publications, given that the test was due to be launched in a matter of months, but there were no preparation coursebooks available at that time. The teachers reported month after month that they were still waiting for materials to appear. They were still without coursebooks when the data collection ended in May 2005.

T2 also sought information from local educational agencies, but her search was unsuccessful. She mentioned the British Council as one such source but they did not supply any details. T4 also believed that the British Council could help students to understand which test was more suitable for their needs (T4: 43.73). The fact that the teachers believed a British organization could help with an American test showed that they were not aware of the different agencies' roles and functions. What was surprising, however, was that T2 could not get help from the local Fulbright office or the American Cultural Center, either (T2: 97.91 and T2: 41.120). She referred to such bodies on several occasions (see also T2: 121.20), so it seems she felt a need for more information than was supplied by the ETS Web site.

Interpersonal Channels

All six teachers reported that there were discussions about the new test taking place in their institutions, but these were not necessarily frequent or productive. Discussions in T3's institution really began only in March and very little discussion took place at all in T5's. The discussions that teachers did participate in would have been with the managers of the institution, the director of studies, or fellow teachers. They also had some discussions with students. We saw in Phase 1 that students supplied their teachers with considerable information about the current test, visiting the teachers after they took the test and recounting what they had experienced.

Students could not share information about the new test as they had not yet taken it, but they passed on news that they had received from their own sources and they asked questions that we presumed would prompt the teachers to try to find more information. T2 reported student questions throughout the data collection period, mainly relating to the Speaking section. T5's students had questions until March. This was around the time the news arrived about the delay in launch date. T1, T3, and T4 reported little to no student questioning. T1 tended to distance herself from the administrative side of the school (T1: 40.141), so if enquiries were being made she may not have been aware of them.

The Impact Study

We cannot ignore the role our project played in raising the teachers' awareness. Most teachers mentioned the study when we asked them about their sources of information (T1: 1.66, T2: 44.46, T3: 127.141, T4: 3.40, and T6: 76.18). While some teachers referred to information they had gained by being involved in Phase 1, others said that they had searched for information to complete the tasks in Phase 2. T3 said that she found the tasks "very useful" (T3: 70.7). T1 said that the tasks had forced her to think about the new test in detail (T1: 64.48 and T1: 120.109). Although the teachers might eventually have gone through some of the processes we asked them to go through (e.g., thinking about the differences between the tests, studying scoring criteria and practicing awarding scores), our tasks prompted them to do this earlier than they might have done and asked them to look at the new features in depth. T2 told us

If I hadn't done this kind of research and if I hadn't talked to you, I think I wouldn't feel prepared at all for my new classes. (T2: 126.250; see also T1: 120.64 and T3 127.142).

We were initially concerned about the influence our project might have on the teachers we were studying, but we argue that it would be difficult to carry out research into participants' awareness and attitudes without raising their consciousness in some way. As discussed in the Methodology section in this report, we have tried to minimize this influence in various ways.

Summary

The teachers made considerable use of ETS sources but we do not know whether they thought these were preferable to other sources or whether they were simply the most accessible sources at the time. The teachers used other mass media sources and interpersonal sources to

differing degrees. Their involvement in this research also played a role in their awareness-raising.

It is important to record whether commercial coursebooks take over as the strongest communication channel once they become available to the teachers. Such a reliance on coursebooks is not unusual. Read and Hayes (2003, p. 165) reported in their study of preparation for the International English Language Testing System (IELTS) in New Zealand that in 90% of the cases they looked at commercial exam preparation books were employed. Andrews (1994), Lam (1994), and Roberts (2002) also discussed how heavily such coursebooks are used in test preparation classes. These researchers carried out their research within the Asian context, but we believe that the phenomenon is also common in Central and Eastern Europe, given the results of Phase 1.

Spratt (2005) mentioned that the Lam and Andrews studies explored a situation shortly after a test reform, when it might have been expected that teachers needed extra support while they were becoming acquainted with test requirements. She wonders whether reliance on coursebooks continues once the test is familiar to teachers or whether the reliance is simply the “fruit of uncertainty” (2005, p. 11). Long-term impact studies are the only way of finding out for certain.

One of the most important factors in the diffusion of an innovation is a clear message about the aim and nature of the intended change, transmitted through efficient communication channels. We were interested in how clear the message was to the teachers in this study, which channels they chose to use to learn about the message, and which channels seemed the most effective. It seems that the messages concerning the mechanics of the new test (timing, number of questions, sections, item types, etc.) were relatively clear and these are the aspects of the test that the teachers had few problems grasping. What was not clear was whether the teachers would understand messages about the test construct, if these messages in fact reached them before the projected launch in mid-2006. We have seen in other parts of this report that the teachers were searching for materials that would give them a concrete representation of the sorts of things that would be tested in the new TOEFL and that some of them mentioned that they wanted officially scored sample responses to the writing and speaking tasks so that they could get a clear idea of the new test standards. There was not much material of this sort on the Web site during Phase 2, nor were there clear models of classroom practice that could show the teachers how they might

best prepare their students for the upcoming changes. The TOEFL Web site has expanded considerably since the last round of data gathering in May 2005, and at the time this report was written (December 2005), there are many coursebooks for the teachers to choose from. It will be one of the aims of Phase 3 to discover how great a role the Web site and the coursebooks play as the teachers begin to design their new classes and whether the influence will be beneficial or detrimental.⁹

Other Factors Facilitating or Hindering Change

Background

We saw in Phase 1 that the advisers who contributed to the design of the new TOEFL hoped that the test would produce positive washback on classroom practice. The general features they hoped to see included an emphasis on academic language and skills and a reduction in memorization and test-taking techniques as preparation methods (Wall & Horák, 2006). The specific features they hoped to see included the teaching of integrated skills, the teaching of speaking, the studying of longer and more complex reading passages, and approaches to the teaching of writing that developed discourse-level skills—paraphrasing and summarizing. However, only 4 of the 10 advisers we consulted said that they had been involved in discussions about *how* this positive washback would be achieved. Their responses generally showed an assumption that if the design of the test was right, then the impact that was desired would automatically follow. Only one of the advisers mentioned the need for test preparation materials (in the form of model tests), workshops for teachers, and information about the test development process that would be available to the public. Apart from this contribution, there was no mention of the mechanism needed to make change happen or to assist what in the literature is called the *diffusion of innovation* (Rogers, 1983).

We refer again to Henrichsen's (1989) model, which can be found in Figure 1. The Process section of the model shows the variety of factors that contribute to educational change. These include the channels of communication that transmit the message from the innovators to the receivers (discussed in the Characteristics of Communication section in this report), and a number of factors within the innovation itself (the new test), the resource system (ETS), and the user system (the teachers' educational context). Once the receivers (the teachers) have become aware of the innovation and formed attitudes towards it they will make a decision about whether to adopt it or not. In most studies of educational innovation this decision relates to the adoption

of a new approach to teaching, a curriculum, a set of methods or ideas about content. It could also relate to the adoption of a new approach to assessment or a new test or set of testing procedures. The situation we are studying is more complicated, however. Although the teachers in the study are learning about the new TOEFL, the important decision is not whether to adopt the TOEFL (the fact that they are teaching preparation classes means that their institutions have decided that giving TOEFL courses is financially viable) but rather whether to teach in a way that is consistent with what was intended by the original TOEFL advisers. (It must be remembered, however, that the teachers themselves are not aware of the advisers' intentions.) What we wish to establish is which factors will influence the teacher's future classroom practice. If their teaching includes the features that the TOEFL advisers intended, then which factors will have facilitated this? If it resembles the teaching we saw in Phase 1, with no integrated skills work, little speaking, and lots of working to the coursebook, then which factors will have hindered the appearance of positive washback?

The Henrichsen factors that seemed to be most important during Phase 2 came from the categories Within the Innovation and Within the User System.

Factors Within the Innovation

The three most important factors in this category were complexity, explicitness and form.

Complexity

Rogers and Shoemaker define complexity as "the degree to which an innovation is perceived as difficult to understand and use" (1971, p. 22). Fullan adds that "any change can be examined with regard to difficulty, skill required, and extent of alterations in beliefs, teaching strategies, and use of materials" (2001, p. 78). In the case of the new TOEFL complexity could be defined in terms of language difficulty (are the teachers themselves capable of achieving good results?) and teaching difficulty (do they have the knowledge and skills necessary to extrapolate from what they see in sample material and design lesson plans [content and methodology] that will develop essential skills in their students?).

With regard to language difficulty none of the teachers seemed to have problems with the demands of the new TOEFL. Two were native speakers of American English and the four local teachers all had a high level of English language proficiency. Only one teacher mentioned having difficulties with the test, saying that she sometimes could not find the correct answer to inference

questions. She was referring to the current TOEFL test, however, and furthermore, was talking about questions she found in preparation coursebooks rather than the test itself (T5: 129:211). Only one of the teachers had actually taken the current TOEFL and none of them had taken the new version, but we believe, from our communication with them in the Phase 1 interviews and the written conversations we had with them in Phase 2, that they were all probably capable of handling the language challenges that the new test would offer them.

With regard to teaching ability, all four of the local teachers had formal training and teaching qualifications, and one of the expatriates had recently completed a TEFL certificate course and was working on an MA in an aspect of language education. Although only one of the teachers (T4) ran classes that were in any way communicative in Phase 1, the teacher's responses to our Task 5 suggested that they were familiar with a number of communicative task types and used some of them in their non-TOEFL classes. The teachers seemed more open than we expected to the idea of using these task types in their future TOEFL classes. What we were not able to establish in this phase, however, was whether they were capable of doing so.

What was noticeable in this phase, as compared to Phase 1, was that none of the teachers seemed to doubt their technical ability to prepare students for an Internet-based TOEFL. Indeed, most of them sounded quite confident when discussing the types of software that might be useful for supporting their students in the future (T1: 64.40, T3: 7.51, T3: 42.81, T3: 110.61, T4: 92.31, T4: 99.115, T:4 123.23, T5: 11.69, and T5: 118.119). T4 even spoke of designing software to help students with their writing. Some of the Phase 1 teachers had lacked confidence in their ability to deal with computers. The fact that the Phase 2 teachers seemed more confident may be a function of our methodology, since everyone who volunteered to participate in Phase 2 had enough confidence in their technical ability to agree to communicate via e-mail or MSN Messenger.

When teachers talked about the complexity (or in their terms, *difficulty* or *confusion*) of the new TOEFL, it was mostly in relation to what to they perceived would be difficult for their students, not themselves. If they referred to difficulties for themselves it was mainly because they did not yet have enough information about some aspect of the test they considered important.

Explicitness

Explicitness refers to the clarity with which an innovation is described and to whether it is well worked out as a notion (Henrichsen, 1989, p. 84). According to Dow, Whitehead, and Wright (1984, cited in Henrichsen, 1989, p. 84) potential users of an innovation should understand its rationale, philosophy and specific goals and objectives.

A great deal has been written about the rationale and philosophy of the new TOEFL, including the framework documents, which explained early thinking behind the test and the search for constructs on which to base it. There have also been many ETS conference presentations outlining the goals and objectives of the new design. The important question is whether these ideas have been clearly communicated to the teachers. We did not ask the teachers about these issues in Phase 2 as it was apparent from the beginning that they had not read enough about the new test to be able to form a solid impression. We concentrated instead on investigating the explicitness of the official information regarding the test's content and format.

We felt this was important because we, ourselves, had found it difficult at the beginning of Phase 1 (January to March 2003) to piece together what the new TOEFL would look like. We wanted to find out as much as possible about the test as we were planning to incorporate some aspects of the design into our data collection instruments. We had to consult a number of sources as there was no single source that could give us all the details. There were some aspects that were not clear even at the end of Phase 1 (June 2004), and when we received feedback on our Phase 1 report in December 2004 we learned that there had been more developments in the test in the second half of the same year. One of our concerns throughout Phase 1 was whether ordinary TOEFL teachers, who did not have the resources that we had to check our understanding of the new test, would have the clarity they needed to be able to start planning their new courses.

There was much more information on the TOEFL Web site when we began collecting data for Phase 2 (January 2005). ETS had also made a number of other publications available, both in print form and via the Web. At the time of this writing (December 2005), there are about a dozen publications available. What is not known, and time constraints have prevented us from following this up in this phase, is whether the information in all these documents is exactly the same, or whether, as we found out in Phase 1, it is necessary to read several documents to get a full picture of the new test.

It was seen in the Characteristics of Communication section in this report, however, that the teachers in our sample relied mainly on the TOEFL Web site. It is important to analyze how explicit the information was that the teachers referred to at the time of data collection. An analysis of the descriptions they gave of the new test in Task 2 (February) suggested that the documents must have been explicit when it came to basic features such as number of sections, number of questions, and time limits as most of the teachers were able to report these accurately. It would not have been possible to check the teachers' understanding in every interview; however, we are able to report the types of questions they had each month and to see whether these suggest any lack of explicitness. The questions are presented in Table 7.

Table 7
Teacher's Questions About the New TOEFL

| Question | How many teachers asked? | | | | |
|---|--------------------------|------|-------|-------|-----|
| | Jan. | Feb. | March | April | May |
| Reading | | | | | |
| Will the items test synonyms or guessing the meaning of words in context? | | 1 | | | |
| Will wrong answers be penalized? | | | | | 1 |
| Listening | | | | | |
| Will there be non-native voices in the listening passages? | | | | 1 | |
| Will wrong answers be penalized? | | | | | 1 |
| Writing | | | | | |
| What are the criteria for scoring writing? | 1 | 1 | | | |
| Is my standard for scoring okay? | | | 1 | | |
| Are there still three basic topics for writing? | | | 1 | | |
| Why is there no 6 on the rating scales? | | | 1 | 1 | |
| Who scores writing—a human? | | | | 1 | |
| What's the required standard for writing for the United States? | | | | 1 | 1 |

(Table continues)

Table 7 (continued)

| Question | How many teachers asked? | | | | |
|---|--------------------------|------|-------|-------|-----|
| | Jan. | Feb. | March | April | May |
| Speaking | | | | | |
| What do the speaking tasks look like? | 2 | | | | |
| Is the Speaking section taken separately like TSE? | | | | 1 | |
| What are the criteria for scoring speaking? | 3 | 1 | | | |
| What's the weighting of the different criteria for speaking? | | | | 2 | |
| What's the weighting for pronunciation? | | | | 2 | 1 |
| What is the weighting for accent? | | | 1 | | |
| Will non-American pronunciation be acceptable? | | | | | 1 |
| What's the required standard for speaking for the United States? | | | | 1 | 2 |
| Is my standard of scoring okay? | | | 1 | | |
| What can such short tasks show about a person's speaking ability? | | | | 1 | |
| How do I teach students to do independent speaking (if it involves their own content and vocabulary)? | | | | 1 | |
| General | | | | | |
| Why are independent tasks necessary if integrated tasks test the same things? | | 1 | | | |
| What's the average performance like on integrated tasks? (teacher seeking a notion of standard) | | | 1 | | |
| Does the TOEFL test language or IQ? | | | 1 | | 1 |
| What's the standard necessary to get into U.S. universities? | | | | 1 | |
| Who should set the standard for U.S. universities? | | | | | 1 |
| What's the difference between TOEFL NGT and TOEFL iBT? | | | | 1 | 1 |
| Why has the TOEFL changed? | | | | | 1 |

(Table continues)

Table 7 (continued)

| Question | How many teachers asked? | | | | |
|---|--------------------------|------|-------|-------|-----|
| | Jan. | Feb. | March | April | May |
| Training | | | | | |
| Am I eligible to attend workshops? | 1 | | | | |
| How can we learn about other teachers' ideas for classes? | | 1 | | | |
| What do preparation classes look like in other countries? | | | 1 | | |
| Can you give me details of conferences? | | | | | 1 |
| Administrative | | | | | |
| How long is the test result valid for? | 1 | | | | |
| How can an Internet-based test be given in a country without the technical resources? | | | 1 | | |
| What will happen in my country before the new test is launched in 2006? | | | | 1 | |

Note. From tracking questions and Tasks 1-5, January-May 2005.

It can be seen from the table that the teachers' questions were varied, covering all skill areas as well as training and administrative matters. The fact that they had questions in the beginning does not mean that the descriptions they read were not explicit. We would judge that they were quite well written. The teachers may have had questions simply because they were trying to process too much new information at the same time (information overload) or because some of them were receiving messages from other sources and may not have known how to put everything all together in one coherent package. Most of the questions they had at the beginning of the study disappeared by March. The questions that remained in May related to the weighting of criteria for the scoring of speaking (especially pronunciation) and to standards. Although the teachers had practiced scoring writing and speaking performances as part of our Tasks 3 and 4, they did know whether they had scored them correctly or how good a student's performance had to be in order to be acceptable to higher education institutions in the United States. These would seem to be two areas where more explicitness would be useful.

Form

Although the teachers were fairly knowledgeable about the test by June, they were not able to present many concrete ideas about how they would conduct their classes in the future. They were asked several times how the changes in the TOEFL might affect their practice, but most of their responses had to do with the content of their classes—the skills they would concentrate on, the time they would spend on certain skills as opposed to others—rather than the types of activities they would introduce and the way they would manage their classrooms. As we have mentioned so often, the teachers were waiting until commercial preparation materials arrived in their institutions before they committed themselves to thinking about their future teaching.

It is not unusual that teachers only start appreciating what an educational change implies when it has been translated into materials. Henrichsen cites Richards' (1984) claim that “methods that lead to texts have a much higher adoption and survival rate than those which do not” (Henrichsen, 1989, p. 85). The teachers made many comments about the importance of obtaining materials before they could start their general planning (T1: 6.83, T1: 125.210, T2: 59.164, T2: 109.67, T3: 98.120, T3: 116.191, and T5: 100.49) and before they could make plans for each skill area. In some cases, what they wanted were passages that they could work with. T6 said she would have to search the Internet for reading passages if she could not obtain published materials, but “I wouldn't know where to begin for listening” (T6: 94.245). T5 was already searching for materials that would help her to teach note taking (T5: 93.164). Most of the teachers wanted scored samples of writing, either so that they could familiarize themselves with TOEFL scoring standards (T2: 84.31 and T4: 92.36) or so that they could help their students to understand the standards (T1: 71.117, T3: 91.124, and T4: 92.98). T5, however, wanted to see sample questions more than sample responses so that she could “recommend patterns of structure to use in writing” (T5: 93.111).

Materials for integrated tasks were in demand, with T3 asking for as many examples as possible of reading and listening tasks on the same topic accompanied by related questions (T3: 91.81). T2 was interested in

...more sample integrated tasks in which the students' responses are good concerning grammar, structure, and spelling but fail to connect the points made in the lecture and reading (T2: 84.34).

T1 and T5 expressed their desire to see more sample speaking tasks (T1: 40.97 and T5: 112.68).

Most teachers seemed to prefer printed materials to online or other computer-based materials, despite the fact that the test itself was to be delivered via the Internet. One factor influencing this was probably portability (T2: 90.194). Another may have been cost. T2 was concerned that there were not enough free samples on the TOEFL Web site and that students (and presumably teachers?) would have to pay what for them was a lot of money to get access to more materials.

Factors Within the User System

The most important factors in this category were classroom factors, institutional factors, and teacher factors.

Classroom Factors

The classroom factors that were most likely to have an impact on the type of teaching teachers do in the future were time and space.

Time. The courses offered by the six teachers varied in length from 36 to 80 hours (a teaching hour was only 45 minutes in some institutions), spread out over a period of from 1 month to a year. Most of the teachers mentioned that there was too little time available to deal with all the aspects of language they would like to deal with or to include all the types of activities they would like to include. The new TOEFL was going to increase demands on time as the teachers would also have to fit speaking and integrated skills into their syllabuses. Their plans for coping included spending less time on grammar (T2: 29.31 and T3: 122.75), assigning the reading part of integrated tasks for homework (T6: 68.186), and not using some of the more communicative task types we asked them about in Task 5 (T4: 128.226).

Space. In one institution, the director of studies was considering running a single course for students preparing for the current TOEFL and those preparing for the new TOEFL, even though the teacher who would be running the course thought this made little pedagogical sense. One of the director of studies' justifications was that there were not enough classrooms to hold two parallel classes (the other was that there was only one TOEFL teacher, who also had other obligations—T1: 40.105). In at least one other institution, it was likely that students studying for

the new test would be working in standard classrooms rather than at computers, as there were not enough computers.

What I think is most likely is that students will be informed of the computer nature of the test but will be advised that if they want to practice for this particular test, they will have a traditional, paper-and-pencil classroom set-up and will be responsible for familiarizing themselves with the computer test in their own time. (T6: 9.33)

Institutional Factors

The institutional factors included type of institution, type of clients, management factors, and resourcing.

Type of institution. We expected that the type of institution that the teachers worked in might have some influence on the plans they were making for their future test preparation courses. Our sample included an institution that was a national education information center (T1), one that was also a Prometric testing center (T5), and one that was part of what we understood in Phase 1 to be a large and profitable chain of language schools (T6). We predicted that the education information center and the Prometric testing center would have access to more information about TOEFL than the other institutions and would therefore be more ahead in their planning. This was not the case during this data collection period. We also imagined that the institution that was part of a large chain of schools would benefit from opportunities to share ideas and experience across the organization. In fact, this institution and the chain of schools with which it was affiliated closed about two thirds of the way through the data collection period, apparently due to bankruptcy (we hope to continue to follow T6 as she seeks other teaching positions in the future, though this has already taken her to another country—Russia). T4 joined a new teaching institution, though he continued to work in his original institution as well. The new institution is an educational support organization that is more interested in developing its clients' individual potential than in offering standard test preparation courses, although it does not rule out the possibility of offering preparation courses in the future if there is demand. It was clear in Phase 1 that T4 was not enthusiastic about giving courses that focused solely on test preparation, so his move to the new institution seems logical. One of the advantages of the new institution is that it will allow him to offer classes in which learners “analyze their own styles, what helps them learn, and they will compare that with others” (T4: 31.114). It is important to

monitor whether this approach can prevail in classes where students are mainly interested in getting good test results.

Type of clients. We saw in Phase 1 that the clients for TOEFL preparation courses were mainly young professional people and university students who were interested in studying abroad. The numbers of students taking the TOEFL in these countries were not very great but the institutions in our sample were able to run classes with adequate numbers most of the time. Three of the countries represented in Phase 2 have now joined the European Union (EU). This means that students can study in other EU countries more easily and cheaply than before. This may affect their motivation to take the TOEFL in the future:

I have to wonder if, because TOEFL is changing and the new one may be harder to take, it will actually become *less* popular here. Already it's more common to take the Cambridge exams, and with this country now a part of the EU, more students may choose to study in Britain and Ireland and allow the Cambridge exams to suffice, especially if they have trouble finding a way to practice on the computer and have to travel to Berlin or Vienna to take the test. (T6: 101.185)

T3's country was not yet a member of the EU, but she claimed that the probability that they would be joining soon was already affecting the students' interest in the TOEFL.

There are now less students preparing for TOEFL. This is because a lot of students already went to study in the U.S.A. Now most of our students want to study in Europe since this country is expected to become a member of the EU. (T3: 121.179)

It is clear that in a situation where the number of potential clients is dwindling institutions will have to make major efforts to attract and retain the few that remain. One institution was preparing for clients from a completely new market: teenagers in state schools who realized that they were not learning enough in their regular classrooms to benefit from short preparation courses just before taking the TOEFL. The teenagers wanted a longer preparation course that they would start while they were still in school (T2: 41.182). It is important to monitor whether the age and level of maturity of these new students will affect the type of teaching that is offered in the future.

Management factors. There were two management factors that were influencing what the current TOEFL courses looked like and were likely to influence the new courses as well. These were the type of management structure that was in place in the institutions and the specific characteristics of the people who were working as managers or directors of studies.

We use the term *management structure* to refer generally to how responsibilities were divided up within the institutions and, more specifically, to whether the teachers were expected to implement decisions made by their managers or whether they were allowed to design courses the way that they wanted. We have already seen that T1 (whose director of studies was considering offering preparation for both versions of the TOEFL in the same course) perceived decision-making to be out of her hands (T1: 40.105). She was happy in most instances not to be involved in management issues, but in this particular case she felt that the wrong decision was being made and it annoyed her (T1: 40.89). T6, although approving of a decision to lengthen TOEFL courses from a semester to a year, did not have much confidence in the management of her institution (T6: 9.26). Her doubts proved correct when the institution and the chain it was part of closed with little warning (T6: 76.109). When asked what planning there had been for new TOEFL courses, she replied:

Towards the end, they were putting out fires rather than fireproofing by planning ahead for the new test. (T6: 76.150)

T2, T3, and T5 did not have much communication with their directors of study, although T3 had been asked to be involved in the coursebook selection process when her director of studies (and the owner of the institution) was too busy to manage this (see Phase 1 report). T5 was happy to leave the technical side of operations to her manager while she herself did the course planning (T5: 124.68). T2 would have preferred more contact with her director of studies, but he was too busy with his own duties (T2: 97.121). This was unfortunate as he seems to have been the only director of studies with EFL experience.

T4 made uncomplimentary remarks about the institution where he worked in Phase 1, claiming that its management gave no support to TOEFL teachers:

(They) basically have no requirements on teachers for TOEFL courses. I can't speak for other teachers, but I was only given a copy of a TOEFL book and a classroom. The rest was up to me. (T4: 31.52)

As a manager himself in his new institution, T4 was determined to ensure course quality, but it was clear that he would face tough decisions when trying to implement his educational philosophy in a context where space and time constraints also applied.

The fact that most of the directors of studies did not have EFL experience was important when it came to the transmission of information about the new test. T5, for example, reported that her director of studies (who was not an EFL professional) had not passed on information about integrated skills in the new TOEFL, concentrating instead on the testing of speaking. T5 implied that this was because her director of studies did not have the background to understand how significant the integrated skills development was (T5: 44.84). The fact that it is the directors of studies who attend the conferences rather than the teachers (T1: 96.29, T2: 97.121, and T5: 44.34) is potentially important. It is worth investigating in the future whether the directors always bring back the information that they should and whether they pass it on correctly.

Resourcing. In the main, the institutions did not have the financial resources to send teachers to conferences or courses (T2: 38.48, T2: 41.160, T2: 41.154, and T5 44.174). T3 said that she welcomed the idea of Internet-based training because it would save her institution time and money (T3: 122.201). There was also too little funding to pay for the teachers to sit through the new TOEFL (T3: 127.16 and T5: 44.109). The teachers' lack of first-hand experience with the test may affect their teaching in the future by making them even more dependent on commercial publications.

Two other aspects of resourcing were relevant during Phase 2: the institutions' technical arrangements and whether they had a library. The technical arrangements, in particular the availability of reliable computers and Internet connections, are likely to be very important given that the new TOEFL is now Internet-based and students will probably want to use practice material via computer. In our 5 months of contact with the teachers, there were various technical hitches, both during the time we were communicating with them (difficulties in connecting with the Internet and remaining connected) and when they were trying to access Web-based materials. T2 had serious problems both when trying to take the tour of new TOEFL on the ETS Web site (T2: 72.26) and when trying to access the online practice test that was the basis for our Tasks 3 and 4. She was not able to hear one of the sample speaking tests so we were not able to determine whether she could use the relevant scoring criteria appropriately (T2: 109.12 and T2: 115.48). She was worried that her future students would not be able to practice listening because

of problems with equipment (T2: 109.30). T6 repeatedly voiced her concern about how students in her country could prepare for the new test properly, given the poor level of technical equipment and Internet access (T6: 76.18, T6: 101.87, and T6: 130.181).

T6 also raised the issue of whether institutions had adequate libraries. This was important not just because students need to find materials but because teachers needed materials on which to base classes. The library at T6's institution was not adequate, so she was planning to find material in two different ways. The first was to access her former university's Web pages (she was a recent graduate from an American university and still had permission to do this), where she could find texts that her teachers had required their students to read as part of their assignments. The second was to use a resource that she and local colleagues had put together as part of a self-development effort:

My own colleagues have all created Web sites that we are linking together and they will have "authentic English" of their own that I can use for my classes. (T6: 119.163)

What is interesting here though is that T6's fall-back options require access to the Internet, which, as we have seen, she felt students could not take for granted.

Teacher factors. Henrichsen (1989) believed that teacher factors should be investigated during the antecedents phase of the process of innovation (under characteristics of the users), before the innovation is introduced into the educational context. Wall (1999) argued that it was important to continue investigating teacher factors and student factors during the process phase as well, given that these can change during the long period that is generally needed before an innovation finally gets adopted or rejected. The teacher factors that seemed most relevant in Phase 2 were motivation to learn about the new TOEFL, teaching knowledge, teaching experience, and confidence. Table 8 shows our impressions of these factors for all six teachers, as well as our judgement of how aware the teachers were of developments in the new TOEFL.

Table 8***Teacher Factors That May Influence Future TOEFL Practice***

| Teacher | Motivated to learn about the TOEFL | Teaching knowledge | Teaching experience | Confidence | Awareness of the new TOEFL |
|---------|------------------------------------|--|---|---|--|
| T1 | Very motivated | Sound. Familiar with most Task 5 activities. | 2 years Recently taken over TOEFL class. | Growing in confidence, but still not confident about designing materials. | Good awareness from start. No misunderstandings. |
| T2 | Very motivated | Seems sound. Familiar with most Task 5 activities. | 16 years Advanced and Business English. | Confident about designing materials, though she believed she was not authorized to do so. | Much parroting of TOEFL Web site at beginning, with some misunderstandings but these seemed to clear up later. |
| T3 | Very motivated | Seems sound. Familiar with most Task 5 activities. | 6 years. Serving as academic director, plus administrative and pastoral roles. | Fairly confident about designing materials but worried about time. | Good awareness from the start. No misunderstandings. |

(Table continues)

Table 8 (continued)

| Teacher | Motivated to learn about the TOEFL | Teaching knowledge | Teaching experience | Confidence | Awareness of the new TOEFL |
|---------|------------------------------------|---|---|--|--|
| T4 | Less motivated | Seems sound. Familiar with most Task 5 activities. | 16 years. Now managing and teaching in educational support organization. Not interested in teaching that is aimed solely at test preparation. | Confident in own understanding of U.S. universities (from United States) and own teaching ability. Beginning to design Web-based materials for various clients. | Less awareness than others all through the data collection period. Inaccurate understandings and some unsureness. Lack of attention to detail. |
| T5 | Very motivated | Sound. Working towards PhD—academic approach. Familiar with most Task 5 activities. | 24 years. Teaching EAP and other test preparation courses. | Confident in understanding of U.S. universities (had studied in United States) and own teaching ability. Confident about designing materials. | Good awareness though minor misunderstandings at start. Much querying of information and deep thinking about constructs. |
| T6 | Very motivated | Seems sound. Working towards MA—very reflective. | 2 years. General English and test preparation courses. | Confident in understanding of U.S. universities (from United States) and own teaching ability, especially in area of writing. Confident about designing materials. | Good awareness throughout, with many questions and much exploration of ideas from her recent MA work. |

As we stated earlier in the Complexity section in this report, the teachers in this sample generally seemed competent and confident when they spoke about the new TOEFL, though this was more evident at the end of the 5-month data collection period than at the beginning. The exception was T4, who often gave the impression that he had not spent as much time as we expected completing the tasks and whose responses were not as accurate as the other teachers'. This could have been a reflection of his general attitude to tests and test-preparation classes, which was also apparent in Phase 1. He regularly mentioned his dislike of courses with the sole aim of enhancing students' chances of acquiring a certificate, and he explained many times that his priorities were general educational development and helping learners to recognize and realize their own potential.

Summary

This study presents a number of factors relating to the test itself and to the educational environment that are likely to play important roles in the shaping of the impact of the new TOEFL. What is clear is that most of the teachers (T4 is a possible exception) were fairly aware of the content and format of the new test by June 2005 and were, in spite of the postponement of the launch date, beginning to think about the resources they needed to muster to design courses that would help their students with the new test demands. What will not be clear until later, however, is how all of these factors will interact and what their ultimate effect will be on teaching and learning. One of the major factors seems to be form—or materials that represent the contents of the new test and give teachers ideas about how to prepare their classes. It is important to analyze the messages transmitted by the official ETS materials and the way they are interpreted by publishers and retransmitted via commercial coursebooks. These analyses, combined with interviews and observations in the future, should allow us to determine whether the original intentions of the TOEFL advisers lead in the end to the desired outcomes in TOEFL preparation classes.

Discussion

This report has presented the findings of an investigation into six teachers' awareness of changes in the new TOEFL test, their reactions to what they understood of the changes, and their early plans to prepare future test preparation courses. We have analyzed the teachers' responses to five sets of tracking questions and five tasks focusing on different aspects of the new test and

TOEFL teaching and their comments in 10 interviews per person via computer-mediated communication.

We believe that the methodology used in our investigation was appropriate and effective for the purposes we had in mind. By working with a small sample of teachers who could devote 25 hours of their time to our questions and tasks, we were able to probe understanding and attitudes in a way that would not be possible in a large-scale survey. We accept that it is difficult to generalize from the findings of qualitative work such as ours but it was not our intention to make wide-ranging statements about all teachers in the region we were studying. What we hoped to do was produce a report that would provide interesting or useful insights to ETS and other test producers and that would further thinking about the factors that may facilitate teachers' (and as a result, students') appreciation of test demands and their application of what they know to their classroom practice.

The teachers' awareness of the changes in the TOEFL was quite low at the beginning of the study. All six knew that the new test would include a Speaking section and some were aware that there would be integrated Writing and Speaking sections, but they only had a general idea of what these sections would entail. Their awareness grew during the data-gathering period, due partly to the tasks that we set them, and at the end of the period they were aware of the main differences between the current test and the new one and of many of the surface features of the latter. They had gone through the process of scoring sample writing and speaking performances and consequently had some understanding of the criteria that would be used for scoring in the future, but they did not have a good grasp of how to use the criteria and would need more exposure to them and practice with them before they could give their students clear messages about the standards that the criteria represented. They did not pay much attention to the Reading and Listening sections during this time, perhaps because our tasks did not direct them to these sections but more likely because they did not see great differences between these sections and the corresponding sections on the current TOEFL. There were a few comments about the input passages and item types in the Reading section, and some awareness and enthusiasm for the fact that note taking would be allowed in the Listening section, but there was far less concern about or interest in the consequences of the changes to these sections than to the addition of integrated skills and, above all, of speaking to the overall test structure.

The teachers' reactions to the new test were mostly positive, especially towards the idea of testing speaking (although one teacher was against the idea of testing this skill and another did not see how brief speaking tasks could indicate anything significant about the type of speaking students would have to do in the target language situation). The integrated writing task was also received favorably, as was the idea that students would be able to take notes during the Listening section and not have to rely on their memory. The teachers felt that these innovations would lead to changes in their classes, but most of them could only envisage changes in general terms and were waiting for test-preparation materials to appear that would help them to decide on the details. They consulted the ETS Web site and had access to a few other ETS resources, but it was clear that they were counting on commercial publications to ease their burden of planning. There was interest in other forms of teacher support, but the institutions that the teachers worked for were generally not able or prepared to assume the costs of the ETS support that the teachers were aware of, whether this took the form of teacher training seminars in other countries or the online practice test that had to be paid for. The teachers themselves did not go to conferences and therefore would not receive detailed information about the tests in this way; they would instead have to depend on their directors of studies receiving and transmitting relevant information to them accurately. There had already been at least one instance of a teacher not hearing about an important change because her director of studies had not realized its importance.

The ETS Web site was clearly the most important official channel of communication about the new test, but the interesting question was going to be whether it would provide enough free information in the future for both teachers and students and whether it would give teachers the opportunity to practice scoring sample performances in a way that would increase their skill and confidence. It was inevitable that the teachers would also refer to the non-ETS Web sites that they mentioned so frequently in Phase 1. These had not begun to present much new-TOEFL-related information during the period of the study, but it was likely that when they began transmitting information the teachers would accept it fairly uncritically. (Discussion on L-TESTL, a major language testing discussion list, has recently highlighted how much information of dubious quality can be transmitted by these unofficial sources, a cause of concern to those convinced of the importance of the consequential aspect of validity.) It must be stressed again, however, that whatever appeared in the way of commercial preparation materials would provide a more powerful impact on teaching than the Web sites, as these would provide teachers with

ready-made solutions to the problem of finding materials. Whether these materials would accurately reflect the nature of the test or provide appropriate ideas about how to conduct test preparation classes is another question, and one we could not answer during the data-gathering period as no preparation materials had yet appeared in the Central and Eastern European region.

One of the points that emerged from our Phase 1 survey with ETS advisers was that they wanted test preparation in the future to be more communicative. The Phase 2 teachers were drawn to the idea of using more authentic materials and practicing skills that would be useful in the target language use situation, and most of them envisaged focusing more on speaking in the future. Quite a few of the teachers expressed an interest in incorporating more communicative task types in their teaching. If intentions to change in the desired direction were an indication of positive test washback, then the new test had begun having impact during this period. What remained to be seen, of course, was whether the teachers would develop the greater understanding we feel they needed to be able to select appropriate materials (would these exist by the time they were needed and would they be accessible?), whether they would receive the guidance necessary to present the materials in effective ways (would commercial publishers see this as part of their mission?), whether they would have the skills necessary to implement their nascent ideas (they spoke of communicative intentions but did they have the knowledge or attitudes to teach communicatively?), and whether their institutions would provide the support and facilities necessary to carry out a different sort of teaching.

When we designed the Phase 2 study, we thought that we would be gathering data right up to the time that the teachers would have to begin teaching their new courses. The delay in the launch date of the new test meant that the pressure that the teachers should have been feeling to prepare their classes was lessened. This, plus the fact that there were no commercial materials to work with, meant that we were not able to see during the time of the study how the teachers' preliminary thinking would be transformed into concrete plans. We learned near the end of the study that the new TOEFL would be introduced into this region in mid-2006 so the preparations we thought would be occurring during the latter stages of the Phase 2 data-gathering period would be taking place as this report was being written. We felt that it was necessary to continue the Phase 2 work so that we could capture and describe the changes that would have taken place since June 2005 and to investigate, in particular, how the appearance of new and fuller information on the ETS Web site and the appearance of commercial materials would affect the

teachers' plans and actual classes. Would their understanding of the constructs underlying the new test, the criteria used for scoring, the details of the format of all the sections, and their understanding of standards increase by the time the test was launched? Would they be able to figure out how to effectively develop their students' skills for the integrated sections and the testing of speaking? Would the steps they had made since June 2005 lead them toward the type of teaching envisaged by the advisers to the new TOEFL, or would their classes be off-target as a result of a dependence on information and materials from unofficial (or even official) sources? This would become the focus of Phase 3 of the Impact Study.

In the introduction to this report, we stated that among the main influences on our Phase 2 work was Messick's notion of consequential validity (1996). It has been our intention throughout the report to understand which of the teachers' ideas concerning their future teaching could be directly linked to changes in the new TOEFL. Not surprisingly, given the nature of the tasks we set, much of what the teachers told us regarding their intentions will have roots in the changes they knew about and understood. The fact that their awareness and understanding were not perfect could be explained by various elements present in Henrichsen's hybrid model of the diffusion/implementation process (1989), which has also influenced our work greatly. Chief among these would be the degree of explicitness present in the explanations of the test provided by the main channel of communication (the ETS Web site). Of importance in the future would be the form that support materials (official ETS materials, but especially commercial preparation materials) would take and factors within the teaching context such as resourcing. Teachers were seeking information about the new test, but it would not be obvious until later whether they would enjoy clarity or suffer from either false clarity or painful unclarity (Fullan, 2001, p. 77). They had positive intentions, but it would not be known until later phases of the Impact Study whether these would manifest themselves in appropriate focus and methodology. The latter is particularly important if the test is to have the impact that was originally intended, and here the issue of communication takes on utmost importance.

A number of useful ideas have emerged from our work with the teachers in Phase 2. Test producers might wish to take these into consideration as they prepare or revise their plans for disseminating information about their tests or for investigating whether the impact they wish their tests to have are working the way they should. These ideas include

- surveying Web site users to elicit comments about the organization of the information on the site and ease of access,
- creating a Web-based teachers' forum that would include
 - frequently asked questions and responses from TOEFL experts,
 - new questions, arising as the test and the challenges of preparing students become better known,
 - samples of student writing and speaking, which teachers can use to develop their understanding of the scoring rubrics, and
 - standards (teachers score these without support to begin with but can later access experts' annotations), and
 - suggestions about where to find the type of materials that forms the input for integrated tasks.
- surveying the major test preparation coursebooks to see whether they are doing an adequate job of informing teachers about not only test format but the constructs underlying test design.

These ideas are in accord not only with the views about communication that arise so frequently in this report but also with an idea taken from one of the major influences on our work, Chapman and Snyder (2000), who spoke of the importance of transmitting messages accurately and efficiently:

changing a test is unlikely to have impact on teachers' instructional practices unless teachers know the changes have been made and understand what actions on their part might prepare students for those changes. Widespread dissemination of information about the knowledge and thought processes being sought by revision to a test is probably more important than the specific changes themselves. (p. 462)

We mentioned in the introduction to Phase 1 (Wall & Horák, 2006) that the impetus for this research was ETS's interest in tracing the impact of the new TOEFL on teaching and learning. We hope that the two phases of research that have been completed so far have made a contribution to understanding of the process of introducing a testing innovation. We look

forward to future phases of the research that will investigate further the relationship between testing and teaching. The next phase, Phase 3, will focus on the messages transmitted by test preparation coursebooks and whether these materials help or hinder the aspirations of the test designers.

References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing, 13*(3), 280–297.
- Andrews, S. (1994). Washback or washout? The relationship between examination reform and curriculum innovation. In D. Nunan, R. Berry, et al. (Eds.), *Bringing about change in language education: Proceedings of the international language in education conference 1994* (pp. 67–81). Hong Kong: University of Hong Kong.
- Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 listening framework: A working paper* (TOEFL Monograph No. MS-19). Princeton, NJ: ETS.
- Beaumont, J. (2005). *NorthStar: Building skills for the TOEFL iBT—intermediate*. New York: Pearson -Longman.
- Beauvois, M. H. (1997). High-tech, high touch: From discussion to composition in the networked classroom. *Computer Assisted Language Learning, 10*(1), 57–69.
- Burnett, C. (2003). Learning to chat: Tutor participation in synchronous online chat. *Teaching in Higher Education, 8*(2), 247–261.
- Butler, F., Eignor, D., Jones, S., McNamara, T., & Suomi, B. (2000). *TOEFL 2000 speaking framework: A working paper* (TOEFL Monograph No. MS-20). Princeton, NJ: ETS.
- Chapman, D. W., & Snyder, C. W. (2000). Can high-stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development, 20*, 457–474.
- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and teacher education, 15*, 253–271.
- Clyde, L. A., & Klobas, J. E. (2000). Shared reflection by email: Its role in new information technology education and training. *Education for Information, 18*, 273–287.
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London: Routledge Falmer.
- Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper* (TOEFL Monograph No. MS-18). Princeton, NJ: ETS.
- Dow, I. I., Whitehead, R. V., & Wright, R. L. (1984). *Curriculum implementation: A framework for action*. Toronto, Canada: Ontario Public School Teachers' Federation.

- Enright, M., Grabe, G., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework: A working paper* (TOEFL Monograph No. MS-17). Princeton, NJ: ETS.
- ETS. (2002). *LanguEdge courseware—Handbook for scoring speaking and writing*. Princeton: Author.
- ETS. (2005a). *Make the connection—Join the world’s most advanced testing network*. Princeton, NJ: Author.
- ETS. (2005b). *TOEFL iBT at a glance*. Princeton, NJ: Author.
- ETS. (2005c). *TOEFL iBT tips—How to prepare for the next generation TOEFL test*. Princeton, NJ: Author.
- Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). London: Cassell.
- Greenfield, R. (2003). Collaborative e-mail exchange for secondary ESL: A case study for Hong Kong. *Language Learning and Technology*, 7(1), 46–70.
- Henrichsen, L. E. (1989). *Diffusion of innovations in English language teaching: The ELEC effort in Japan, 1956–1968*. New York: Greenwood Press.
- Herring, S. (1999). Interactional coherence in CMC. *Journal of Computer Mediated Communication*, 4(4), 1–25.
- Hilke, R., & Wadden, P. (1997). The TOEFL and its imitators: Analyzing the TOEFL and evaluating TOEFL-prep texts. *RELC Journal*, 28(1), 28–53.
- Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 framework: A working paper* (TOEFL Monograph No. MS-16). Princeton, NJ: ETS.
- Jarvenpaa, S. L., Knoll, K., & Leidner, D. L. (1998). Is anybody out there? Antecedents of trust in global virtual teams. *Journal of MIS*, 14, 29–38.
- Jepson, K. (2005). Conversations—Negotiated interaction—In text and voice chat rooms. *Language and Learning Technology*, 9(3), 79–98.
- Kung, S.-C. (2004). Synchronous electronic discussions in an EFL class. *English Language Teaching Journal*, 58(2), 164–173.
- Lam, H. P. (1994). Methodology washback—an insider’s view. In D. Nunan, R. Berry, et al. (Eds.), *Bringing about change in language education: Proceedings of the international language in education conference 1994* (pp. 83-102). Hong Kong: University of Hong Kong.

- Lincoln, Y., & Guba, E. (1985). *Naturalistic enquiry*. Beverley Hills, CA: Sage.
- Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), *The politics of reforming school administration* (pp. 83–121). London: The Falmer Press.
- Mann, C., & Stewart, F. (2000). *Internet communication and qualitative research—A handbook for researching online*. London: Sage.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256.
- Read, J., & Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. In R. Tulloh (Ed.), *IELTS research reports 2003* (vol 4., pp. 154–205). Canberra: IELTS Australia.
- Reese, J. (2002). Web-enhanced foreign language learning: C'est si bon. *Multimedia Schools*, 9(6), 43.
- Richards, J. C. (1984). The secret life of methods. *TESOL quarterly*, 18(1), 7–23.
- Roberts, M. (2002). TOEFL preparation: What are our Korean students doing and why? *The Korea TESOL Journal*, 5(1), 81–106.
- Rogers, E. M. (1983). *The diffusion of innovations* (3rd ed.). London: Macmillian.
- Rogers, E. M., & Shoemaker, F. F. (1971). *Communication of innovations: A cross-cultural approach*. New York: Free Press.
- Samuda, V., Johnson, K., & Ridgway, J. (2000). *Designing language learning tasks: A guide* (vol. 1). Lancaster: Lancaster University, Dept. Linguistics and Modern English Language.
- Scientific Software Development. (2000). Atlas-ti (Version 4.2) [Computer software]. Berlin: Author.
- Solorzano, H. (2005). *NorthStar: Building skills for the TOEFL iBT—High intermediate*. Parsippany, NJ: Pearson Longman.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5–29.
- Torii-Williams, E. (2004). Incorporating the use of e-mail into a language program. *Computer Assisted Language Learning*, 17(1), 109–122.

- Wall, D. (1997). Test impact and washback. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language education: Vol. 7. Language testing and evaluation* (pp. 291-302). Dordrecht, the Netherlands: Kluwer.
- Wall, D. (1999). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Unpublished doctoral dissertation, University of Lancaster, UK.
- Wall, D. (2005). *Studies in language testing: Vol. 22. The impact of a high-stakes examination on classroom teaching*. Cambridge, UK: Cambridge University Press.
- Wall, D., & Horák, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 1, The baseline study* (TOEFL Monograph No. MS-34): Princeton, NJ: ETS.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318–333.

Notes

- ¹ We will refer to the TOEFL iBT as *the new TOEFL* throughout this report, as this is the phrase that we used with our participants during the research we describe.
- ² This reference and those that follow will include the transcript number and the line number in which the information can be found. This reference refers to Teacher 3, Transcript 127, Line 142.
- ³ The observation schedule can be found in Wall & Horák, 2006, pp. 170–178.
- ⁴ At the time we submitted this report for publication, it was possible to find mention of all these features on the TOEFL Web site, but the site was not as complete and informative when we were collecting data.
- ⁵ It is important to record that the new writing section is scored by human raters in the operational test, but there was some confusion among the teachers about whether the rating would be made by humans or by an automated rating tool. This is likely to have been caused by the fact that the writing samples the teachers produced when they took the online writing test on the ETS Web site as part of Task 3 were indeed scored by a program.
- ⁶ The students read a passage and then listen to a lecture. They can look at the passage again to compare what it says with the content of the lecture.
- ⁷ NGT stands for Next Generation TOEFL, the name that was used for the new TOEFL while it was in development.
- ⁸ No publication details available.
- ⁹ There was a potential problem in that a software package called LanguEdge (ETS, 2002) was still being offered as part of the ETS suite of materials until May 2005 to help prospective candidates to familiarize themselves with the new version of TOEFL. This was in fact an out-of-date publication that offered tasks that were in the same paradigm as the new TOEFL but did not reflect them exactly. This was not a problem for teachers in this study as none of them had seen or referred to the publication, but it may have been confusing to other teachers who bought the publication thinking that it would answer their questions about the new test.

Appendix

List of Codes Used in TOEFL Impact Study, Phases 1 And 2

| Codes | Antecedents |
|---|----------------------------|
| Traditional pedagogic practices re: current TOEFL | |
| Aim | Course aims |
| Typ | Course type |
| TSE | Test of Spoken English |
| TWE | Test of Written English™ |
| Vers | Version of TOEFL taken |
| Ct | Content—general |
| CAss | Classroom assessment |
| CGr | Grammar |
| CLang | Re: Language areas general |
| CLs | Listening |
| CMat | Re: Materials |
| CRd | Reading |
| CSp | Speaking |
| CTTT | Re: Test-taking techniques |
| CVo | Vocabulary |
| CWr | Writing |
| Mth | Methodology—general |
| MGr | Grammar |
| Mint | Integrated skills |
| MLang | Re: Language areas general |
| MLs | Listening |
| MMan | Re: Classroom management |
| MMat | Re: Materials |
| MRd | Reading |
| MSp | Speaking |
| MTTT | Re: Test-taking techniques |
| MVo | Vocabulary |
| MWr | Writing |
| Rol | Re: Role of teacher |
| Characteristics of the user system | |
| Crm | Classroom factors |
| Cult | Cultural factors |
| Econ | Economic factors |
| EdAd | Education administration |
| Geo | Geographical factors |
| Pol | Political factors |
| Sch | School factors |
| SchT | Technology in school |
| SchTr | School-based training |

(Table continues)

Table (continued)

| Codes | Antecedents |
|------------------------------|---|
| Characteristics of the users | |
| TAb | Teacher's abilities |
| TAbT | Teacher's technological abilities |
| TACrmT | Teacher's attitude towards classroom teaching |
| TAEd | Teacher's attitude towards education |
| TAEng | Teacher's attitude towards English |
| TAEx | Teacher's attitude towards exams |
| TAIds | Teacher's attitude towards new ideas |
| TALT | Teacher's attitude towards language teaching |
| TATC | Teacher's attitude towards TOEFL classrooms |
| TAtt | Teacher's attitude towards TOEFL |
| TAttN | Teacher's attitude towards the new TOEFL |
| TAw | Teacher's awareness of current exam |
| TAwN | Teacher's awareness of the new TOEFL |
| TCIG | Teacher's goals for class |
| TEcon | Teacher's economic situation |
| TInt | Teacher's interests |
| TLEd | Teacher's level of education |
| TPL | Teacher's personal life |
| TPsG | Teacher's personal goals |
| SAb | Student's abilities |
| SAbT | Student's technological abilities |
| SACrmT | Student's attitude towards classroom teaching |
| SAEd | Student's attitude towards education |
| SAEng | Student's attitude towards English |
| SAEx | Student's attitude towards exams |
| SAIds | Student's attitude towards new ideas |
| SALT | Student's attitude towards language teaching |
| SATC | Student's attitude towards TOEFL classrooms |
| SAtt | Student's attitude towards TOEFL |
| SAttN | Student's attitude towards new TOEFL |
| SAw | Student's awareness of current exam |
| SAwN | Student's s awareness of the new TOEFL |
| SCIG | Student's goals for class |
| SEcon | Student's economic situation |
| SInt | Student's interests |
| SLEd | Student's level of education |
| SOOC | Student's out of class preparation activities |
| SPL | Student's personal life |
| SPsG | Student's personal goals |

(Table continues)

Table (continued)

| Codes | Antecedents |
|----------------------------------|--|
| Characteristics of the users | |
| DAbT | Director of studies' technical abilities |
| DACrmT | Director of studies' attitude towards classroom teaching |
| DAEx | Director of studies' attitude towards exams |
| DAIds | Director of studies' attitude towards new ideas |
| DALT | Director of studies' attitude towards language teaching |
| DATC | Director of studies' attitude towards TOEFL classrooms |
| DAtt | Director of studies' attitude towards TOEFL |
| DAttN | Director of studies' attitude towards the new TOEFL |
| DAw | Director of studies' awareness of TOEFL |
| DAwN | Director of studies' awareness of the new TOEFL |
| <i>Ftr EAP</i> | <i>Features of EAP/advanced general English classes</i> |
| <i>FtrNon-TOEFL</i> | <i>Features of non-TOEFL exam classes</i> |
| <i>FtrTOEFL</i> | <i>Features of TOEFL classes</i> |
| Codes | Process |
| Characteristics of communication | |
| Comm | Communication |
| <i>Delay</i> | <i>Delayed launch of the new TOEFL</i> |
| <i>TQs</i> | <i>Teacher queries re: the new TOEFL</i> |
| <i>MIS</i> | <i>Misapprehensions</i> |
| <i>TSpec</i> | <i>Teacher speculation re: the new TOEFL</i> |
| Receiver | |
| Awareness/interest | |
| <i>TAwNSp</i> | <i>Teacher's awareness of the new TOEFL—Speaking section</i> |
| <i>TAwNRd</i> | <i>Teacher's awareness of the new TOEFL—Reading section</i> |
| <i>TAwNWr</i> | <i>Teacher's awareness of the new TOEFL—Writing section</i> |
| <i>TawNLs</i> | <i>Teacher's awareness of the new TOEFL—Listening section</i> |
| <i>TawNInt</i> | <i>Teacher's awareness of the new TOEFL—Listening section</i> |
| <i>CritSp</i> | <i>Speaking criteria/scoring scales/rubric</i> |
| <i>CritWr</i> | <i>Writing criteria/scoring scales/rubric</i> |
| <i>Perc</i> | <i>Perceptions of TOEFL (contrast attitudes and awareness)</i> |
| <i>SReac</i> | <i>Student reaction to news of the new TOEFL</i> |
| Evaluation | |
| TAtt | Teacher's attitude towards TOEFL |
| TAttN | Teacher's attitude towards the new TOEFL |
| <i>TAttNSp</i> | <i>Teacher's attitude towards the new TOEFL—Speaking section</i> |
| <i>TAttNRd</i> | <i>Teacher's attitude towards the new TOEFL—Reading section</i> |
| <i>TAttNWr</i> | <i>Teacher's attitude towards the new TOEFL—Writing section</i> |

(Table continues)

Table (continued)

| Codes | Process |
|-----------------|--|
| | Evaluation |
| <i>TAttNLs</i> | <i>Teacher's attitude towards the new TOEFL—Listening section</i> |
| <i>TAttNInt</i> | <i>Teacher's attitude towards the new TOEFL—Listening section</i> |
| <i>TEvN</i> | <i>Teacher evaluation of the new TOEFL</i> |
| <i>SEvN</i> | <i>Student evaluation of the new TOEFL</i> |
| | Factors that facilitate/hinder change |
| | Characteristics of the innovation (i.e., the new TOEFL) |
| Comx | Complexity |
| Expl | Explicitness |
| Flex | Flexibility |
| Fm | Form |
| Obs | Observability |
| Orig | Originality |
| Pra | Practicality |
| Prim | Primacy |
| RelAd | Relative advantage |
| <i>Comps</i> | <i>Comparisons with other exams</i> |
| Stat | Status |
| Tri | Trialability |
| | Characteristics of the resource system (i.e., ETS) |
| Cap | Capacity |
| Hy | Harmony |
| Op | Openness |
| St | Structure |
| <i>Tech</i> | <i>Technological features</i> |
| | Characteristics of the user system |
| Crm | Classroom factors |
| Cult | Cultural factors |
| Econ | Economic factors |
| EdAd | Education administration |
| Geo | Geographical factors |
| Pol | Political factors |
| Sch | School factors |
| SchT | Technology in school |
| SchTr | School-based training |
| <i>TLU</i> | <i>Target language use</i> |
| | Characteristics of the users (Wall version of model) |
| | All teacher, student, and director of studies codes as used in antecedents |
| <i>TBLs</i> | <i>Teacher's beliefs about construct of listening</i> |
| <i>TBRd</i> | <i>Teacher's beliefs about construct of reading</i> |

(Table continues)

Table (continued)

| Codes | Process |
|--|--|
| Characteristics of the users (Wall version of model) | |
| All teacher, student, and director of studies codes as used in antecedents | |
| <i>TBSp</i> | <i>Teacher's beliefs about construct of speaking</i> |
| <i>TBWr</i> | <i>Teacher's beliefs about construct of writing</i> |
| <i>TBGr</i> | <i>Teacher's beliefs about construct of grammar</i> |
| <i>TBInt</i> | <i>Teacher's beliefs about integrated skills</i> |
| <i>TBVo</i> | <i>Teacher's beliefs about construct of vocabulary</i> |
| <i>TBLang</i> | <i>Teacher's beliefs about language in general</i> |
| <i>TConf</i> | <i>Teacher's confidence</i> |
| Consequences ^a | |
| <i>Plns</i> | <i>Plans re: introduction of the new TOEFL courses</i> |
| <i>CtN</i> | <i>Content—the new TOEFL general</i> |
| <i>CRdN</i> | <i>Content of TOEFL classes re: the new TOEFL reading</i> |
| <i>CGrN</i> | <i>Content of TOEFL classes re: the new TOEFL grammar</i> |
| <i>CLsN</i> | <i>Content of TOEFL classes re: the new TOEFL listening</i> |
| <i>CSpN</i> | <i>Content of TOEFL classes re: the new TOEFL speaking</i> |
| <i>CWrN</i> | <i>Content of TOEFL classes re: the new TOEFL writing</i> |
| <i>CtNon-LangN</i> | <i>Content of TOEFL classes re: the new TOEFL other than language</i> |
| <i>CTTTN</i> | <i>Content of TOEFL classes re: the new TOEFL test taking techniques</i> |
| <i>CMatN</i> | <i>Materials for teaching the new TOEFL</i> |
| <i>WebMats</i> | <i>Web-based support materials for the new TOEFL</i> |
| <i>MRdN etc.</i> | <i>Methodology in the new TOEFL classes re: reading</i> |
| <i>MGrN</i> | <i>Methodology in the new TOEFL classes re: grammar</i> |
| <i>MLsN</i> | <i>Methodology in the new TOEFL classes re: listening</i> |
| <i>MSpN</i> | <i>Methodology in the new TOEFL classes re: speaking</i> |
| <i>MWrN</i> | <i>Methodology in the new TOEFL classes re: writing</i> |
| <i>MTTTN</i> | <i>Methodology in the new TOEFL classes re: TTT</i> |
| <i>Rol</i> | <i>Re: Role of teacher</i> |
| <i>Item</i> | <i>Familiarization with item types to be found on the TOEFL</i> |
| <i>Fam</i> | <i>Familiarization with test in general</i> |
| <i>Fback</i> | <i>Feedback</i> |
| <i>FbackWri</i> | <i>Feedback to students on their writing</i> |
| <i>FBackSp</i> | <i>Feedback to students on their speaking</i> |
| <i>TSupp</i> | <i>Teacher support</i> |
| <i>TT</i> | <i>Teacher training</i> |
| <i>WB</i> | <i>Washback</i> |
| <i>Impl</i> | <i>Implications</i> |

(Table continues)

Table (continued)

| Codes | Process |
|-------------------------------|---|
| Background info ^a | |
| <i>SData</i> | <i>Info on types of students</i> |
| <i>TData</i> | <i>Info about Ts since Phase 1</i> |
| <i>CrseDate</i> | <i>Course dates</i> |
| <i>CrseLgth</i> | <i>Course length</i> |
| Tracking changes ^a | |
| <i>QInfo</i> | <i>Tracking question: Do you have any news sources of info?</i> |
| <i>QInst</i> | <i>Tracking question: Has it been discussed in your institution?</i> |
| <i>QMonth</i> | <i>Tracking question: Has anything happened this month?</i> |
| <i>QNew</i> | <i>Tracking question: Have you learned anything new?</i> |
| <i>QSs</i> | <i>Tracking question: Have the Ss asked anything?</i> |
| <i>QWorries</i> | <i>Tracking question: Do you have any concerns re: the new TOEFL?</i> |

Note. Italics indicate codes that were created in Phase 2.

^a Additional categories not included in the Henrichsen (1989) model.



**Test of English as a Foreign Language
PO Box 6155
Princeton, NJ 08541-6155
USA**

To obtain more information about TOEFL programs and services, use one of the following:

**Phone: 1-877-863-3546
(US, US Territories*, and Canada)**

**1-609-771-7100
(all other locations)**

**E-mail: toefl@ets.org
Web site: www.ets.org/toefl**

*America Samoa, Guam, Puerto Rico, and US Virgin Islands