ETS TOEFL

Quality Beyond Measure.

TOEFL iBT® Research Report
TOEFL iBT–17

The Impact of Changes in the TOEFL® Exam on Teaching in a Sample of Countries in Europe: Phase 3, The Role of the Coursebook Phase 4, Describing Change

Dianne Wall

Tania Horák

October 2011

The Impact of Changes in the TOEFL® Exam on Teaching in a Sample of Countries in Europe: Phase 3, The Role of the Coursebook Phase 4, Describing Change

Dianne Wall and Tania Horák
Lancaster University, United Kingdom

RR-11-41

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.

Abstract

The aim of this report is to present the findings of the 3rd and 4th phases of a longitudinal study into the impact of changes in the TOEFL® exam on teaching in test preparation classrooms. Phase 1 (2003–2004) described the type of teaching taking place in 12 TOEFL preparation classrooms before the introduction of the new TOEFL. Phase 2 (2004–2006) followed 6 of the Phase 1 teachers as they became aware of the requirements of the new test and faced the challenges of designing new courses to help students to prepare for it effectively. The objectives of the Phase 3 study (2006–2007) were to analyze the coursebooks that 4 of these teachers were using as they continued to prepare students for the TOEFL computer-based test (CBT) and began to plan courses for the TOEFL iBT®, and to find out how the teachers were using the coursebooks as they developed their courses and planned individual classroom sessions. The coursebook analysis revealed that the TOEFL iBT coursebooks differed considerably from the TOEFL CBT coursebooks in terms of content, with the inclusion of integrated writing tasks and independent and integrated speaking tasks and the absence of attention to grammatical form on its own. They did not differ greatly in terms of their general methodological approach, however. Information about how the teachers used their coursebooks was gathered via tracking questions and tasks eliciting self-report data. The coursebooks seemed to be playing an important role in shaping the teachers' understanding of the requirements of the new TOEFL, and the teachers depended on them heavily as they developed their courses and planned their lessons. The objectives of Phase 4 (2007–2008) were to interview and observe 3 of the same teachers, to find out what their preparation classes looked like 1 year after the introduction of the TOEFL iBT in their countries. While some aspects of teaching seemed not to have changed greatly, considerable changes occurred in the amount of attention the teachers paid to the development of speaking and to the integration of different skills. The teachers differed from each other in how much they had changed their methods to develop their students' language skills. The report concludes with a discussion of the role the new test, new coursebooks, and other factors in the educational context played in shaping current practices in these TOEFL preparation classrooms.

Key words: washback, impact, TOEFL, CBT, TOEFL iBT, coursebook, Europe

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 (2010-2011) members of the TOEFL Committee of Examiners are:

Alister Cumming (Chair)	University of Toronto
Carol A. Chapelle	Iowa State University
Barbara Hoekje	Drexel University
Ari Huhta	University of Jyväskylä, Finland
John M. Norris	University of Hawaii at Manoa
James Purpura	Columbia University
Carsten Roever	University of Melbourne
Steve Ross	University of Maryland
Mikyuki Sasaki	Nagoya Gakuin University
Norbert Schmitt	University of Nottingham
Robert Schoonen	University of Amsterdam
Ling Shi	University of British Columbia

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

Acknowledgments

We wish to express our thanks to the TOEFL® Committee of Examiners and the TOEFL program at ETS for funding what turned out to be a 5-year study into the impact of the new TOEFL test on teaching practices. We are grateful to the members of the research subcommittee for their support during the course of the project, and to Dr. Mary Enright for her encouragement and help throughout the process.

We would also like to thank our colleagues in the Language Testing Research Group at Lancaster University for their valuable feedback on our work over the years.

Finally, we wish to acknowledge the help of all the 12 teachers who took part in the Phase 1 study (2003-2004), and of the 7 teachers who continued to work with us in later phases. We are especially grateful to the "survivors"—the 3 teachers who stayed with us until the very end of the project in early 2008. We thank them for allowing us into their classrooms and for answering countless questions, in interviews and in virtual communication, about their beliefs, attitudes, and teaching practices.

Table of Contents

List of Tables

Background

The aim of this report is to present the findings of Phases 3 and 4 of a longitudinal investigation into the impact of changes in the TOEFL® exam on teaching in test preparation classrooms. Phase 3 focused on the role of commercial coursebooks in disseminating information about the new TOEFL, and Phase 4 focused on describing the type of teaching that was taking place in three test preparation classrooms approximately 1 year after the new TOEFL was launched in the countries represented in our sample.

The study found that there were important changes in the teaching of the three teachers who participated in Phase 4, particularly in terms of the content of their teaching. These changes were uniform across the teachers and can be seen as positive, in the sense that they correspond to the impact on teaching content intended by the designers of and advisors to the new TOEFL. There was more variation in the methods the teachers used deliver their lessons, however, with two teachers using methods that encouraged more interaction and communication than previously, while the remaining teacher continued using the same methods she had used before the launch of the new TOEFL.

Before discussing the details of the Phase 3 and Phase 4 studies, we briefly review the rationale for revising the TOEFL, give an overview of the TOEFL Impact Study as a whole, and summarize the findings of the first two phases of the investigation.

Rationale for Revising the TOEFL Exam

McNamara, writing in 2001, described the then-current TOEFL as being "based on models of language and its measurement dating back to the 1960s" (p. 2). The test had undergone some revisions since its creation in 1964, but it reflected a structuralist view of language well into the *communicative* era, with considerable weighting on language knowledge, the receptive skills of reading and listening (tested separately), and a form of writing that only partially represented the demands placed on students in tertiary level academic settings. Speaking was not assessed in the TOEFL itself, but in the TSE®, an associated test that was not required by many receiving institutions. Taylor and Angelis (2008) described the 1980s as a time when those in charge of TOEFL development began "to wrestle with the need for integrative measures requiring constructed responses and the complexities introduced by communicative competence theory" (p. 37).

Taylor and Angelis (2008) referred to a number of projects undertaken in the 1990s to investigate and possibly redefine the purpose of the TOEFL, explore various operational issues, and determine what the goals of any new version of the test should be, in terms of construct and design. They explained that in addition to the goal of creating a test that would reflect modern theories of communicative language use relevant to an academic context, there was a tacit goal of producing a test that would be "more aligned with current language teaching practice and thus create a test with more positive washback than the current TOEFL" (p. 42).

Intensive research activity took place in the late 1990s, leading to the production of a general framework for the test design (Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 2000) and more specific framework documents for each of the macro-skills that would be tested in the future: reading (Enright, Grabe, Koda, Mosenthal, Mulcahy-Ernt, & Schedl, 2000), listening (Bejar, Douglas, Jamieson, Nissan, & Turner, 2000), writing (Cumming, Kantor, Powers, Santos, & Taylor, 2000), and speaking (Butler, Eignor, Jones, McNamara, & Suomi, 2000). These framework documents explored the constructs and content that might be covered in a new TOEFL and recommended further research that would, by the early 2000s, lead to decisions about the final shape of the new test.

The major changes that were eventually decided on were

- elimination of a separate structure (grammar) section,
- addition of an integrated writing task (listening and reading inputs leading to a writing output),
- addition of a speaking section, testing this skill on its own and in an integrated manner (with reading and listening inputs), and
- note-taking would be allowed throughout the whole of the test.

These changes and others are explained in more detail in later sections of this report.

Alongside the research focussing on construct and design issues, work was commissioned to explore questions relating to the washback that the new TOEFL might produce and how this could be investigated over time (Bailey, 1999). The resulting report was one of a series of reports informing the validation process that accompanied the development of the new TOEFL (Chapelle, Enright, & Jamieson, 2008a) and addressing that part of the process concerned with the consequential aspects of the test's validity (Messick, 1989).

TOEFL Impact Study in Central and Eastern Europe

The TOEFL Impact Study (hereafter referred to as *the Impact Study*) was commissioned by the TOEFL Research Subcommittee in 2002, with the brief of determining whether changes in the new TOEFL would lead to changes in classroom practices. The researchers (hereafter referred to as *we*) were asked to set up the study in Central and Eastern Europe, an area that was felt to have had limited communication with American educational institutions such as ETS and that could serve as a "test case for the extent of and barriers to the diffusion of knowledge about innovations in the test and implications for teaching" (Wang, Eignor, & Enright, 2008, p. 299). We envisaged a long-term study that would be divided into several phases, to provide not only an accurate description of the teaching and learning taking place at different times, but also to allow the results to be used by TOEFL management in a formative way. The first phase would be a baseline study, to describe the type of teaching that took place before the introduction of the new TOEFL and, ideally, before the release of any details that might tempt teachers to begin altering their approach to teaching. Further phases would be added as the situation required to determine whether key events, such as the release of sample materials, resulted in changes in teaching practice and to recommend ways in which teacher support efforts might be strengthened in the future. The final phase would provide descriptions of teaching after the introduction of what came to be known as the TOEFL iBT®. It would also provide explanations for why changes might or might not have taken place, again with the intention of feeding results back into the test design and dissemination process. In this way, the new test could benefit from contact with its users and could respond to their needs.

It was not clear at the time the study was commissioned when the launch date for the new TOEFL would be. A phased rollout meant that it was not until mid-2006 that the test went operational in the countries we were studying. The phased introduction gave us an opportunity not only to describe the teaching that was taking place before the test was generally known about, but also to study the reactions of a sample of teachers and their institutions during the 2-year period between our baseline visit and a second visit after the new test had "settled in."

The study eventually included four phases, as shown in Table 1.

Table 1

TOEFL Impact Study: Phases 1 to 4

Phase	Name	Dates	Sample	
			No. of teachers	Countries
1	Baseline study [a]	January 2003–June 2004	12	7
2	Transition study—coping with change	September 2004–March 2006	6	5
3	Transition study—the role of the coursebook	April 2006–March 2007	4	4
4	Describing change	April 2007–March 2008	3	3

Note. A timeline showing the details of the entire Impact Study can be found in Appendix A.

[a] The original baseline study (Wall & Horák, 2006) was based on interviews and observations with 10 teachers in 6 countries. At the request of the TOEFL Research Subcommittee, we observed two more teachers in a seventh country in October 2004. The findings from these teachers matched those from the original 10, so for ease of reference we here refer to the baseline study as having dealt with 12 teachers in 7 countries.

The Impact Study was unique in that the teachers who participated in Phases 2 to 4 had all been visited in Phase 1. Two of the three teachers who participated in Phase 4 had participated in all of the earlier phases, and the third had participated in Phases 1 and 3. We were therefore able to gather a considerable amount of data from the same individuals over the course of several years. This sustained contact has enabled us to write with confidence about the teachers' experience over the course of the full study and to feel secure about our conclusion that the changes apparent in teaching practices in Phase 4 of the study can be linked to changes in the new TOEFL. (The reasons for attrition between the stages are explained in the Methodology section for Phase 3.)

Phase 1 Findings

The main aims of the first phase of the Impact Study were

- to determine what sorts of impact the designers of the new TOEFL meant it to have, and

- to describe the characteristics of TOEFL preparation classes before the introduction of the new examination.

From the beginning of the study, the term *impact* was taken to mean the same as *washback* (also known as *backwash*), which in its most general sense refers to "the effect of testing on teaching and learning" (Hughes, 2002, p. 1). Some researchers make a distinction between impact and washback, using impact to refer to the effects a test might have on the general educational context or even society more generally and washback to refer to the test's influence on what takes place in the classroom (Wall, 1997). We use the terms interchangeably in this report, however, given that the original invitation from TOEFL management used the term *impact* but also made it clear that what was to be investigated were changes that might occur in TOEFL preparation courses after the introduction of the new TOEFL.

A full account of the Phase 1 investigation can be found in Wall and Horák (2006) and relevant details are referred to below. It is useful to give a summary of the findings, however, in order to set a context for the work that followed.

In order to address the first aim, we surveyed the framework documents that laid the foundations for the new test and contacted experts who had contributed to its design. Although there seemed to have been a general desire for the new test to have a beneficial effect on teaching, there were no detailed statements in the framework documents of what this washback should look like, and the experts were unable to recall discussions in which washback had been discussed in a thorough way. We summarized our findings in this way:

> There was a general hope that the new TOEFL test 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)

It was of interest that only a few of the experts said that they had been involved in discussions about how to go about achieving positive washback. Several of their responses suggested a belief that if the test design were right, then beneficial washback would follow automatically. Only one expert mentioned the need to produce test preparation materials, to prepare workshops for teachers, and to make information about the test development process available to the test users.

In order to address the second aim we identified a sample of 10 teachers in six different countries, six of them local to the area they were teaching in and four of them American or British expatriates. We designed and piloted interview schedules for the teachers, their students, and their directors of studies, and an observation schedule to use when visiting their TOEFL preparation classes and general advanced classes. We found that the teachers had little to no awareness of the upcoming changes in the TOEFL, so we could safely assume that the teaching they told us about and that we observed could serve as a baseline against which we could measure possible future changes. In general, their teaching was coursebook based and teacher dominated, with very little resemblance to the communicative approaches encouraged by modern teacher educators and recent materials. The main characteristics of their teaching were as follows:

Listening. Teachers did not know how to break listening down into teachable subskills, and they had few techniques for developing listening as opposed to assessing it. They seemed to believe that students would improve their listening through a process of osmosis, through copious practice inside and outside the classroom.

Grammar. Teachers generally expected students to have attained a certain level of grammatical knowledge before they entered TOEFL courses, but this did not eliminate the need for considerable review of grammatical points, especially those believed to be "tricky." This review took the form of coursebook exercises, some drilling, and a focus on grammar when marking student writing.

Reading. Teachers knew more about the subskills for reading (e.g., skimming, scanning, referencing, inferencing) than they did for listening, and they practiced them via exercises in their coursebooks. They made little use of modern techniques to activate schemata or to encourage the discussion of ideas. Some teachers relegated reading to homework, which meant that they could not be sure that their students were reading quickly or selectively. Much attention was paid to improving vocabulary, which was considered to be a key challenge in reading.

Vocabulary. Vocabulary was thought to be crucial not only to reading, but also to other skills tested on the TOEFL. The two main means of helping students with vocabulary were distributing lists of words and phrases and encouraging students to pay attention to vocabulary in their outside reading. The teachers had few techniques for developing word skills and the burden

seemed to be on the students to expand their vocabulary on their own. Many relied on practice materials on CDs they had bought themselves or that they had found on the Internet.

Writing. The teachers devoted a great deal of classroom time to writing, as it was generally felt that students had not received adequate training in this skill at school. They concentrated on the structure of essays, using a formulaic approach that was presented in their coursebooks, and paid less attention to the content of the writing. Most teachers expected the students to write at home, and the feedback they gave was based less on the TOEFL writing rubrics (rating scales) than on advice given in the coursebooks or their personal experience as students in academic settings.

Speaking. English was the medium of teaching in nearly all the courses, but this emphasis occurred because teachers wanted to give their students practice listening to the language or because the teachers were expatriates who had not learned the local language. Little attention was paid to developing speaking as a separate skill and the main reason for this decision was that speaking was not tested on the TOEFL.

None of the teachers had received special training for teaching TOEFL classes, and most of them stated in their interviews that they depended on their coursebooks for information about the test itself and for practice material. The coursebooks were mainly designed for students to use on their own, so there was no advice about techniques teachers could use to promote learning in the classroom. The teachers consulted ETS materials, including the TOEFL Web site, and other Web sites, but their coursebooks remained the major influence on their teaching.

The findings of the Phase 1 study were submitted to ETS in early 2004. The research subcommittee approved a proposal for a second phase of research and also requested that visits be made to teaching institutions in a country in Western Europe. We made a visit to a seventh country in October 2004 and interviewed and observed two teachers in one institution (a second institution had agreed to participate in the study, but they informed us when we were already in the country that they no longer wished to cooperate). The institution we visited was probably the best resourced of all the institutions in our study, and it offered strong support to TOEFL teachers through in-house training opportunities and encouragement from management. We found, however, that the classes were very similar to the classes we had observed at most of the other sites in Phase 1. They were managed in a teacher-centered lockstep way, aiming at familiarization with test text- and task-types via practice materials mirroring the TOEFL.

Although the teachers claimed to have knowledge of communicative teaching methods, they did not feel these were suitable for test preparation classes.

Phase 2 Findings

The main aim of the second phase of the Impact Study was to monitor a number of the teachers who had participated in Phase 1, to find out how they were reacting to news about the new TOEFL, and to learn how this news was affecting the plans they were making for preparation courses for the future. The research focused on these questions:

- How aware were teachers of the differences between the old and new versions of the TOEFL?
- What was their attitude to the new test as they understood it?
- What were the implications of what they understood for their preparation classes in the future?

A complete account of the Phase 2 study can be found in Wall and Horák (2008), but a summary is provided here to facilitate comparisons with the findings of Phases 3 and 4, which are presented in separate sections of this report.

The sample for this phase consisted of six teachers from five different countries. These were the teachers who were still teaching TOEFL classes when Phase 2 began and who also had the interest and the technical possibilities (mainly the ability to access and use the Internet) of working with us at a distance. The nature of our research questions meant that we would need to be in close contact with the teachers over quite a long period rather than, as in Phase 1, collecting data from them in a single visit. We corresponded with the teachers over a period of 5 months, from January to May 2005. This was the time when TOEFL management was beginning to release information about the new test but when the teachers and their institutions did not know when the test would be launched in their countries.

We used two means of data collection:

- Monthly tracking questions sent by e-mail, and followed up either by e-mail or by computer-mediated interviews using MSN Messenger
- Monthly tasks, followed up by MSN Messenger interviews discussing the task responses

The tracking questions asked teachers what, if anything, they had learned about the new TOEFL since our last contact with them, where they had gotten their information from, whether they or their students were experiencing any problems understanding or reacting to the new test, and what plans the teachers had for preparation classes in the future. The tasks were designed to probe the teachers' views of what test preparation classes should consist of, their awareness of the then-current computer-based TOEFL (CBT) and the new TOEFL iBT, their understanding of the TOEFL iBT integrated writing task and scoring rubrics, their understanding of the TOEFL iBT speaking tasks and scoring rubrics, and their views of the types of content and teaching methods they might use in their future test preparation courses.

By the end of Phase 2 (March 2006) we had found the following:

- The teachers had experienced difficulties in the first few months of 2005 because of a lack of information about the test and because they did not know when it would be introduced in their countries.
- They had many questions about the content and the format of the new test.
- Their awareness and understanding increased once the tasks we set for them forced them to think carefully about the differences between the old and the new tests.
- Their attitude toward the new test was generally positive. They liked the idea of authentic materials, tasks that represented the demands of the target language use situation, the inclusion of speaking, and the marking rubrics for writing and speaking.
- They were not sure, however, how they should go about teaching speaking and integrated skills.
- They had not had much practice using the marking rubrics and were not confident about how to incorporate them into their teaching.
- Several teachers were interested in using more communicative tasks in the future.
- All of the teachers were worried about delays in the appearance of TOEFL preparation coursebooks in their countries.

One of the main themes emerging from the study was the importance of getting clear and accurate information about the new test. The TOEFL Web site was the teachers' main source of

information, though they did not report using all of the information that was available. They were exposed to the online practice test through their work with us, but it is debatable whether they would have paid for access to the practice test if they had not gained it through our study. Some teachers were aware of the TOEFL workshops, but only one teacher had the funding to attend one. The teachers sought information from non-ETS Web sites but these were not very helpful, and their attempts to obtain information from educational agencies such as the local Fulbright offices were disappointing. The teachers placed great hope in the appearance of new test preparation coursebooks, but these were slow in arriving and few were available at the end of the data collection period. It was not clear at that time which books the teacher would use and whether these would provide adequate information about the constructs underlying the test or any advice about how to organize teaching.

Teachers needed to think about a number of factors when planning their future courses, not just what the test would look like. Amongst these were factors relating to the *user systems* (Henrichsen, 1989) they worked in (time-tabling constraints, classroom conditions, institutional priorities, client characteristics and demands, division of labor and power relationship within the institutions [especially between the directors of studies and the teachers], and resourcing [for teacher training, computers, Internet connections and libraries]).

Also important were the teachers' own characteristics: their knowledge of teaching, the types of experience they had gathered over their career, their level of confidence, and their motivation.

It was still not certain at the end of the Phase 2 study when the new test would appear in the countries in our sample. Therefore, it was not until Phase 3 that we were able to explore how the plans the teachers were beginning to put together would actually work out in practice.

Organization of This Report

The rest of this report is dedicated to a description and discussion of Phases 3 and 4 of the Impact Study.

The next section presents an account of the research undertaken in Phase 3, which focused on the coursebooks that the teachers were using shortly after the launch of the new TOEFL in their countries and on the effect that these coursebooks were having on their teaching. It contains a brief review of the literature on the role of the coursebook in language teaching and in the creation of test washback, a description of the methods used in Phase 3, an analysis of the

coursebooks in use during this phase, and a discussion of the teachers' reactions to their coursebooks and how they used them in their classrooms.

The following section presents an account of the Phase 4 study. It includes a description of the methods that were used to collect data and an account of the findings relating to the teaching of the four skills tested in the new TOEFL and to the teaching of grammar and vocabulary. It presents an analysis of the means of communication that the teachers used to keep themselves informed about the new TOEFL and how they should teach TOEFL classes. It also deals with three further themes that emerged from the Phase 1 study: the use of computers in the classroom, the types of assessment that were carried out in the teaching institutions, and the types of teacher training that the teachers could access. There is also a discussion of other factors influencing the type of teaching that was taking place during this phase.

The report concludes with a discussion of whether the impact that the new test was meant to have on teaching (as identified in Phase 1 of this study) had appeared by the end of Phase 4.

The Phase 3 Study

Aims of the Study

The first aim of Phase 3 was to carry out a detailed analysis of the test preparation books that were being used by the teachers in our sample before and just after the launch of the new TOEFL in their countries. The second aim was to find out how the teachers were using their coursebooks as they began teaching groups of students preparing for the new test.

The focus on coursebooks seemed reasonable in the light of the findings of Phases 1 and 2, which showed that test preparation coursebooks were "at the heart of the majority of the courses investigated" (Wall & Horák, 2006, p. 78). Coursebooks provided the syllabus for teaching in most of the Phase 1 classrooms, and most teachers worked through them systematically. Only 1 teacher of the 12 teachers we interviewed had actually taken the TOEFL as a learner. For the others the coursebooks functioned as their main source of information about the content and format of the test and how it would be marked. We found in Phase 2 that the biggest worry teachers had when they were trying to decide what to include in their TOEFL iBT preparation courses was whether they could find suitable coursebooks to guide their course design. Coursebook producers seemed to have a great deal of influence on what the teachers taught and what the students studied. It was therefore interesting to learn that the coursebooks were not always selected on the basis of a informed analysis but rather for reasons such as price

or because their use in a given institution gave that institution a "market edge" over similar institutions in the same context.

The focus on coursebooks also seemed suitable given the claims made about the importance of coursebooks in recent literature on language teaching and testing.

The Role of Coursebooks in Language Teaching and Testing

As can be seen in the following short review, although numerous theorists have argued that coursebooks can limit teachers' creativity and encourage conservatism and rigidity in teaching, others argue that most teachers are happy to take advantage of coursebooks that save them the trouble of having to set syllabuses, design materials, and plan classroom activities on their own. The coursebook is also important as it represents a compromise between what is theoretically desirable and what classroom teachers are able to understand and implement, especially in situations where professional development opportunities are not available and there are practical considerations to consider. The coursebook can assume particular importance if it is designed for test preparation purposes, as it may be the main source of information teachers have regarding the test construct, format, tasks and criteria for marking.

The importance of the coursebook in language teaching. A search of the literature on language teaching confirms that it is not unusual for teachers to depend heavily on their coursebooks, despite warnings from some quarters that these can "absolve teachers of responsibility" (Swan, 1992, p. 33, as quoted in Hutchinson & Torres, 1994, p. 315) and lead to a situation in which teachers become mere managers of "a preplanned classroom event" (Littlejohn, 1992, p. 84, in Hutchinson & Torres, 1994, p. 316). Thornbury (2000) argued that an overreliance on commercial coursebooks runs counter to beliefs that language learning depends on teacher-learner and learner-learner interaction, that the learners' experiences and concerns provide valid content for this interaction, and that one of the main roles of the teacher is to optimize the "language learning affordances" (p. 3) that emerge from talk produced in the classroom. (These language-learning affordances are similar to learning opportunities noted by Allwright, 2000.) Such ideas have much in common with Breen's (1987) notion of the "process syllabus" (p. 169), involving negotiation of content, input materials, and techniques for teaching and assessment. Long and Crookes (1992) provided further explanation of the process syllabus, and also elaborated on the "procedural syllabus" and "task-based language teaching," noting that all three approaches reject the idea of a predetermined "synthetic" syllabus (one in which the

learner is expected learn a language in parts, and then put the parts together "when the time comes to use them for communicative purposes" (p. 28).

Other writers see the coursebook not as a restraining influence but as a necessary and valuable support for overworked teachers. Hutchinson and Torres (1994) argued that discussions of process and negotiation ignore the reality of most teachers, who have little time outside classroom hours to devote to designing their own teaching programs or instructional materials. Their main need is not for "maximum freedom. . . but for a predictable and visible structure within the lesson and across lessons" (p. 321). Hutchinson and Torres claim that while no textbook is perfect, textbooks in general provide the structure and security that teachers (and learners) need to be able to work confidently. They provide not only "something to negotiate about," but also a representation of what goes on in the classroom for other stakeholders in the educational context ("accountability") and an orientation in relation to "what is expected of them (the teachers), what is regarded as acceptable or desirable in terms of content, what objectives should be reached, how much work should be covered in a given time, etc." (p. 320).

Hutchinson and Torres (1994) further claim that textbooks are especially important when change is being introduced into an educational system, as they keep the disturbances caused by change "within manageable limits" (p. 321). This idea complements Henrichsen's (1989) notion of "form" (p. 85) in educational innovation and is in accord with Spratt's (2005) suggestion that teachers may depend on their coursebooks more when these are the only representation they have of what change should look like. Spratt used the term "a fruit of uncertainty" (p. 12) to indicate the function of coursebooks in periods of change and suggested that teachers may rely on them less when they have become accustomed to what is required in a new approach to teaching.

In a recent analysis entitled "Advances in Materials Design," Waters (2009) explained that the term *advances* could be understood in two different ways. The applied linguistics perspective viewed advances in materials design as being the successful application of "advances in academic theorising and research concerning language, language learning and education" (Waters, 2009, p. 312). Such advances would include the beliefs or principles underlying the process, procedural, and task-based approaches mentioned earlier. The second perspective is a more "audience-based one" which caters for the needs of "end-users of teaching materials" (teachers and learners) as these users perceive them (p. 312). Waters reviewed earlier surveys of

published teaching materials (Clarke, 1989; Rossner, 1988), which judged the extent to which modern textbooks succeeded in incorporating theoretical notions such as authenticity (of text, task, and context) and purposeful communication. Waters summarized Rossner's conclusions by writing that "a more 'traditional' focus is perceived to have remained intact, despite the addition of a communicative 'overlay'" (p. 313). He summed up Clarke's conclusions in a similar way: that "the majority of the teaching materials reviewed were seen to have failed to live up to the theoretical ideals of the communicative approach" (p. 314). Rather than criticize the textbook authors and publishers who produce such materials though, Waters emphasized that their job is "a difficult, complex and highly-skilled process, involving, in particular, the notion of a *compromise* between what might be theoretically desirable and what is practicable and appropriate in audience terms" (pp. 323–234).

Waters (2009), like Hutchinson and Torres (1994), called for more research into teachers' and learners' attitudes toward different types of textbook design, as well as into how textbooks are actually used in and out of the classroom, stating that this research would provide useful data "for informing optimization of their design" (Waters 2009, p. 324).

The importance of the coursebook in language testing. Spratt (2005) surveyed approximately two dozen studies of test impact and washback written in the last two decades and categorizes the findings according to whether they involved impact on curriculum, teaching materials, methodology, attitudes, or learning. It is the impact on teaching materials that is most relevant to the present study. Spratt sorted this impact into four different categories: 1) the production and marketing of materials to facilitate preparation for high-stakes tests, 2) the use of these materials, 3) the users' views of such materials, and 4) the content of the materials. Spratt referred to a number of studies that recorded a heavy use of test-related materials in the classroom, amongst them Lam (1994); Andrews (1995); Andrews, Fullilove, and Wong (2002); Cheng (1997); and Read and Hayes (2003). The first four studies were carried out in Hong Kong, where a series of new tests had been introduced into the educational system, and it was this connection between new tests and heavy test-related materials usage that led Spratt to the idea that teachers may depend on their coursebooks more during periods of change (p.12). This notion does not seem to apply, however, to the situation described by Read and Hayes (2003) or the context investigated by Alderson and Hamp-Lyons (1996), where the tests for which teachers were preparing their students were already embedded in the educational context.

Spratt's (2005) discussion of teachers' and learners' views of test-preparation materials included references to Lumley and Stoneman (2000), who found that while teachers in a tertiary setting in Hong Kong were pleased that their materials had the potential for taking learners "beyond test preparation" (p. 75; Spratt, 2005, p. 12), the learners themselves were very test-focused and lacked interest in developing non-test-related strategies or abilities. In contrast, Alderson and Hamp-Lyons (1996) found that teachers said they used certain test-preparation materials because their learners insisted on it, while the learners themselves claimed that there were other more interesting ways to prepare for the test they were facing.

Spratt's (2005) discussion of the content of test-preparation materials highlighted Hamp-Lyons' (1998) survey of five TOEFL preparation coursebooks, which revealed the extent to which the content of the coursebooks related directly to the test for which they were providing support. The messages we felt were strongest in the Hamp-Lyons discussion were her disapproval of the coursebooks on the grounds that they did little to develop the learners' language ability beyond the test-taking requirements (compare this with Lumley & Stoneman, 2000, above) and her views that some of the coursebooks were either hovering on the edge of or had crossed over into the realm of "unethicality." She based this judgment on frameworks devised by Mehrens and Kaminsky (1989) and Popham (1991), which judged as unethical those materials that aim to boost scores without necessarily encouraging a mastery of the domain being tested.

Wadden and Hilke (1999) responded to Hamp-Lyons' (1998) article, criticizing it for reaching general conclusions on the basis of only a small sample, discussing their own (very different) findings in an earlier survey (Hilke & Wadden, 1997), and questioning Hamp-Lyons' arguments that the coursebooks she had analyzed would not help learners to develop any but test-specific abilities and that they were unethical. Wadden and Hilke believed that it was necessary to "critically educate students as to which materials are the most accurate, representative and appropriate for their own interests and to encourage and empower them in achieving their own educational goals" (p. 270).

In addition to criticizing Hamp-Lyons (1998), Wadden and Hilke (1999) criticized ETS for selling its own test-preparation materials (unspecified) and institutions of higher learning for "indiscriminately and imprudently use the TOEFL as their principal initial criterion" (p. 269), thus encouraging students to concentrate on gaining test-specific competence rather than

developing further their general language proficiency. Hamp-Lyons' reply (1999) made clear that her comments had been about certain TOEFL preparation materials, not all. The one point she and Wadden and Hilke agreed on was that research should be carried out into the efficacy of TOEFL preparation materials. They did not agree on who should carry out the research. Wadden and Hilke were in favour of independent researchers, but Hamp-Lyons challenged the notion that any single researcher could be truly independent, suggesting that the Teaching English to Speakers of Other Languages (TESOL) organization should be in charge of this endeavor.

Space limitations prevent further detailed discussion of the impact of tests on coursebooks, apart from brief references to several empirical studies emphasizing the trust that teachers and learners place in commercial materials. Wall (1999, 2005) described how a large sample of teachers in Sri Lanka preferred to use commercial test preparation books rather than design their own materials because they felt the authors of the commercial materials had privileged knowledge about the contents of the O-level examination. Roberts (2002) contacted eight TOEFL preparation institutions in Toronto and discovered that seven of them used TOEFL preparation coursebooks that "tended to reinforce a non-communicative approach to language education" (p. 84). The students trusted these books, however, believing that the authors were authorities. Finally, Zacharias (2005) found that teachers in Indonesia preferred to use coursebooks from international publishers rather than books produced locally, trusting the English of native-speaker writers more than that of local writers.

This brief survey has shown that despite movements within applied linguistics that question the appropriacy of preset syllabuses and that criticize the idea of using textbooks to determine what should be taught and how, there are also voices defending the teachers' desire for books that can ease their planning burden. It is also important to remember the publishers' desire and need to produce materials that teachers understand and feel comfortable using, even if this means sidelining advances in language or learning theory.

Research Questions

Given the findings concerning the importance of coursebooks in Phases 1 and 2 of the Impact Study and the attention they have received in the literature on test impact and washback, it was decided to devote Phase 3 to an analysis of the TOEFL preparation coursebooks being used in a sample of the original Impact Study teaching institutions. We learned at the start of this phase that the teachers we were working with were still teaching CBT classes and therefore

using CBT coursebooks, but they were also introducing TOEFL iBT courses and had either chosen or were in the process of choosing TOEFL iBT coursebooks. We decided to analyze both types of coursebooks to determine whether the TOEFL iBT books offered any type of content focus (language, skills, or other information) that was different from the CBT books. We also wished to see whether the TOEFL iBT books offered any means of presenting or practicing content that were different from the CBT coursebooks. We were particularly interested in means that might resemble the communicative or academic approaches mentioned by some of the expert advisors when they were questioned about intended test washback in Phase 1.

We also wanted to find out how teachers were reacting to their new coursebooks and whether the coursebooks were affecting the way they conducted their classes. It seemed possible to us that the new coursebooks might reflect the new test accurately but that the teachers might not understand the books or might use them in inappropriate ways. If the coursebooks reflected the test well and the teachers used them as intended then the coursebooks would be serving as an effective link in the process of creating test washback. If the coursebooks did not reflect the test well and/or if the teachers did not use them in the way intended, this outcome would weaken the potential of the new TOEFL to cause positive changes in test preparation classrooms.

Our research questions were as follows:

- Do TOEFL iBT coursebooks differ from CBT coursebooks in terms of their content focus?
- Do TOEFL iBT coursebooks differ from CBT coursebooks in terms of the means used for presenting and practicing content?
- How do teachers react to the TOEFL iBT coursebooks?
- Do the TOEFL iBT coursebooks affect the way the teachers deliver the test preparation classes?

Methodology

Sample of participants. We worked with four teachers in Phase 3. Three of these teachers had participated in both Phases 1 and 2. They were joined by another teacher who had participated in Phase 1 only. This teacher was interviewed and observed after we submitted our Phase 1 report, when the TOEFL Research Subcommittee asked us to visit a Western European country to see whether any difference could be noted between practice there and in the Central

and Eastern European countries we had visited earlier. The teacher could not do the Phase 2 work as he was still doing his Phase 1 work when Phase 2 was taking place with other teachers.

It is important to comment on the decline in sample size in every phase of the Impact Study, including Phase 4, when the number of teachers went down to three. The decline in the number of teachers reflects the nature of much English language teaching (ELT) teaching in the private sector, with its fluctuating demand for courses, sometimes difficult teaching conditions, and a transient population of teachers. Our numbers dropped from 12 in (the extended) Phase 1, to 6 in Phase 2, 4 in Phase 3, and 3 in Phase 4. We were, in fact, quite pleased that so many teachers were able to stay on with us until the end of the study, especially since those who remained had proven to be good informants, providing sensitive and coherent responses to our interviews and the other tasks we set them.

As in previous phases, the teachers were paid for the time and effort they put into the study.

Table 2 presents the details of the sample.

Sample of institutions. We were asked at the start of the Impact Study to concentrate on countries in Central and Eastern Europe. We defined this region as countries that had been members of the former Soviet bloc or that had opened up since the fall of the Berlin Wall. Teacher 1's (T1), Teacher 2's (T2), and Teacher 3's (T3) institutions were located in this region. We were later asked by ETS to add a country in Western Europe. Teacher 4's (T4) institution was located in this region. All four institutions were located in either the capital city of their country or a major town.

All the institutions were operating in the private sector. They offered a range of language courses, both for general language development and test preparation. The test preparation courses were aimed mainly at adult learners wishing to gain a particular TOEFL score, usually, but not exclusively, for studying abroad. T1's institution was also a national education information center for students aiming to study in the United States. T3's institution was also an information center and a Prometric testing center. T4's institution was the largest of the three and was part of a larger educational institution (a private university) offering courses in a wide variety of subjects. T1's and T4's institutions both became TOEFL testing centers in late 2006 or early 2007, between our Phase 3 data-gathering activities and the start of Phase 4.

Table 2

Phase 3—Teacher Details

Teacher ID	Gender	Age (approx)	Native (N) or nonnative English speaker (NNS)	Years teaching English	Years teaching TOEFL	Highest academic qualification	Type of institution
T1	F	20s	NNS	4	4	University graduate, and teaching qualification	Language school, national education information center, and TOEFL testing center (from Phase 4)
T2	F	30s	NNS	8	5	University graduate, and teaching qualification	Language school
T3	F	40s	NNS	26	11	University graduate, and MA in teaching arts	Language school, information center, Prometric testing center
T4	M	30s	NNS	16	11	University graduate, and MA in English language teaching	Language school, affiliated with a private university, and TOEFL testing center (from Phase 4)

Note. T1, T2, T3, and T4 = Teacher 1, Teacher 2, Teacher 3, and Teacher 4.

Sample of coursebooks. The teachers were asked to send us a list of the coursebooks they were using to prepare students for the CBT and a list of the coursebooks they planned to use once they began offering TOEFL iBT preparation classes. We received details of 14 coursebooks all together: 8 for CBT and 6 for TOEFL iBT. The details of all the coursebooks are presented in Table 3, along with information about which teachers (T1 to T4) were using or were planning to use each coursebook. (Note: Each coursebook was assigned a code number [see Column 1], and we refer to code numbers rather than book titles throughout the rest of this report.)

The coursebook analysis. The coursebook analysis was carried out using a framework that we designed for the purpose. We drew on a number of sources in order to decide which elements of the coursebooks we should be examining and describing.

The first set of sources were analyses that we had carried out in Phase 1 when we investigated what sorts of impact the new TOEFL test was meant to have and how this test differed from the earlier versions of the TOEFL. We had gathered information about intended impact from the TOEFL 2000 framework documents and from a survey of experts who had served as advisors during the test development process. We also used a table that we had drawn up in mid-2004 when we compared all three versions of the test: the paper-based version (PBT), the CBT, and the TOEFL iBT. What we were trying to determine at that time was whether the new version was really very different from the earlier versions and whether the new elements in its design had the potential to cause changes in classroom teaching. The results of these two surveys and the comparative table can be found in Wall and Horák (2006, pp. 126 and 136–143). We incorporated the points we considered relevant into the language skills sections of our coursebook analysis framework. (Further details about the framework are given below.)

The second set of sources we consulted were publications in the field of materials evaluation: Bonkowski (1996), Breen and Candlin (1987), Cunningsworth (1984), Dudley-Evans and Bates (1987), Ellis (1997), Garinger (2001), Hutchinson (1987), Hutchinson and Waters (1987), Littlejohn (1998), Miekely (2005), Skierso (1991), and Williams (1983). All of these sources offered ideas for describing and evaluating materials (including coursebooks) and helped us not only to build up the language skills sections of our framework (listening, reading, writing, and speaking), but also to design sections dealing with the treatment of grammar, supplementary

Table 3

Phase 3—Coursebooks Analyzed

Code	Publisher	Author	Title	Teachers			
				T1	T2	T3	T4
CBT1	Arco	Sullivan, P. N., Brenner, G. A., & Zhong, G. Y. Q. (2003)	*Master the TOEFL CBT 2004*			✓	
CBT2	Barron's	Sharpe, P. J. (2001)	*How to Prepare for the TOEFL* (10th ed.).	✓			✓
CBT3	Cambridge University Press	Gear, J. & Gear, R. (2002)	*Cambridge Preparation for the TOEFL Test* (3rd ed.)		✓		✓
CBT4	Kaplan	Shanks, J. (2004)	*TOEFL CBT Exam* (3rd ed.)		✓		✓
CBT5	Longman	Philips, D. (2001)	*Longman Complete Course for the TOEFL Test—Preparation for the Computer and Paper Tests*	✓	✓		
CBT6	Macmillan	Mahnke, K. M., & Duffy, C. B. (1996)	*Heinemann ELT TOEFL Preparation Course*				✓
CBT7	Peterson's	Rogers, B. (2003)	*TOEFL CBT Success 2004*		✓	✓	
CBT8	Princeton Review	Miller, G. S. (2002)	*Cracking the TOEFL*	✓			✓
iBT1	Kaplan	Hudon, E., Clayton, I., Weissgerber, K., & Allen, P. (2005)	*TOEFL iBT With CD-Rom*		✓		
iBT2	Pearson Education	Philips, D. (2006)	*Longman Preparation Course for the TOEFL Test: iBT*	✓	✓	✓	
iBT3	Pearson Education	Solórzano, H. (2005)	*NorthStar: Building Skills for the TOEFL iBT –High Intermediate*.	✓			✓
iBT4	Pearson Education	Fellag, L. R. (2006)	*NorthStar: Building Skills for the TOEFL iBT – Advanced*	✓			✓
iBT5	McGraw Hill	Educational Testing Service (ETS; 2006)	*The Official Guide to the New TOEFL iBT*	✓	✓	✓	✓
iBT6	Thomson Heinle	Rogers, B. (2007)	*The Complete Guide to the TOEFL Test: iBT Edition*			✓	

resources, and teachers' guides. Bonkowski's (1996) instrument was particularly useful as it had been designed for use with coursebooks preparing students for the International English Language Testing System (IELTS) examination, a test whose purpose is similar to the TOEFL's.

The third set of sources was work published by Hilke and Wadden (1997), Hamp-Lyons (1998), and Wadden and Hilke (1999). This work addressed issues directly related to TOEFL coursebooks. Hilke and Wadden provided a detailed survey of 10 coursebooks being used in the mid-1990s. The authors examined the grammatical structures and question types the coursebooks offered and compared these with the coverage of grammar and the distribution of item types in the structure and written expression section of the PBT. We could not find a similar analysis for the CBT, so we used this analysis to build up the structure component section in our own framework. The debate between Hamp-Lyons and Wadden and Hilke concerning the ethicality of TOEFL coursebooks prompted us to include a section on this theme at the end of the framework. We included one of the two measures Hamp-Lyons had employed: a 7-point scale proposed by Mehrens and Kaminski (1989). We did not share Hamp-Lyons' concerns about whether test preparation coursebooks were doing a disservice to learners, but we did wish to see whether the coursebooks we were examining would fall into the same category as the coursebooks that she examined.

The finished framework contained eight sections, which are presented in Table 4. The framework was trialed and revised, then used to analyze all 14 books in our sample. We analyzed the CBT books first, both because we needed this information in order to interpret the responses the teachers gave to our Task 1 (see below) and because it was not until several months into Phase 3 that the teachers were able to give us the details of the TOEFL iBT books they would be using. (See Wall & Horák, 2008, for an account of how coursebooks were late in appearing in several of the countries in our sample.)

The second stage in the process was to transfer certain features from the individual coursebook analyses to a composite table, which would allow a comparison of all the coursebooks together. The features we transferred related to new elements in the TOEFL iBT test (e.g., length of reading passages similar to TOEFL iBT). This table would allow us to see what the coverage was in both the CBT and the TOEFL iBT books for elements that were considered to be innovative in the TOEFL iBT test. (This table is presented as Table 6.)

Table 4

Details of the Coursebook Analysis Framework

	Name	Number of questions	Contents
A	Bibliographic details	8	Title, author, publisher, year, edition, ISBN, number of pages, which version of TOEFL the book is intended for
B	Overview	23	Basic structure: number of units, organization of units, content of teacher's notes, number and type of practice tests, information on scoring, self-study features, test-taking strategies, grammar reference section, other support features (e.g., , webpage), information about the test and how to apply, other features (e.g., tutorial, CD users' guide)
C	Listening component	23	Characteristics of input (e.g., number of passages, length, authenticity), characteristics of tasks (e.g., context provided, number of questions, time limits, purpose given, question types), other features (strategies for building subskills, test-taking strategies, using CD)
D	Structure component	14	Amount of grammar, types of grammar covered, question types, test-taking strategies, recycling,
E	Reading component	21	Characteristics of input (e.g., number of passages, length, difficulty level, topics, genre, sources, glossaries), characteristic of tasks (e.g., number of questions, time limits prereading exercises, purpose given, subskills practiced, question types), other features (strategies for building subskills, test-taking strategies)
F	Writing component	25	Characteristics of input (e.g., nature of prompts, nature of tasks, existence of reading and listening input), characteristics of tasks (e.g., skill-building v practice exercises, do tasks resemble TOEFL tasks, length of output, type of output), other features (pair work or group work, model answers, criteria for assessing writing, test-taking strategies)
G	Speaking component	31	Characteristics of input (e.g., nature of prompts, existence of reading and listening input), characteristics of tasks (e.g., number of tasks, skill-building v practice exercises, specialist knowledge required, work on pronunciation, which variety of English allowed, models of desired output)
H	Other general features	12	Treatment of vocabulary, existence of answer keys and explanations for right and wrong responses, work on note-taking, nature of CD, increase in demands as students proceed through book
I	Overall evaluation	3	Balance between TOEFL information and general language development work, accuracy of reflection of TOEFL, ethical versus unethical test practice scale used by Hamp-Lyons (1998)

The third stage was to transfer other features from each of the individual analyses to a second composite table. This would help us to see whether the TOEFL iBT coursebooks represented a different sort of teaching approach than that represented in early TOEFL preparation coursebooks. Courses for the PBT and CBT had been characterized as dry and predictable, focusing mainly on the development of test-taking techniques and test practice (Alderson & Hamp-Lyons, 1996; Wall & Horák, 2006). One of the reasons given for revising the TOEFL was to generate positive washback on teaching (Wang et al., 2008, p. 42). Given the dependence on coursebooks that teachers had shown in Phases 1 and 2, it seemed important to find out whether the coursebooks that they were likely to be depending on in the future displayed any features beyond simple test preparation and practice. The results of this analysis are presented in Table 7.

Consulting the teachers. The second part of the data collection involved consulting the teachers in our sample to probe their understanding of and attitudes toward the coursebooks they were using and to learn about the role these coursebooks played in their classroom teaching. We believe that even the coursebooks that most faithfully mirror tests will only be successful mediators of test washback if teachers understand the messages they convey, feel favorably disposed to the messages, and are able to take them up in their classrooms. This view stems from the experience of one of the researchers during an investigation into washback in another setting (Wall, 1996, 2000 and 2005). It became clear to them that many teachers could not respond appropriately to changes in a new curriculum and the accompanying high-stakes examination because (a) they did not fully understand the view of language skills underlying the examination, (b) what they did understand did not necessarily correspond to their own view of language, and (c) they did not have the technical skills or resources to teach in ways that were different to what they were used to. Chapman and Snyder (2000) reviewed other studies with similar findings, and note that there is no direct link between the introduction of a new test and improved teaching or learning. They proposed a model of "linkages" between high-stakes tests and instructional practice, which includes not only resources (e.g., appropriate materials), but also cautions against assuming that "teachers and school administrators will know how to channel the additional resources. . . in ways that will improve instruction to levels that can be detected on a standardized test" (p. 466).

We used three methods for gathering data from the teachers: tracking questions, tasks, and computer-mediated interviews. Table 5 presents the timing of the data-gathering activities.

Table 5

Phase 3—Data Collection Activities

Month (2006)	Type of activity
April	Tracking questions—Set 1
May	
June	Task 1
July	Interview 1
August	
September	Tracking questions—Set 2
October	Task 2
November	Interview 2

Tracking questions. We sent out two sets of tracking questions: the first at the start of Phase 3, in April 2006, and the second halfway through the phase, in September 2006. The purpose of the first set of questions was to find out whether the teachers' teaching situation had changed since we were last in contact with them (12 months earlier in the case of Teachers 1, 2, and 3, and 18 months earlier in the case of Teacher 4). This set of questions covered the following topics:

- Whether the TOEFL iBT had been launched in their countries
- How many and what types of TOEFL courses (CBT or TOEFL iBT) they had taught since we were last in contact with them
- Which materials (not restricted to coursebooks) they were using in their TOEFL iBT courses, who selected them, and why
- Whether they had received any training to support their TOEFL iBT teaching
- Whether their TOEFL iBT students had any worries about the test
- Whether they themselves or their institutions had any worries about the test

The second set of questions covered these topics:

- Whether the TOEFL iBT had been launched yet (it had not reached any of the countries at the beginning of Phase 3)
- Whether the teachers had received any new information about the test and whether this had affected their teaching
- How many and what types of TOEFL courses (CBT or TOEFL iBT) they had taught since April
- What challenges they had faced since the launch of TOEFL iBT in their country
- Which coursebooks they were using in the TOEFL iBT courses and which aspects were helpful or problematic
- What they thought about the TOEFL iBT courses they were teaching, and why
- Whether they had received any support (of any kind, not just training) to help them with their TOEFL iBT teaching
- What (if anything) their former students had told them about the TOEFL iBT test
- What their opinions were about TOEFL iBT and whether they had any worries related to the test

We sent the tracking questions to the teachers by e-mail, and they sent their responses back in the same way. We wrote back to them as necessary to ask for clarifications and for information they might not have provided.

Tasks. The second means of gathering data from the teachers was to set tasks for them that would help us to understand how they used their coursebooks in their classrooms. The best way of collecting this information would have been to observe the teachers in action, but budget restrictions made it necessary to find an indirect way of investigating their practice. Teachers 1, 2, and 3 were used to working with tasks as we had given them five tasks to do in Phase 2. We explained this way of working to Teacher 4, who had not been part of Phase 2, and he was confident that he, too, could participate in this type of activity.

Task 1: Attitudes toward and use of TOEFL computer-based test (CBT) coursebooks. The first task was sent to the teachers in June 2006. The purpose of the task was to find out about the teachers' attitudes toward and their use of CBT coursebooks. Although the aim

of the Impact Study was to find out about the possible effects of the TOEFL iBT on teaching, including on coursebooks, we believed it was necessary to establish what teachers thought about their CBT coursebooks and how they used them so that we had a point of comparison when we talked about the role of the TOEFL iBT coursebooks in their courses. Without this point of comparison it would be difficult to support statements that the TOEFL iBT had provoked change in classroom practices.

Task 1 was in four parts. The first part contained questions about how the teachers had chosen their coursebooks and what they considered the positive and negative features of each coursebook to be. The second part asked the teachers to describe how they would use specific sections of their coursebooks when preparing students for different sections of the CBT. They were to identify, for example, a few pages they might use to prepare students for the listening test, explaining how they would present the material, what they would do, and what their students would be expected to do in the lesson. The third part of the task asked the teachers to give specific details about this lesson, if they had not already given them (e.g. student interaction patterns, how students would check their responses to exercises, the type of feedback the teachers would give), and they were asked directly how closely they would follow the coursebook (would they follow it exactly, add material, delete material, alter material, etc.). The fourth part asked them how much the TOEFL preparation coursebooks influenced their teaching.

Task 2: Attitudes toward and use of TOEFL iBT coursebooks. We originally envisaged Task 2 as a replication of Task 1 but focusing on TOEFL iBT coursebooks instead of CBT coursebooks. However, the replies the teachers gave to our questions in Task 1 were briefer than we had hoped for, giving us the impression (which we also sometimes had in Phase 2) that they were not used to reflecting on or analyzing their own teaching without considerable prompting. We decided to ask them more direct questions in Task 2 and to relate these queries to concrete teaching situations.

Task 2 was in three parts. The first part asked the teachers to send plans for three lessons they were actually intending to teach (as opposed to descriptions of lessons they might possibly teach, as in Task 1). They were given specific questions to answer (e.g., what content would they cover, what sorts of activities would they organize, what sorts of interaction would they encourage, which materials would they use, which resources would they draw on, and which other factors would they take into consideration when planning). The second part asked them to

send us a description of what actually happened when they taught the lessons, including comments on how they used their materials, who did what during each activity, whether the lessons went as expected, and whether they were satisfied with the lessons and the materials. The third part asked them to choose one section of one of the lessons they had taught—a section that represented the role that their coursebooks typically played in their TOEFL teaching. They were also asked to choose a metaphor to represent the role of the TOEFL preparation coursebook in their classes. They had to complete this sentence: *The TOEFL coursebook is. . .* with one of the following ways of completing it: *an instruction manual for a piece of equipment, a recipe book, the instructions for how to assemble/build something, a bible,* or *a reference book.* They could choose another metaphor if they preferred.

Computer-mediated interviews. The third means of collecting data from the teachers was through long-distance interviews, using MSN Messenger. (See Wall & Horák, 2008 for the rationale for using computer-mediated communication.)

We conducted two main interviews with each teacher. The first was several days after they had completed Task 1, and the second, several days after they had completed Task 2. We studied the responses they had given to the task and then formulated questions that would help us to understand what they were telling us when the meaning was not clear.

Teacher data analysis. The teachers' responses to our tasks and the MSN Messenger interviews were in written form: 51 electronic files containing 80,139 words in all. All the data were loaded into Atlas–ti, the same qualitative data analysis package that was used in Phases 1 and 2 of the Impact Study.

The coding scheme was based on Henrichsen's (1989) *diffusion/implementation process*, a framework that divides the process of innovation into three stages (the antecedent situation, the process itself, and the consequences of the process) and shows how factors within an innovation (in this case, the new TOEFL test) and other factors within the context work together (or do not, as the case may be) to produce consequences in the educational system. We had used the Henrichsen framework from the start of the Impact Study, as we saw the introduction of a new test with the intention of creating positive impact as an instance of introducing an innovation into an education system with the intention of creating positive change. Phase 1 of the Impact Study was a description of what Henrichsen called the *antecedent situation* (we used the term *baseline study*), and Phases 2 and 3 aimed to document the factors affecting the *process* part of his model.

To the 215 codes that were used in Phases 1 and 2, 68 new codes were added in Phase 3. These codes related to the aims of the new TOEFL iBT courses, the content and teaching methods being used, the coursebooks that the teachers were using (both CBT and TOEFL iBT), the teachers' views of the courses and the coursebooks, and the challenges they were facing as they made the transition from CBT to TOEFL iBT preparation work. (See Appendix B for a list of the codes that were introduced in all the phases.)

All of the data were coded by both researchers, who first worked independently and then discussed their results to further refine the coding scheme. We did not calculate the degree of inter-rater agreement, but there were few instances in which we differed in our understanding of what the teachers meant to say. This was due both to the fact that we had developed the codes over several years of working together and discussing their definitions frequently and to the nature of the questions and the tasks in this phase, which produced mainly factual and narrative information.

Analysis of Coursebooks

Content. Table 6 presents the results of our analysis of the content of the 14 coursebooks. The first column presents features that were announced as being new in the TOEFL iBT (apart from the penultimate row, which relates to the treatment of grammar in isolation—a feature of CBT). We analyzed all of the CBT and TOEFL iBT coursebooks separately, but we found that there were no differences within the group of CBT coursebooks or within the group of TOEFL iBT coursebooks, so we presented the results under two headings only (CBT and TOEFL iBT). The presence or absence of TOEFL iBT features is shown by checkmarks (✓) or crosses (X) respectively.

What Table 6 shows is a clear mirroring of TOEFL iBT features in the TOEFL iBT coursebooks and an absence of these features in the CBT coursebooks. The most striking difference between the CBT and TOEFL iBT coursebooks is that the CBT books do not present speaking tasks or criteria for judging speaking and integrated writing tasks or activities for developing note-taking skills. Also important is the absence of grammar sections in the TOEFL iBT books (see other notable features at the bottom of Table 6). Only one of the books included exercises on grammar, but these were in an appendix rather than in the book itself.

Table 6

Analysis of Coursebooks—Presence or Absence of TOEFL iBT Features

Features	Coursebooks	
	TOEFL CBT $n = 8$	TOEFL iBT $n = 6$
Reading		
Length of reading texts similar to TOEFL iBT (600-750 words)	X	✓
Paraphrasing is tested	X	✓
Some words are glossed in reading texts	X	✓
Listening		
Listening section includes longer (than in CBT) conversations of 3 minutes approx.	X	✓
Listening section includes no short (2-turns) dialogues	X	✓
Varied native English accents included (not only North American)	X	✓
Pragmatic understanding is tested	X	✓
Speaking		
Speaking skills included	X	✓
Independent speaking tasks resemble TOEFL iBT tasks (prompt leading to monologue)	X	✓
Integrated speaking tasks resemble TOEFL iBT tasks	X	✓
TOEFL iBT criteria for scoring speaking are described (scale 0-4)	X	✓
Writing		
Integrated writing tasks resemble TOEFL iBT tasks	X	✓
TOEFL iBT criteria for writing described (scale 0-5, not 0-6 as in CBT)	X	✓
Integrated tasks		
Note-taking skills are included	X	✓
New question types		
Listening: Excerpts from the passage are replayed before the question is given	X	✓
Listening and reading: completing category or summary charts (table)	X	✓

Features	Coursebooks	
	TOEFL CBT *n* = 8	TOEFL iBT *n* = 6
Information for students about TOEFL iBT		
Listening is no longer computer-adaptive	X	✓
Note-taking is allowed	X	✓
Candidates *must* type written responses	X	✓
In integrated tasks candidates can see reading passage on screen during time for writing response	X	✓
Suggested length of writing task is 300 words	X	✓
Other notable features		
Grammar section is included (this is a feature of CBT)	✓	X
Practice tests look like iBT (papers, order of papers, length of input texts, time allowed, output expected, etc.)	X	✓

Note. ✓ = presence of Internet-based features, X = absence of Internet-based features, CBT = computer-based test.

The only TOEFL iBT feature that did not appear in the TOEFL iBT coursebooks was a range of native-speaker English accents. ETS announced early on that the TOEFL iBT would include a variety of native accents in the future, not just North American accents as in previous versions of the TOEFL. The coursebook publishers did not seem to pick up on this feature, however, perhaps because the practice tests available on the ETS Web site at that time included only North American accents. As the Web site provided the only official guidance available (no detailed specifications were available to the public), it seems logical that publishers would have followed this model when producing their preparation materials.[1] Overall, then, the new coursebooks seemed to reflect accurately the content of the test they represented.

Approach to teaching. As stated earlier, the second analysis examined the means the coursebooks used to present and practice language and language skills. Our reason for looking at this aspect of the coursebooks was to respond to the concerns expressed in the framework documents about the effects of earlier versions of the TOEFL on teaching (e.g., "that discrete-point test items, and the exclusive use of traditional, multiple-choice items to assess the receptive

skills, have a negative impact on instruction" [Jamieson et al., 2000, p. 3]), to the hopes expressed in statements like "TOEFL preparation courses will more closely resemble communicatively orientated academic English courses" (Bejar et al., 2000, p. 36), and to the expectations that "research can be designed to investigate washback effects on what examinees study and to determine whether the emphasis on communicative learning increases once the new test is operational" (Cumming et al., 2000, p. 49). Unfortunately no definitions were given for *communicative* in the frameworks. As language teachers and teacher educators, we were well aware that the definition of communicative was infinitely expandable, meaning different things to different people. Richards and Rodgers' (2001) survey of approaches in language teaching made clear how varied the factors are that can be appealed to when deciding whether teaching qualifies as communicative or not: theories of language, theories of learning, program design factors such as objectives, types of syllabus, types of learning and teaching activities, learner roles, teacher roles, the roles of materials, and so on. Such diversity led Richards and Rodgers to declare that "there is no single text or authority on it, nor any single model that is universally accepted as authoritative" (p. 155).

Our own view of a communicative language approach included notions such as focusing on meaning as well as form; developing sociolinguistic, discourse and strategic competences as well as linguistic competence; and negotiating meaning through interaction. We were interested in other features as well, but we felt that it was unrealistic to expect many of these characteristics in test preparation coursebooks. We were influenced in this regard by the views reported by Waters in an earlier draft of his 2009 article (reviewed above), namely, that it was unlikely that many advances in the academic conceptualization of language or language learning would appear in commercial coursebooks because publishers would not be sure they would be acceptable to teachers, who have their own specific needs and constraints. We also knew from our work in Phases 1 and 2 that TOEFL preparation teachers felt pressured to provide the type of teaching that would, in the eyes of their students, be directly related to their goal of doing well on the test, with no unnecessary distractions. We therefore adopted a conservative view of what positive impact might mean in presentation and practice terms, looking for points that would have some relationship with the notion of communicative competence and communication but that would be readily appreciated by the type of students we observed in Phase 1: instrumentally driven, with little time for expressing their own meanings or negotiating meanings with others,

and desiring quick returns for the investment they were making by enrolling in a test preparation course. The features we looked for are listed in Column 1 of Table 7.

Table 7

Analysis of Coursebooks—Means Used to Present and Practice Language

Features	Coursebooks	
	TOEFL CBT	TOEFL iBT
	Listening	
1. Context established before listening	Not included in any of the CBT coursebooks	Included in iBT3 and iBT4, but not in the other TOEFL iBT coursebooks
2. Listener asked to predict content of passage	Not included in any of the CBT coursebooks	Not included in any of the TOEFL iBT coursebooks
3. Questions provided prior to exercise	Included only in CBT4 (Students may look ahead in the other books if they wish, but the intention is that they should not—the same as in the test.)	Not included in any of the TOEFL iBT coursebooks (Students may look ahead if they wish, but the intention is that they should not—same as in the test.)
4. Free (not controlled) exercises included	Not included in any of the CBT coursebooks	Not included in any of the TOEFL iBT coursebooks.
5. Strategies for building subskills	Included in 5 of 8 CBT coursebooks	Included in all TOEFL iBT coursebooks
	Reading	
6. Context: Source of texts obvious/ stated	Not included in any of the CBT coursebooks	Included in iBT3 and iBT4, but not in the other TOEFL iBT coursebooks
7. Reader asked to predict content of text	Not included in any of the CBT coursebooks	Included in iBT3 and iBT4, but not in the other TOEFL iBT coursebooks
8. Free (not controlled) exercises included	Not included in any of the CBT coursebooks	Not included in any of the TOEFL iBT coursebooks
9. Strategies for building subskills	Included in 5/8 of the CBT coursebooks, though judged not to be very helpful in 2 cases	Included in 3/6 of the TOEFL iBT coursebooks

Features	Coursebooks	
	TOEFL CBT	TOEFL iBT
	Writing	
10. Skill-building exercises	CBT1 consisted only of practice tests, so no skill building possible. Included in 5/7 of the other CBT coursebooks	Included in 4/6 of the TOEFL iBT coursebooks
11. Kind of written responses required made clear	Included only in CBT5	Included only in TOEFL iBT3
12. Work in pairs/ groups suggested	Included only in CBT5	Included in 3/6 of the TOEFL iBT coursebooks
	Speaking	
13. Skill-building exercises	Not included in any of the CBT coursebooks	4/6 TOEFL iBT coursebooks included such exercises
14. Kind of spoken responses required made clear	No work on speaking	Not included in the TOEFL iBT coursebooks, apart from in general terms
15. Work in pairs/ groups suggested	No work on speaking	Included in 3/6 of the TOEFL iBT coursebooks
	Grammar	
16. Grammar dealt with throughout the book (not just one section)	Grammar generally dealt with in separate section	Grammar exercises included in iBT2, but in an appendix rather than the main book itself
17. Exercise types beyond those in TOEFL	Included in 3/8 of the CBT coursebooks	Not applicable, as TOEFL iBT does not have a grammar section
18. Recycling of grammar points	Grammar generally not recycled	Grammar not recycled
	Vocabulary	
19. Exercises/ tasks to develop vocabulary	Included only in CBT4	Included only in iBT6
20. Vocabulary recycled across units	Vocabulary not recycled	Vocabulary not recycled

Features	Coursebooks	
	TOEFL CBT	TOEFL iBT
21. Advice on how to develop depth/ breath of vocabulary	Included in 4/8 of the CBT coursebooks, though judged to be minimal in 2 cases	Included only in iBT6
	Other	
22. Explanations of all suggested responses (correct and incorrect) provided	Included in 4/8 of the CBT coursebooks	Included in 3/6 of the CBT coursebooks
23. Study support materials included—e.g., study plans/schedules, information on colleges	Included in 4/8 of the CBT coursebooks	Included only in iBT2

Note. Coursebooks are identified in Table 3. CBT = computer-based test.

The list consisted mainly of features that could help the students to develop the strategic element of communicative competence (Canale & Swain, 1980). For listening and reading, we asked whether the students were given practice using contextual features to anticipate or disambiguate language (Features 1 and 6), whether they would be encouraged to use their background knowledge to predict what they might hear or read (2 and 7), and whether they would be given questions ahead of listening to allow them to listen selectively (3). Taylor and Angelis (2008) wrote that "many individuals were dissatisfied because of the perceived negative effects of the multiple-choice TOEFL on language instruction" (p. 48). It was this expression of dissatisfaction that led to the inclusion of Features 4 and 8, regarding whether students would be able to expand their responses (exercising creativity, or perhaps risk-taking) rather than being restricted by the question formats found on the test. We included strategies for building sub-skills (5 and 9) in response to Hamp-Lyon's (1998) concerns that test preparation books often only assessed whether students could answer testlike questions rather than help them to develop the abilities they needed to do so.

We also included skill-building exercises (10 and 13) under writing and speaking, for the same reasons given above. The other features listed for writing and speaking are making clear

the kind of response required (11 and 13), which relates to sociolinguistic and discourse competence, and working in pairs and groups (14 and 17), which relates, if only in a limited way, to negotiation of meaning.

We were initially interested in the way that grammar and vocabulary were presented in both the old and the new coursebooks. We had noticed during earlier phases of the study that grammar was often dealt with in an isolated way, in a separate section of the coursebook rather than integrated with skills work throughout the book. It seemed unusual to find exercise types that did not mimic the item types on the test. We decided to check whether the new coursebooks provided any other approach to working with this aspect of language (16 to 18) and to check whether the coursebooks offered any developmental work for vocabulary (10 to 21). These issues were not central to the notion of communicativeness; however, notions such as integration of form and use, and recycling of language points, are generally considered useful features in modern language teaching approaches (Willis, 2008).

The final features in the framework (22 to 23) have to do with the support offered by the coursebooks to the students and the teachers. Feature 22 would help the learner to benefit from his or her wrong responses by offering explanations for why they were wrong. Feature 23 would help teachers by giving them extra information about language features or teaching methods. Neither of these features are exclusive to any particular approach to teaching, but we included them here as features that could enrich the learning experience beyond the monotonous routine we observed in many classes in Phase 1—consisting only of familiarization with test formats, answering exercises, and noting whether the answers were correct or incorrect.

We carried out a detailed analysis of all eight CBT coursebooks and all six TOEFL iBT coursebooks, and then summarized what we found for each type of coursebook.

What Table 7 indicates is that CBT coursebooks and the TOEFL iBT coursebooks did not differ greatly in terms of the approach they took to presenting and practicing language and skills content. The content itself differed, as we saw in Table 6, but the coursebooks dealt with it in quite similar ways. Under listening, for example, neither type of coursebook paid much attention to establishing a context for listening (only iBT3 and iBT4 did this), asking listeners to predict what they would hear, or encouraging them to read the questions before they heard a passage so that they could listen purposefully (only CBT4 did this). None of the coursebooks included exercises where student could express their own ideas rather than responding to controlled exercises. The

only notable difference was in the percentage of books offering strategies for building subskills: Only 5/8 (63%) of the CBT books did this, as compared to 100% of the TOEFL iBT books.

The pattern for reading was similar. Again, neither type of coursebook paid much to attention to establishing a context for reading (only iBT3 and iBT4) or asking readers to predict what they would be reading (again, only iBT3 and iBT4). None of the books included exercises where students could respond freely rather than in a controlled way. Little difference was found in the percentage of books offering strategies for building subskills, with 5/8 (63%) of the CBT books doing this, as opposed to 3/6 (50%) of the TOEFL iBT books.

Under writing, the percentage of books offering skill-building exercises was similar on both sides, and only one book on each side made clear what kind of writing response was required. A difference was noticed, however, in the type of interaction suggested for writing exercises: Half the TOEFL iBT books included suggestions for students to work in pairs or groups, while only one of eight CBT books did this.

We have already seen that none of the CBT books offered speaking exercises. Of the TOEFL iBT books, two-thirds provided skill-building exercises and half included suggestions for students to work in pairs or groups.

We have also already seen that little separate teaching of grammar was included in the TOEFL iBT coursebooks. The only TOEFL iBT book that included grammar exercises presented them in an appendix rather than in the main book itself. The question about whether there were any exercise types beyond those given in the TOEFL was not applicable as no grammar questions were included on the TOEFL iBT.

As for vocabulary, only one CBT coursebook and one TOEFL iBT coursebook included exercises on vocabulary, and neither recycled the vocabulary in other parts of the book. Advice about how to develop vocabulary depth and breadth was given in four (50%) of the CBT coursebooks, although it was judged to be not very helpful in two of the books. There did seem to be a difference in the CBT and TOEFL iBT coursebooks in this regard, as only one of the latter offered advice in this area.

To summarize this section then, although the CBT and TOEFL iBT coursebooks differed in content (Table 6), there did not seem to be a great deal of difference in the means used to present and practice language and language skills (Table 7). Only the iBT3 and iBT4 books stood out as representing a slightly different approach to teaching in that they encouraged

students to think about the context of the listening and reading they presented, and they included some work that required students to predict the content of the texts they were about to read. This finding suggested that if we saw differences in classroom teaching during the later stages of the Impact Study, they were likely to be in the content of the teaching rather than in the manner of presenting the content—if what the literature suggested about teachers' dependency on coursebooks for planning and conducting lessons proved true.

Teachers' Views of Coursebooks

The aim of this section is to examine what the teachers in our sample told us about the way they viewed the role of coursebooks in language teaching in general, the way they viewed coursebooks in TOEFL preparation courses, the reasons they had for selecting or rejecting particular coursebooks for their TOEFL iBT courses, and their reasons for not producing their own materials.

The role of the coursebooks in language teaching in general. We first wanted to establish what the teachers' views were on the role of coursebooks in general to see whether their views of the role of coursebooks in TOEFL classes followed logically from more fundamental beliefs they had or whether their views contradicted their beliefs in any way. In the first interviews (July 2006), we asked how they viewed the use of coursebooks in class in general.

The teachers fell into two groups with opposing opinions. T3 viewed coursebooks as a necessary evil." She used them because she thought her students felt more secure in a class organized around a coursebook. She also stated that it was a university requirement to have a coursebook so she had no choice but to use one (34:46. This reference and those that follow include the transcript number and the line number in which the information can be found. This reference is to Transcript 34, Line 46). In contrast, T1 and T2 felt positive about using coursebooks since these gave structure to courses and, according to T2, they could also offer guidance to novice teachers (21:30). T4 was also positive about using coursebooks but stated that no book was perfect and teachers always had to stay true to their objectives (44:21).

We wondered whether the teachers' attitudes might have been affected by what they had learned about coursebooks while they were training to be teachers. They all confirmed that this topic was on their training syllabus, but none of them could elaborate on what they had learned.

This lack of detail is not surprising considering that they had all been trained several (or in the case of T3, many) years earlier. The only comments they did make were similar in nature, with T4, for example, saying he had been advised that coursebooks were only a "tool to accomplish goals" (44: 49) and T3 reporting that she had been told to be selective in their use (34:10).

The role of coursebooks in TOEFL preparation classes. We next investigated how the teachers viewed the role of coursebooks in their TOEFL preparation classes. We first asked the teachers whether they had decided on the aims of their course first and then chosen their coursebooks, or whether they had chosen the coursebooks first and then designed their courses around them. The teachers were divided in their responses. T2 and T3 had chosen their coursebooks first and saw them as a core around which they designed their courses (T2, 17:199; T3, 30:230). Both teachers were working in small institutions and were the only TOEFL teachers on the staff. T1 was also the only TOEFL teacher in her institution, but she had decided on her aims first and only then chosen her coursebook. She may have been influenced by the CBT teacher who had served as her model when she began teaching TOEFL. That teacher had decided on her aims but was not able to find one coursebook that suited all her purposes. She ended up putting together a collection of materials from different sources, which she photocopied for the students. T1 was lucky enough to find a coursebook that was appropriate for what she wanted to achieve, with some supplementing (13:74). T4 was from a large institution and worked with a team of colleagues to design both the CBT and TOEFL iBT courses (38:97). He stated that for TOEFL iBT they "first set the goals which reflected our aim to prepare our students to deal successfully with the test," and then "tried to select the best book to fulfill our aim" (43:14).

Three of the teachers stated that their TOEFL iBT coursebooks were playing an important role in their actual teaching (T1, 7:4; T2, 20:4; and T3, 33: 4). One of the clearest functions the coursebooks served was providing the teachers with information about the test. T1, for example, stated:

> Ninety percent of what I know about the test is the knowledge acquired from the books used in the course, Internet and similar. The other 10% is the knowledge I gained from practical experience, my interaction with the students preparing, from observing them, thinking about ways to help them, learning to approach them and their weaknesses in the best way. (14:20)

T1 relied heavily on coursebooks when doing her lesson planning (13:86). T2 used the term *backbone* to describe their function in her teaching (20:15). T3 relied on her coursebooks for the answers to exercises, especially for reading and grammar practice (30:168). She complained, however, that "in many cases books limit my choice in lesson or topic selection" (33:25), and that using a coursebook was like "having another teacher in the classroom" (35:68). In other words, they could be intrusive.

T4 claimed not to be influenced by his coursebooks; nevertheless, he put great store in them, trusting the expertise of the authors (50:223). He felt that coursebooks were more necessary in examination preparation classes than in general classes, since the goals of preparation classes were so specific (44:2). He was also under more pressure in examination preparation classes, whereas in general classes he felt "more relaxed and perhaps more creative. . . to use materials chosen or even developed by me" (44:28).

Reasons for selecting or rejecting specific coursebooks. The teachers gave various reasons for selecting or rejecting specific coursebooks, as is shown below, but there were four themes that stood out as common across their explanations. The first theme, which had actually emerged in Phase 2 and was repeated in the early stages of Phase 3, was that they were not interested in using CBT preparation materials for TOEFL iBT courses (e.g., T2, 15:149; T3, 28:144; T4, 38:133). They did not consider that there was enough similarity between the two versions of TOEFL to make this strategy worthwhile. However, at least one of the teachers changed her mind by the end of Phase 3, feeling that CBT materials could usefully be employed to prepare students for the TOEFL iBT independent writing section (T2, 22:144).

The second general theme was that the teachers made a distinction between "theory" (explanations of what was being tested and how it would be tested) and practice material, and valued the latter over the former. T1 had rejected two coursebooks on the grounds that they contained too much theory and not enough practice material (1:08, 1:13), although she later used extracts from both of them. T2 noted repeatedly that a good TOEFL coursebook should include plenty of exercises, especially for reading and listening (17:139 and 157; 21:122, 170, and 204; 23:40; 27:147; 27:207 and 317). T3 made similar comments (T3, 30:123), adding that it was when coursebooks lacked practice material that teachers had to supplement them with other titles (30:424). T4 also noted that while he thought his coursebooks were very good, there was simply not enough practice material in them (39:52; 47:28; 48:61; 48:134; 48:206).

The third general theme was that the teachers looked favorably upon TOEFL iBT coursebooks if they had had a positive experience with CBT coursebooks from the same publisher. The iBT2 book had an advantage over others in this regard (T4, 45:67). The fourth theme was that the teachers respected books that had an endorsement from ETS. Here the iBT5 book had the advantage. T1 stated that she had "compared (iBT2) against (iBT5) primarily, as they are the test-makers after all" (14:282).

In some cases, however, the choice of coursebook was not in the individual teacher's hands. In T3's institution the director of studies ordered books from a publisher with whom they had a long-standing relationship, and the selection appeared to have been made on financial as much as pedagogical grounds (30:79, 34:163). In T4's case the coursebooks were selected by the director of studies and piloted by several teachers before being approved (44:151). T4's institution was the largest of the four being studied and it had enough resources (economic and human) to operate in this way. However, even here there were practical issues to consider. For instance, a coursebook that had otherwise been deemed excellent would not be used for the TOEFL iBT course as it was not possible to fit the contents into the 60-hour courses offered in the institution (39:65).

Table 8 shows in more detail the teachers' reasons for selecting or rejecting specific TOEFL iBT coursebooks. (Similar information was gathered about CBT coursebooks, but space restrictions do not permit an analysis here). The code numbers for the coursebooks are listed in the left-hand column. Note that the first book listed, iBT0, was not amongst those we analyzed earlier as the teachers did not have access to it until quite late in Phase 3. The code numbers for the teachers are given across the top of the table. The information in each cell begins with a note indicating whether the coursebook in question was selected for use or rejected by the teacher, and whether this decision was made at the start of or later in Phase 3. *At the start* means up to June or July 2006, and *later* means from that time up to November of the same year.

What Table 8 shows is that although some common themes held true across the teachers, some teachers had individual preferences that they might not have shared with other teachers. T1, for example, reacted quite negatively to the iBT1 coursebook, while the other teachers viewed it positively. T3 felt that the iBT5 coursebook did not help her to see the difference between lower- and higher-level speaking performances, while T4 felt that it gave a good picture

Table 8

Teachers' Reasons for Selecting or Rejecting Specific Coursebooks

Coursebook	T1	T2	T3	T4
iBT0 (This book was not analyzed in Phase 3 because it was not available to the teachers until late in study.)	Selected later focus on "new skills" that are useful for integrated sections of TOEFL—note- taking, paraphrasing, summarizing. (14:225)		Selected later useful for supplementing stock of practice tests (23:39)	Selected later useful for supplementing stock of practice tests (50:159)
iBT1	Rejected at start "weird and terrible" (1:11) poor reviews at start (1:23) not as focused on test as iBT2 (13:23) confusing layout and organization (8:374)	Selected at start clear explanations, presentation comprehensive (15:126) hoped to use as core book (15:120) but later decided to use iBT2. iBT1 used as source of extra practice material (22:134, 27:159)	Selected later good for skills development (36:266)	Doubts at start but reconsidered later too difficult for his students, but might be useful for extra practice material in the future (39:62)
iBT2	Selected at start widely available at the start trusted publisher because of experience of using CBT material (3:18) organized in logical fashion, so makes lesson planning easy (10:60) deals with question types well (13:11) Doubts: Easier than TOEFL? (4:343) though could be used with other books	Selected at start widely available at the start	Selected at start widely available at the start offers "a lot of materials, exercises and skills (tricks and strategies)" (3:22) more focused on TOEFL iBT tasks than other titles (3:22) organized in logical fashion (3:27) Doubts: Are explanations effective? (3:32) no grammar section, apart from	Considered at start, but rejected later widely available at the start trusted publisher because of experience using CBT book (45:67) Doubts: Easier than TOEFL? (44:79) too much for 60-hour course (39:65)

Coursebook	T1	T2	T3	T4
	treatment of integrated skills (14:248)		in appendix (35:123)	
iBT3 and iBT4: Two books in the same series. iBT3—High Intermediate iBT4—Advanced	Selected at start, but rejected later Different books for different levels (1:16) Useful for planning course (2:100) Doubts: can't use different books with mixed ability class (8:339) material highly integrated so difficult to use any one section on its own (10:93)			Selected at start Offers more than test preparation—e.g., prelistening and prereading activities, so it is unique amongst titles (49:51) Culture notes (49.21) Doubts: not enough material for whole course, so needs supplementing (39:56) models for note-taking—useful for integrated skills work (46:109; 46:199)
iBT5	Selected at start approved by ETS (14:282) "reliable and simple" (9:154) used as benchmark for judging other materials (10:61; 14:281) Doubts: poor reviews at start (1:23) lots of theory	Selected at start approved by ETS detailed information, clear to students (25:12) (later replaced iBT1)	Selected at start approved by ETS Doubts: hard to distinguish between high- and low-level speaking responses (29:48)	Selected at start approved by ETS good descriptions of test (46:140) useful as supplement to iBT0 and iBT4 (39:43 and 57) accurate picture of level expected of students (46:145)
iBT6			Considered at start, but rejected.	

Note. T1, T2, T3, and T4 = Teacher 1, Teacher 2, Teacher 3, and Teacher 4. Coursebooks are identified in Table 3.

of the level expected of students. What must be remembered here is that Phase 3 was a time of transition when the teachers were still learning about the TOEFL iBT and were trying out different coursebooks to see which ones would work in their own situation. What seem like contradictory views about some coursebooks might be natural, given the different perceptions that teachers had about the test, their beliefs regarding teaching, and even factors as seemingly unimportant as the order in which they inspected the individual coursebooks.

Although the teachers were generally positive about the TOEFL iBT coursebooks (e.g., T2, 25:55, 25:119; T3, 30:413), they recognized and made it clear that the books were not without their problems. T2, for example, reported that she had found mistakes in the answer keys of her book (17:269), and T3 was not convinced that one of her books was dealing with the same concept she had in mind when it referred to inferencing (34:277). T1 believed the problems she was having with her main coursebook were because it had (in her eyes) been produced very quickly, to arrive in time for the launch of TOEFL iBT (8:153). She did not see any of the books as a final product and expected all of them to improve in future editions (T1, 1:46).

Reasons for not producing their own materials. Given that all four teachers had considerable experience teaching TOEFL preparation courses and that they all believed there were flaws in their coursebooks, it would seem reasonable to expect them to have produced some materials themselves. None of them attempted to do so, however, apart from putting together the occasional handout, which they did not seem to view as materials production (T1, 2:120, 9:175; T2, 15:139; T3, 28:133, 35:173; T4, 38:124). Three teachers gave reasons that suggested a lack of confidence in their own abilities. T1 said she could not do a better job than the TOEFL coursebook writers (9:181; although she also said in response to another question that TOEFL teachers should be the ones who do coursebook writing, 8:152). T3 also felt that her materials would not be "of the same value" as those produced by the coursebook writers (30:162, 30:179). She did not think she had a good enough "feel" for the standard students were required to reach to succeed on TOEFL iBT, and she felt that her own variety of (which she judged to be more British than American) might cause problems for her students (34:313). She also said that she simply did not have enough time to write materials (35:180). It was surprising to hear her first two reasons, given that she was the teacher with the most teaching experience in our sample; however, she was also the teacher who seemed most able to reflect deeply on her teaching. The problem of not having enough time

would presumably be common to all teachers who work on a part-time basis in different institutions.

T2's institution was the only one where any extra material design activity had taken place. She had commissioned a friend to produce some computer software that would enable her students to experience tasks similar in format to the integrated tasks on TOEFL iBT (22:154).

We have mentioned that teachers took exercises from other coursebooks when they felt their main coursebook did not include enough practice material. They also used other books when they felt that they dealt with particular skills in a better way. T2, for example, did not like the writing section of one of her coursebooks and replaced it with the writing section of another (27:195). T3 did not test her students with the practice tests in her main coursebook as the students also had access to these and could assess them in their own time. When she wanted to test them she used practice tests from sources that they would not have such easy access to (30:108, 34:189).

A paradox. It should be clear from this discussion that all four teachers had given considerable thought to the question of coursebooks and that the decisions they made about which books to buy and which to use for each skill owed much to their own understanding of the requirements of the new TOEFL. What we found interesting here was that the teachers were in a "loop when it came to understanding what the requirements of the new test were, since their vision was shaped not only by information on the ETS Web site but by the very coursebooks they were consulting. None of the teachers seemed to see this as a problem, however.

How TOEFL Coursebooks Were Used in Classes

We explained earlier that it was not possible in Phase 3 to visit the teachers in their own countries and to observe how they were using their new coursebooks. We can therefore only report what they wrote to us in response to questions we sent them about their teaching and in the descriptions they wrote of classes they considered to be representative of their way of teaching TOEFL. We present below what we learned about the amount of attention the teachers devoted the four language skills, grammar, and vocabulary. We then summarize what they reported about how they handled different skills in their TOEFL iBT classrooms and what their reports indicated about their use of their preparation coursebooks.

Proportion of time devoted to skills. Table 9 indicates the percentage of class time the teachers claimed they devoted to each of the four skills, grammar, and vocabulary, both in their

CBT courses (Phases 1 and 2 of this study) and in the early stages of their TOEFL iBT teaching (Phase 3).

The percentage of class time the teachers claimed to be spending on reading, listening, and writing did not seem to have changed much from when they were doing CBT teaching. What this table does not show, however, is how this time was divided between independent and integrated skills. We also do not know how much time the teachers recommended their students spend on writing homework. We saw in Phase 1 that teachers rarely included writing practice (as opposed to explanations about writing) in their lesson plans, preferring instead for the students do their writing tasks at home and hand them in for marking in the next lesson.

The most dramatic changes were in the areas of speaking and grammar. T2 was the only teacher who had included any speaking in her CBT courses, and her figure rose from 5% to 20% when she began teaching for the TOEFL iBT. The other teachers began to pay attention to speaking for the first time when they began TOEFL iBT teaching, and their figures ranged from 10% to 30% of their TOEFL iBT class time. The figures for grammar dropped markedly, to 0% in two cases. The biggest fall was in T4's classes, where grammar had occupied 55% of his CBT time but now occupied a mere 2% of his TOEFL iBT time. (Note: These figures are estimates given by the teachers in response to specific tasks we sent them. The teachers may have given slightly different information in different phases and tasks, but we feel that the figures given here represent general trends during the time we were collecting our data.)

Table 9

Percentage of Class Time Spent on Skills, Grammar, and Vocabulary

Section	T1		T2		T3		T4	
	TOEFL CBT	TOEFL iBT	TOEFL CBT	TOEFL iBT	TOEFL CBT	TOEFL iBT	TOEFL CBT	TOEFL iBT
Reading	25	30	20	20	25	30	20	20
Listening	20	15	20	20	10	20	15	20
Writing	15	25	20	20	20	20	10	18
Speaking	0	30	5	20	0	10	0	20
Grammar	20	0	20	0	30	5	55	2
Vocab	5	0	10	10	10	5	0	20
Other	15	0	5	10	5	10	0	0

Note. CBT = computer-based test. T1, T2, T3, and T4 = Teacher 1, Teacher 2, Teacher 3, and Teacher 4.

Reading and listening. The teachers' descriptions of their classes in Tasks 1 and 2 suggested that there were no great changes in how they taught reading and listening for CBT and TOEFL iBT classes. A typical pattern, explained by T2, was for the teacher to base all her teaching on the coursebook. She would explain how the skills were tested, showing how a particular question type worked (such as inserting text into a passage) or demonstrating a particular subskill (such as scanning). She would then ask the students to work on exercises practicing this feature, lead the group as they checked their answers in plenary, and then help the group with any vocabulary that had caused them problems or that she felt they should focus on as being useful for TOEFL (24:27, 40:24). The TOEFL iBT changes in reading and listening, which involved longer passages in both cases, were barely mentioned by the teachers. They might have overlooked these features or thought them less worthy of comment compared to the bigger changes elsewhere in the test.

T4, however, described a departure from this approach. He described a session in which his students did a prelistening activity, listened to a passage twice (doing different tasks each time), and then listened a third time with an *academic focus*. What T4 meant by this was that they discussed problems they had with the language or looked at unfamiliar vocabulary (48:08). T4 was using the iBT4 book during this session. He stated that there was nothing like iBT3 and iBT5 available during CBT times (50:198), implying that his new approach would not have been easy to implement in his earlier preparation classes. T4 described academic listening not as just asking and answering TOEFL-type questions, as most coursebooks seemed to imply, but rather as getting students to think about the passages they had heard and do something with the information (e.g., comparing and contrasting; 50:169). He also used the iBT3 and iBT4 feature called Culture Notes to help his students to understand aspects of academic life specific to the North American context (49:21). T4's reading lessons followed a similar pattern: prereading vocabulary work, reading a passage at least two times, checking responses, and further vocabulary work, as set out in iBT3 and iBT4 (48:08).

Writing. The course descriptions the teachers sent us suggested that there had been some changes in their teaching of the productive skills since the introduction of the TOEFL iBT. The changes were not evident in all aspects of writing, however. The way T2 dealt with independent writing in her TOEFL iBT classes did not differ greatly from her treatment of writing in her CBT classes. In both cases she explained the main point being targeted, the students did a task

practicing this point, she checked the students' writing in her own time, and then gave the students feedback in the next lesson (24:98). In fact, T2 later used CBT preparation material when teaching the TOEFL iBT independent task, as she felt the tasks were so similar (22:144). T1 worked in a similar way when teaching independent writing. When it came to integrated writing, though, she introduced the notions of paraphrasing and summarizing, skills she had identified as new in TOEFL iBT and that she had made sure to look for when she was selecting her TOEFL iBT coursebook (14:203).

Speaking. It could be seen in Table 9 that all four teachers spent more time on speaking in their TOEFL iBT classes than they had done in the CBT classes, possibly because of their own worries concerning this new skill and because their students were not used to taking tests in speaking. T1 covered some speaking work in every lesson, while she dealt with each of the other skills in every second lesson (10:50). Her method for helping students to practice was to get them to perform one of the TOEFL iBT speaking tasks in the coursebook in front of their peers and then listen to the peers' feedback (11:30, 11:99) and her own (12:94). Although some of T1's understanding of the speaking requirements came from her participation in Phase 2 of this study, much of it came from her TOEFL iBT coursebooks. She particularly valued the marked samples of speaking performances that the coursebooks offered (11:138).

T4 used writing task prompts for both independent and integrated speaking practice and found that this practice worked well for his students (48:196). He used the model set out in his coursebooks (iBT3 and iBT4) for dealing with integrated speaking tasks, which included practice in note-taking (46:107). T2, as already mentioned, had asked a colleague to design some software to simulate a test situation in which students could read, listen, and then record their own voices (22:154, 22:89). All of these attempts to develop the students' speaking abilities represented important changes in TOEFL preparation practice.

Grammar. Grammar teaching was also referred to by the teachers as *structure*, since this was the term used for the relevant section of the CBT. Grammar teaching had taken place in all CBT classes, but it was often on a revision basis (see Wall & Horák, 2006, for further details). T1 stated that there was not enough time for students to study grammar on an TOEFL iBT course (21:327, 21: 342) since it was more important to cover the new components instead. Teachers could help the students to familiarize themselves with TOEFL and could give them tips about test-taking, but "if you don't understand English to a certain level nothing will help you" (T1,

8:106). T4 also felt that there was too little time to cover grammar (44:91). This view represented a dramatic change in his teaching as he had devoted over half his class time to grammar in his CBT courses. These teachers now dealt with grammar on a "need-to-know" basis only, addressing problems arising during the practice of other skills, if time allowed.

T3 was the only teacher who felt that she should be dealing with grammar on a principled basis. She was not sure how to do this, however, as grammar was not included in the coursebooks she used for the TOEFL iBT (35:123).

Note-taking. One of the features we expected to see in the TOEFL iBT classes was note-taking, as this was now allowed throughout the test. We had asked in Phase 2 about the teachers' plans for note-taking since it was not clear whether this was a skill students had already mastered. Some teachers reported plans to teach note-taking and had found materials in anticipation (Wall & Horák, 2008). T1 said that the fact her coursebook covered this skill was one of the reasons she had chosen it (14:225). T4 also taught note-taking since it was part of the approach in the coursebooks he was using (46:107, 46:118).

T3, however, who had reported plans to teach note-taking in Phase 2, reported that she was not actually doing so in Phase 3 as the students found it distracting. She left it up to them to use whatever note-taking skills they already had if they wished to (T2, 27:337).

Conclusion

What conclusions can be drawn from the evidence our data has provided? The main point is that there did indeed seem to be a strong influence from the TOEFL coursebooks on all the teachers—even on T4, who claimed that there was not. The coursebooks played an important role in course design and they were at the heart of each teacher's lesson plans, providing the content material, which reflected the new test, in all cases, and influencing the choice of methods in most cases. While the content the teachers covered differed from the content offered in Phase 1 (e.g., longer reading and listening passages and different question types), the activities they engaged in (teacher explaining new concepts, students doing exercises and checking their answers) seemed mostly similar to what we had observed at the start of our study. T4's approach stood out as different, coursebook-led but providing some opportunities for students to interact with each other, apparently because the coursebook he used contained such features. T1 also showed some innovative touches though, going as far as getting her students to speak in front of the group and assess each other's performances.

The difficulty we had in Phase 3, however, was that we only had the teachers' descriptions of their teaching as data, not our own observations. While the teachers were willing to help us, the conditions of the agreement we had with them meant that each teacher would only provide us with descriptions of two of their TOEFL iBT lessons. The descriptions they provided were not very detailed. We could see that they designed their lessons around their coursebooks, but we could not see whether they were interpreting the messages of the coursebooks (and therefore, presumably, the test) correctly and whether their students were responding appropriately. We were also aware that what the teachers were describing was their practice very soon after the introduction of the new test in their countries (the test was introduced in mid-May 2006 and the teachers sent us their descriptions in October of that year) and that the reliance they were showing on their coursebooks might be what Spratt (2005) termed "a fruit of uncertainty" (p. 12) rather than a long-term trend. Would the teachers be less dependent on commercial materials once they developed their understanding of the requirements of the new test and had more time or more confidence to try alternative materials and methods?

We hoped that by being able to interview the teachers face-to-face and to observe them in Phase 4 we would be able to probe more deeply into their ideas concerning the test, their coursebooks, and their teaching, and thereby gain fuller insights into their actual classroom practice. We also hoped to see whether their dependence on their coursebooks continued a year from the introduction of the test in their countries, or whether they would develop in time their own materials and introduce more innovation into their teaching methods.

The Phase 4 Study

Aims of the Study

The main aims of the fourth and final phase of the TOEFL Impact Study were to investigate whether the approach to TOEFL teaching had changed substantially between 2003 (before the introduction of the new TOEFL iBT test) and 2007 (when the last data were collected) and to determine whether any differences that might exist could be traced back to changes in the test itself.

Before presenting the study, however, it is useful to review some of the key ideas concerning test impact and washback, both in the literature of general education and language education.[2]

Test Impact and Washback

It is now accepted that developers of high-stakes tests should consider the consequences that their tests may have on the educational context and on wider society. Messick (1989) emphasized this need when he included the *consequential aspect* in his expanded view of construct validity. Discussion of test consequences or impact has been taking place in the field of general education for some time (Madaus, 1988; Popham, 1987; Vernon, 1956), but it is only since the 1990s that serious studies have appeared in the literature of language testing. Various articles had been written earlier about how tests could affect teaching either positively (e.g., Pearson's [1988] image of the high-stakes test being levers for change (p. 98) Swain's [1985] notion of "working for washback" (p. 36)) or negatively (e.g., Madsen's [1976] description of how the introduction of a new examination led to "selling English short" [p. 135]), but few publications offered more than expressions of faith or assertions that changes in tests had caused changes in the classroom. Empirical evidence was thin on the ground.

Alderson and Wall (1993) set the agenda for research in this area, problematizing the notion of *washback* (the influence of high-stakes tests on classroom practice) and stressing the need for test developers and researchers to be more specific when setting out to promote or detect test impact in educational settings. They proposed a number of *washback hypotheses*, which made specific some of the types of influence an important test might have: for example, it might influence what teachers teach (the content of the class) or how teachers teach (teaching methods). The hypotheses also illustrated possible focuses for research into the existence of washback in particular settings. Alderson and Wall also argued for a rigorous approach to data collection, advocating the use of classroom observation to complement the use of self-report techniques such as questionnaires and interviews. They encouraged other researchers to read outside the field of language testing for ideas that could aid in the understanding of how tests influenced teaching, indicating that the fields of motivation and innovation in education were particularly fruitful areas to explore.

Two further theoretical discussions of the notion of washback appeared in the early1990s, both commissioned by ETS as part of their validation of what was then known as TOEFL 2000 and was later to become the TOEFL iBT test. Hughes (1993) proposed that there were three main types of washback: washback on participants (anyone "whose perceptions and attitudes towards their work may be affected by a test" [p. 2]), processes ("any actions taken by the

participants which may contribute to the process of learning" [p. 2]), and products ("what is learned…and the quality of the learning [p. 2]). Bailey (1996) expanded this view, specifying four major groups of participants (students, teachers, materials writers and curriculum designers, and researchers) and four types of products (learning, teaching, new materials and curricula, and research results). She attempted to illustrate the relationship between the participants and the products and signaled the potential for feeding the results of these interactions back into test design. We paid special attention to the notion of "processes" during the transition phases (Phases 2 and 3) of the TOEFL Impact Study, investigating the processes one key group of participants—teachers of TOEFL preparation courses—went through as they learned about the nature of the new test, considered which elements should go into the design of new test preparation courses, and decided which teaching methods to use to develop their students' abilities to cope with the new test's demands.

Wall (1996, 2000) made a further contribution to washback research by questioning whether washback could be predicted or controlled. She introduced concepts from the field of innovation in education (e.g., Fullan, 1991; Henrichsen, 1989), to explain how factors other than test design could facilitate or hide the impact that important tests had on teaching. Henrichsen's (1989) hybrid model of the diffusion/innovation process was particularly useful to show the influence of factors in the educational environment before an innovation is introduced (in the TOEFL Impact Study the innovation is the TOEFL iBT test) and how these factors combine with factors in the innovation itself, characteristics of the teachers and learners, and other factors such as the quality of communication concerning the innovation to produce outcomes such as changes in teaching and learning. This model has heavily influenced the TOEFL Impact Study, providing the core of our frameworks for gathering and analyzing data.

A number of studies have been undertaken since the mid-1990s. These studies fall into two main categories:

1. Those that look at the impact of international tests such as TOEFL (Alderson & Hamp-Lyons, 1996; Johnson, Jordan, & Poehner, 2005), the First Certificate in English (FCE; Tsagari, 2006), and IELTS (Green, 2003; Hawkey, 2006; Hayes & Read, 2004)
2. Those that look at tests and other forms of assessment at national level (Andrews et al., 2002; Burrows, 2004; Cheng, 1997, 1998, 2004; Ferman, 2004; Qi, 2004; Shohamy, Donitsa-Schmidt, & Ferman, 1996; Wall & Alderson, 1993; Watanabe, 1996, 2004).

These studies explore various aspects of test impact, including intended and unintended consequences, and consequences that take place before the test is introduced (in test-preparation classes) as well as afterward.

Spratt (2005) surveyed many of the studies that had been produced up to 2003 and identified five areas that were "susceptible to washback" (p. 26). These areas were curriculum materials, teaching methods, feelings and attitudes, and learning. In each category she listed the issues that had been investigated in the work she reviewed:

- Curriculum
    - How much to focus on the exam's content domain as opposed to exam techniques and test wiseness
    - When to teach particular areas of the curriculum
    - How much time to devote to teaching particular areas
- Materials
    - What textbooks to use
    - How much use to make of selected textbooks
    - How much and how to use exam or parallel exam materials
    - How much to use other materials including one's own and the students'
- Teaching methods
    - How much drilling to employ
    - When to employ such methods
    - How much to employ other methods more focused on language development and creativity
    - What kind of exam preparation to employ
    - How much planning time to devote to exam classes
    - What kind of atmosphere to promote in exam classrooms
    - What kind of interaction patterns to encourage in exam classrooms

- Feelings and attitudes
    - What kinds of feelings and attitudes toward the exam to attempt to maintain and promote in students
- Learning
    - The appropriateness of the learning outcomes demonstrated by students(Spratt, 2005, p. 26)

We have addressed many of these issues in the Impact Study, particularly issues having to do with curriculum, materials, and teaching methods. Phase 3 was devoted to a discussion of the materials teachers were using during the transition period between the old and new versions of the TOEFL, and issues related to curriculum and teaching methods have been addressed throughout. We were more concerned with the feelings and attitudes of teachers than of students, however, as practical considerations prevented us from investigating students after Phase 1. Nor have we been able to address the appropriateness of learning outcomes, if this phrase refers to the students' abilities at the end of their preparation courses or to the results when they took the TOEFL.

Cheng and Watanabe produced a collection of studies on washback in 2004. This volume included a review of the notion of washback (Cheng & Curtis, 2004), a survey of methods that have been used in washback research (Watanabe, 2004), and a review of research related to washback and the curriculum (Andrews, 2004). It also included eight case studies about washback in different education contexts in different parts of the world. Cheng's (2004) case study built on earlier work she had carried out (1997, 1998) on the effects of a new examination in Hong Kong on teachers' classroom practice. While the teachers' perceptions of the new exam were accurate and positive, and while they indicated a willingness to change their practices to correspond to what they felt was important in the exam (e.g., more oral and listening tasks, and more real-life tasks), observation late in the study showed that they had not changed many aspects of their teaching, such as teacher talk and delivery modes (Cheng, 2004, p. 162). These results matched a trend seen in much of the research to date, namely, that it is more common to find test washback on the content of teaching (in Alderson and Wall's [1993] terms, *what* the teachers teach) than in teaching methods (*how* they teach). Watanabe's (2004) study furthered his earlier work (1996) on the influence of university entrance examinations on teaching in

secondary level education in Japan and confirms his earlier finding that the examinations do not affect all teachers in the same ways. Watanabe concluded that amongst the factors mediating washback were the teachers' personal beliefs about proficiency, their sometimes mistaken perceptions of what the examinations required, and their own teaching competence. Wantanabe felt that it would not be possible to achieve the washback intended by the examination designers without retraining teachers, including both familiarization with new teaching methods and, importantly, help in changing their perceptions (2004, pp. 139–142).

We have been influenced by many of the studies listed above, but the work that we see as most relevant is the Alderson and Hamp-Lyons (1996) study into TOEFL test preparation classes in the United States. These researchers used teacher and student interviews and classroom observations to try to discover whether differences existed between the way teachers taught when they were conducting ordinary (non-test-preparation) classes and when they were preparing students for the TOEFL. They determined that the test influenced both what and how the teachers delivered their classes, but it "does not explain why they teach the way they do" (p. 295). One of the aims of the Impact Study as a whole has been to seek explanations about why teachers react to the TOEFL in the ways they do.

The main purpose of Phase 4 was to investigate whether any changes appeared in the way our participants conducted their classes between the time of the Phase 1 baseline study (2003–2004) and roughly 1 year after the launch of the new TOEFL test in the teachers' countries (2007). This study responded to McNamara's (1996) claim that "high priority needs to be given to the collection of evidence about the intended and unintended effects of assessments on the ways teacher and students spend their time and think about the goals of education" (p. 22). It also anticipated Cheng's (2008) belief that future washback and impact studies should be "multiphase, multimethod and longitudinal in nature" (p. 359).

The intended effects of the new TOEFL were presented in the Phase 1 baseline study (Wall & Horák, 2006), as were the "antecedent" conditions (Henrichsen, 1989) that existed before the teachers became aware of the changes that were about to occur in the TOEFL. The teachers' reactions to the news they were receiving about the test and their early thoughts about how they would revise the courses they were teaching are documented in Wall and Horák (2008). The teachers' choice of textbooks and how they were beginning to use them have been documented in the description of the methodology in Phase 3. We now report on Phase 4 of the

Impact Study, which gave us the opportunity to see how teachers were teaching after the new test had settled into their contexts and to investigate whether there were any "evidential links" (Messick, 1996) between the new test and the way that teachers were now teaching.

Research Questions

Phase 4 was meant to document the types of teaching taking place a year after the introduction of the new TOEFL in the countries represented in our sample, and to draw on the findings of earlier phases to explain how the teachers' understanding of the new test, the materials they had selected to use with their students, and factors in their own teaching contexts and other factors (such as the quality of communication between ETS and the teaching community) might have influenced their approach to preparing students for the new TOEFL. Phases 1 to 3 had set the scene and introduced the key characters, but the Impact Study would not be complete without a return visit to some of the original teaching sites and interviews with the teachers we had been tracking for nearly 4 years. This visit would allow us to add our own view as independent researchers to the self-report accounts provided by the teachers in Phases 2 and 3. It was hoped that the integration of insider and outsider perspectives would provide a firm platform for any arguments we might make in the end regarding the nature of TOEFL impact.

This phase included five research questions:

1. What did classroom practice look like 1 year after the introduction of the new TOEFL in the countries in our sample?
2. Was the approach to teaching similar or different from the approach that was observed in Phase 1 (2003)?
3. If there were differences in the teaching, could these be linked to changes in the TOEFL test?
4. If there were differences in the teaching, were they in the desired direction?
5. What factors apart from changes in test design might have affected the approach to teaching?

Phase 4 began in April 2007 and data were collected between May and October of the same year.

Methodology

Sample of participants. Phase 4 focused on three of the four teachers we interviewed and observed in Phase 3. The fourth teacher was not able to continue into Phase 4, due to heavy work commitments at her school. The three teachers who stayed on were referred to as T1, T2, and T4 in the Phase 3 study, and we use the same code numbers in Phase 4. Their details can be found in the description of the sample for Phase 3. See Table 2.

Sample of institutions. All three institutions were operating in the private sector. They offered a range of language courses, both for general language development and test preparation. The test preparation courses were aimed mainly at adult learners wishing to gain a particular TOEFL score, usually, but not exclusively, for studying abroad. T1's institution was also an education information center for students aiming to study in the United States, and it had recently become a TOEFL testing center. T4's institution had also recently become a TOEFL testing center. This site was the largest of the three and was part of a larger educational institution (a private university) offering courses in a wide variety of subjects.

We had been asked at the start of the Impact Study to concentrate on countries in Central and Eastern Europe. We defined this region as countries that had been members of the former Soviet bloc or that had opened up since the fall of the Berlin Wall. Our original sample of 10 teachers came from six countries in this region. We were later asked by ETS to add a country in Western Europe to the sample, so we enlisted two more teachers in a seventh country. Of the three teachers and institutions remaining in Phase 4, two were from the original region and one was from the country added later. (We hesitate to name the countries because with the reduction in sample size it is more likely that the institutions and the teachers, to whom we promised anonymity, can be identified.) All three institutions are located in either the capital city of their country or a major regional town.

Data collection. The plan for Phase 4 was to mirror the Phase 1 activities as far as possible. We would travel to the three institutions, interview the teachers about their views of the new TOEFL and their approach to preparing students, and observe their actual teaching. We would also interview the director of studies in each of the institutions to gain a managerial perspective on the introduction of the new test and the type of impact it might be having on the institution as a whole. In none of the institutions was the director of studies the same as in Phase 1. Their responses proved useful, nevertheless, providing us with new information about

the local context. We would not be able to interview students, having had to drop the idea of investigating their views of the test and its effects on them at the beginning of Phase 2, due to practical constraints. The Phase 4 data collection activities are presented in Table 10.

Tracking questions. We e-mailed the teachers a set of tracking questions in the 2nd month of the study. We had not gathered data from them in the previous six months and we needed to establish what the situation was now like regarding their new TOEFL preparation courses. We also asked them a number of questions that had originally been part of the Phase 1 teacher interviews. Some of these questions required only a factual response and we wanted to deal with them on paper so we would not have to spend time on them in the Phase 4 interviews unless the paper responses proved worthy of further probing. Tracking Question 1 (*Have you done any work with any exam board or exam bodies since our first contact with you at the beginning of the project?*) is an example of this sort of question. If the teacher's response was in the negative, there would be no need to ask for more details. Other questions required some thinking time, not because they were difficult questions, but because they might involve the teachers in some calculations. Tracking Question 14 (*What percentage of time do you spend over the length of your course on listening, reading, writing, etc?*) is an example of this type of question. It was useful to send such questions in advance, to avoid wasting time in the face-to-face interview while the teachers worked out their answers. Again, we could ask more about their responses in the interview if we needed to do so.

Table 10

Phase 4—Data Collection Activities

Month	Data collection activity
May 2007	Tracking questions via e-mail
June/July 2007	Face-to-face interviews with teachers
	Face-to-face interviews with directors of studies
	Classroom observations
August 2007, and ad hoc contact up to early 2008	Follow up questions via e-mail

Interviews. The teacher interview was based in part on the responses the teachers supplied to the tracking questions, but it then went on to cover the remaining questions in the Phase 1 interview schedule. The interview was to be conducted after the researcher had observed at least one of the teacher's classes. This arrangement would allow us to ask further questions about what we had seen in the observation. The interview schedule was semistructured, with questions grouped into sets of themes we wished to cover. (See Appendix C for the teacher interview schedule.)

We also made some changes to the interview schedule for the directors of studies, inviting them to comment on whether and how the switchover from CBT to TOEFL iBT had affected their institution as a whole, their staffing, teacher training, resourcing, class sizes, and the content and methodology of classes. (The specific questions we asked can be seen in Section 2 of the teacher interview schedule in Appendix C, as we wanted to gather responses from both parties on the same themes.)

We interviewed the teachers and the directors of studies separately. We interviewed two directors of studies at T4's institution, each responsible for a different aspect of teaching.

Classroom observations. We believed that in Phase 4 it was important to observe the TOEFL classrooms with our own eyes. In Phases 2 and 3 we had attempted to get information about the nature of the TOEFL classrooms from the teachers themselves, using computer-mediated communication. We were able to collect some rich data in both phases, but at times the teachers' responses lacked the depth we needed. This was especially true in Phase 3, when we asked the teachers to describe lessons they were planning and had given to their students. We were well aware that Rogers' (1983, as cited in Markee, 1997) warned about the lack of methodological rigor that can occur "when researchers rely on the subjective recollections of informants instead of objective observational procedures to describe adoption behaviours" (p. 6). It was only by carrying out our own observations that we were would understand, for example, the major role computers were playing in one of the teachers' classrooms, something which was so normal to the teacher herself that she did not dwell on it in her descriptions in Phase 3.

In Phase 1 we had aimed to observe two sorts of classes taught by the same teacher: a TOEFL preparation class and a non-test-oriented class at a similar level. This arrangement would have allowed us to identify each teacher's personal teaching style and to disentangle this style from other factors that might be affecting their classroom practice. This was not possible with all the Phase 1 teachers though, as it was difficult to find institutions where the same teacher taught

both sorts of courses, and it was not possible with any of the Phase 4 teachers. Green (2006, p. 339) noted that few washback studies have managed both types of observations, with Brown (1998), Alderson and Hamp-Lyons (1996), and Watanabe (1996) being notable exceptions.

Transcription, coding, and analysis. All the interview data were fully transcribed and notes from the observations were summarized and typed up into the same format. The responses to the tracking questions and the interview transcripts were analyzed with the help of Atlas-ti, the qualitative data package we had used in all the earlier phases of the study.

The code list included all the codes used in Phases 1 to 3 and a dozen new codes that clarified existing concepts rather than describing new phenomena. The full set of codes can be found in Appendix B. As in all previous phases of our research, the data were coded independently by both members of the research team, and any differences in coding were resolved through discussion. The discussion often resulted in a decision to use both codes, since the codes were not mutually exclusive. We then sent all three teachers a summarized version of the interviews to make sure we had their agreement about the main points arising.

We present our findings in the next seven sections. We first present what we learned about the teaching of each of the skills tested on the TOEFL iBT. We then present what we learned about the treatment of grammar and vocabulary, both of which figured prominently in earlier versions of TOEFL but neither of which was tested separately on the TOEFL iBT.

Each section begins with a reminder of the changes that were introduced in the TOEFL iBT. This is followed by a discussion of the type of impact (if any) the experts behind the design of the TOEFL iBT hoped to see in future language teaching. We summarize the type of teaching that was taking place during Phase 1 and any issues that arose in Phase 2, and then discuss what teaching looked like during Phase 4, the final phase of the study. We conclude each section with our view of whether there were any changes in the teaching and whether these could be traced back to the changes in the TOEFL.

The Teaching of Reading

The changes that occurred in the testing of reading were as follows:

- Length of texts—The TOEFL iBT passages are twice as long as the CBT passages (TOEFL iBT. 700 words; CBT, 250–350 words).

- Number of texts—There are fewer texts in the TOEFL iBT than in the CBT (TOEFL iBT, 3 to 5; CBT, 4 to 5).
- Items per text—There are more questions per passage in the TOEFL iBT than in the CBT (TOEFL iBT, 12–14 items; CBT, 11 items).
- Text types—The TOEFL iBT includes "a broader selection of academic text types, classified by author purpose: (a) exposition, (b) argumentation, and (c) historical biographical/ autobiographical narrative" (Cohen & Upton, 2007, p. 213). The change in text types and the additional length mean that it is possible to offer more complex texts as well, for example, by the inclusion of "multiple-focus passages (compare/ contrast, cause/effect)" (ETS, 2005a, p. 2).
- New question types—The reading to learn questions aim to test the candidate's ability to recognize text organization, distinguish between main ideas and detail, and understand rhetorical function.
- Glossary—Candidates can click on certain "special purpose words" (ETS, 2005b, p. 8) to access a definition or explanation. This facility is supposed to increase the authenticity of the reading experience in that readers would normally have access to a dictionary for technical or unfamiliar terms in the target use situation. The help is only available for some of the words in the text, however, not necessarily all those the candidates might find difficult.

Rationale for change. No major concerns were expressed in the framework for reading (Enright et al., 2000) about the negative effects of the CBT on the teaching of reading, but the team charged with redesigning this section of the TOEFL decided to make the texts and tasks more authentic: "Longer texts better represent the 'academic experiences of students' and. . . they better facilitate the development of Reading to Learn purposes in the test design" (Mary Schedl, personal communication, cited in Cohen & Upton, 2007, p. 213).

In addition, "these new formats were expected to elicit somewhat different 'academic-like approaches' to reading than those elicited by the more traditional formats" (p. 214).

Intended impact. The framework for reading did not mention any specific impact on the teaching of reading, only that it should be more communicative in the future (Enright et al.,

2000, p. 49). The authors recommended that research should be designed to "investigate washback effects on what examinees study and to determine whether the emphases on communicative learning increases once the new test is operational" (p. 49).

The experts we consulted in Phase 1 mentioned three ways they imagined TOEFL iBT would have an impact on the teaching of reading. They predicted that longer texts would be used in class, that these texts would display more complex rhetorical structures, and that teaching would focus on making connections between the parts (Wall & Horák, 2006, p. 15). (Note: The reader is reminded that the search for statements about impact in the framework documents and the survey of experts were two separate operations, run independently of each other. The survey of experts was not meant to result in a working definition of the term *communicative*.)

Findings from Phase 1. The teaching of reading in Phase 1 was coursebook-bound in terms of content and methods. All but one of the 12 teachers followed their coursebooks closely, which meant that the students experienced a great deal of testlike practice (Wall & Horak, 2006, p. 47). The students usually worked individually to respond to exercises in the coursebook. They then checked their answers in plenary and discussed any difficult questions (p Wall & Horak, 2006. 50). The exercises generally mimicked the format of TOEFL. To supplement this routine, all the teachers recommended that their students should read as much as possible outside class time, with the main aim of increasing their vocabulary (Wall & Horak, 2006, p. 51). The teachers claimed that a lack of sufficient vocabulary was the biggest factor stopping students from doing well on the CBT (Wall & Horak, 2006, p. 52).

No teacher reported asking his or her students to do reading as input to any other skills practice, apart from one who sometimes asked her students to read the tape scripts at the back of their coursebook to aid their comprehension of listening passages. This was not integration of skills in the TOEFL iBT sense, however, where the students are expected to process and comment on information from different written or spoken sources.

Findings from Phase 2. Reading was the section of the TOEFL iBT that the Phase 2 teachers commented least on. Two teachers mentioned the fact that reading texts would be longer and two that there would be fewer texts, but they generally seemed to perceive that the new test would be similar to the CBT (Wall & Horák, 2008, p. 41). They felt that the item types were mostly the same, although some mentioned items requiring summarizing, paraphrasing, table completion, and inserting text (three teachers). One teacher was pleased by the idea that the

reading section would no longer rely solely on traditional multiple-choice items. She felt that new item types would offer a more authentic reflection of the situation students would face in an academic institution in North America.

There was a range of opinions concerning which subskills would be tested by the TOEFL iBT reading section, with some teachers feeling that these would remain the same and others believing that the students would have to think harder. One teacher, for example, believed that TOEFL iBT would require "synthesizing, comparison, selection—higher-order skills!" (Wall & Horak, 2008, p. 42). This view seemed in line with ETS intentions, but few of the other teachers took it on board at that time. Another teacher believed that study skills such as using a dictionary would be tested, perhaps because of the inclusion of the glossary. The idea of a glossary was generally welcomed, but one teacher assumed that this meant the reading passages would be more difficult than in CBT.

All of the comments made by the teachers related to the reading section itself, rather than to the reading that would serve as input to the integrated tests of writing and speaking. We do not know why this should have been since the teachers did comment on how listening contributed to integrated writing and speaking.

Findings from Phase 4. The main findings from Phase 4 are presented in Table 11. The first aspect to consider is whether the teachers were aware of how the TOEFL iBT differed from the CBT. If they were not aware of these differences then it would be difficult to attribute any changes in their teaching to the changes in the TOEFL. It is also important to assess the teachers' attitudes toward the new test, since we know from innovation theory that potential users of an innovation (the innovation here being the new test) are unlikely to implement the innovation if they do not see it as an improvement over existing practice (Henrichsen 1989, p. 84). The categories percentage of class time dedicated to skill, materials used, and methods used are criterial features in many washback studies, and advice to students is a concrete manifestation of what teachers believe is salient in the tests they are preparing their students for.

Table 11

Phase 4—The Teaching of Reading

Characteristics of teacher and teaching	T1	T2	T4
Teacher's awareness of new test			
Types of texts	Longer texts than in CBT (7.65)	Longer texts than in CBT (4.63) Teacher didn't feel type of texts had changed (5.300)	Teacher didn't feel texts were more complex (2.1551)
Topics	Academic topics (7.68)	Academic topics—biology, astronomy, history (4.65)	Not as many business-related topics (2.1551)
Subskills			Recognized similarities between CBT and TOEFL iBT—vocabulary items, inference, reference (2:1427) New skills—e.g., summarizing (2:1410)
Other features			Glossary (2:1600)
Teacher's attitude toward new test	Positive	Positive	Positive
Percentage of class time dedicated to skill	35% (7:51, 8:1253) Former students reported that reading section was difficult (8:1267)	20% (5:119) Hardest section to teach, because of different logic (4:339) Homework: Often assigned reading homework (5:84)	20% (1:58) Reading less problematic than other skills Homework: Often assigned reading homework (2:61)
Materials used	Used iBT2 to introduce tasks and to practice them (7:61, 7:124)	Used iBT5 to introduce tasks Used iBT2 and other books for computer practice (4:169)	Used iBT5 to introduce tasks Used iBT4 to develop skills Used iBT0 for test practice Used iBT1 and iBT2 for computer practice (1:244 - 252; 2:2131-2145)

Characteristics of teacher and teaching	T1	T2	T4
			Felt some material in iBT4 not representative of TOEFL iBT. (2:1388) Omitted some exercises from iBT4 because too "technical" 2:1388)
Methods used	Named materials first when asked about methods (7:124) Introduced test section Students did practice exercises Teacher and students checked answers together, discussing some answers when there was a disagreement (8:137) Taught 'principles' underlying questions (8:1273)	Introduced test section Students did practice exercises on computer Students checked own answers on computer (observed) Discouraged looking up meaning of words while taking tests, though checked vocabulary afterward (5:480) Uses prereading exercises in non-TOEFL classes. (5:559)	Named materials first when asked about methods (1:138-145) Did prereading exercises and discussion of topics (observed)
Advice to students	Read outside class, especially academic texts to get used to style (8:1321)	Read outside class to build up vocabulary (5:476) Practice reading on Internet (5:510) Identifying main topic of readings (5:498)	Read outside class to build up vocabulary (2:1543) Read *Popular Science, TIME, Newsweek, The Economist*, newspapers (2:1529)
Has there been any change since Phase 1?	Change in content No great change in methods, though slightly more teacher-student interaction.	Change in content No change in methods	Change in content Change in methods. More communicative, influenced by choice of coursebook.

Note. The numbers in parentheses refer to the transcript and line where information can be found.

T1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4. Coursebooks are identified in Table 3.

As in Phase 2, the teachers did not have much to say about how the TOEFL iBT reading test differed from the CBT version. T1 and T2 mentioned that the texts were longer now, but neither T2 nor T4 felt there was much difference between the types of texts presented. T4 offered comments on the types of subskills that were tested, but what he focused on was the similarity between the tests rather than the differences.

T1 and T2 felt that the reading section was difficult for students, and T1 spent about a third of her teaching time on reading. She had spent less time on reading the first time she ran an TOEFL iBT preparation course, concentrating instead on the new skill of speaking, but the students who then took the TOEFL iBT told her that the reading section was very difficult and she decided to dedicate more time to this skill as a result. T2 felt that reading was the hardest skill to teach. She commented on the problems her students had following the logic of the academic texts they were reading and how she found it difficult to explain to them what they needed to do in order to follow it better. She spent only 20% of her class time on reading, but she often gave the students reading homework. T4 also assigned substantial reading homework. The homework would provide useful reading practice, of course, but the drawback would be that the teachers would not know whether the students paced themselves when reading—a skill that would be useful when taking the TOEFL iBT itself in the future.

In terms of materials, T1 depended on a single coursebook for explanations and practice material, while the other two teachers used a combination of coursebooks for different purposes. When asked about the methods they used for teaching, both T1 and T4 offered the names of their coursebooks. Further probing and some observation work revealed more details about their activities in the classroom. T1 and T2 basically followed a presentation and practice routine, explaining the requirements of the reading section and the types of questions it contained and then getting the students to do many practice exercises. T1's students worked from their coursebook, writing answers either in the book or on paper, but T2's students worked on the computer, as if they were in a real testing situation. T1 got her students to give their answers to the whole group and they discussed any problematic questions. T2's students checked their answers individually, using the checking facility built into the software. There was an interesting contradiction in T2's approach to reading. On the one hand she encouraged her students to ignore unknown words as they were reading (5:48), but on the other she asked for a translation of many of the new vocabulary items after the students finished their test practice. She did not make a

distinction between words that might be useful for other texts in the future and some words (e.g., the names of different sorts of fish) that were specific to a particular text and were unlikely to appear in any other.

While the difference between T1's and T2's teaching in Phases 1 and 4 was slight, considerable difference was noticed between how T4 was teaching then and now. His approach seemed to be influenced greatly by the coursebook he was using (iBT4), which included suggestions for prereading work, pair and group work tasks, integrated skills work, and discussions of the content being covered. The reading activity we observed hardly differed from the type of reading work we would see in any modern general English classroom. T4 felt that the TOEFL iBT had freed him to teach in the way he had been trained to. This observation was in contrast to T2, who felt that different approaches were needed in test preparation classes and other classes:

> In general, in the TOEFL classes you can't see a lot of methodology … because it's simply a course to prepare the students for the TOEFL and to improve their score and skills with whatever we can…. In general English classes, well, a lot of other techniques could be used. (T2, 5.581)

All three teachers advised their students to read widely outside the classroom. T1 wanted her students to get used to an academic style of reading and encouraged them to read academic texts in their first language as well as in English. Both T2 and T4 emphasized the need to build up vocabulary. T2 encouraged her students to read English on the Internet, saying they were used to such reading from their normal schooling. T4 encouraged his students to read popular science and news magazines and newspapers.

Summary. Clearly changes occurred in the content of all three teachers' classes. The changes seemed logical and predictable given that the teachers were all using new coursebooks that reflected the design of the TOEFL iBT. The experts consulted in Phase 1 had imagined washback in the form of longer texts, texts with more complex rhetorical structures, and teaching focusing on making connections between the parts. If the coursebooks were accurate in their representation of the types of texts used in the TOEFL iBT and if their exercises demanded attention to discourse features, then it could be said that the new test had had a positive influence on the content aspect of teaching.

No change was seen, however, in the methods that two of the teachers used to teach reading. The third teacher showed considerable change, however. Whereas his classes in Phase 1 had consisted mainly of input and test-like practice, he now offered a wider range of activities and more student-to-student communication. This change seemed to be a function of the particular coursebook he was using.

The Teaching of Listening

The changes that occurred in the testing of listening were as follows:

- Passage types—TOEFL iBT included only lectures and extended conversations, while CBT included mini-lectures, short conversations, and dialogues. All TOEFL iBT passages are academic or academic-related. Lectures may include some interaction between lecturer and students.
- Number of passages—There are 4 to 6 lectures and 2 to 3 conversations in TOEFL iBT, as opposed to 11 to 17 dialogues, 2 to 3 conversations, and 4 to 6 mini-lectures in CBT.
- Length of passages—The TOEFL iBT lectures are longer (TOEFL iBT 3 to 5 minutes, CBT 2 minutes), and the TOEFL iBT conversations are longer (TOEFL iBT 3 minutes, CBT 2 minutes).
- The language on the TOEFL iBT is modeled on the Spoken and Written Academic Language (SWAL) corpus (Biber et al., 2004).
- Accents—One lecture in each version of the TOEFL iBT is delivered in a British or Australian accent (not just North American accents).
- Replay questions—The relevant section of a passage is played again before a question is given.
- Note-taking is allowed throughout the TOEFL iBT listening section.

ETS reported that there were also new questions which aimed to "measure understanding of a speaker's attitude, degree of certainty, purpose, or motivation" (ETS, 2005a, p. 2). It is not yet clear, however, how this question type differs from the CBT listening question type that asked *What does the man/woman mean?*

Rationale for change. The major reason for changing the listening section was to make the listening passages and tasks more authentic. According to Enright (2004), "an important goal was to develop listening materials that reflected the types of spoken discourse that occur in academic settings" (p. 148).

Enright (2004) added in a footnote that it was more difficult to create this sort of authenticity in listening tests than in reading tests. With reading it was possible to use extracts from longer extant sources, but it was necessary to create listening materials from new.

It was decided to use the SWAL corpus to maximize authenticity:

> Prospective listening texts were analyzed using diagnostic tools developed by Biber. The texts were then compared with the corpus to determine how closely they corresponded to authentic corpus texts with respect to major characteristics and, in some cases, modified to increase their semblance to real-world academic discourse. (Enright, 2004, pp. 148–149)

This decision went some way to address earlier criticism that the listening passages were unrealistic and unnatural (see Buck, 2001, p. 223, for example).

Intended impact. As was the case with the reading framework, there was no mention in the listening framework of listening-specific impact. The main message about impact was general: that TOEFL preparation courses should come to resemble "communicatively-oriented academic English courses." The authors hoped for "an assessment that satisfies the demands of several constituencies without sacrificing construct representation" (Bejar et al., 2000, p. 36). There were also no comments about listening-specific impact from the experts we consulted as part of the Phase 1 baseline study (Wall & Horák, 2006, p.15).

Findings from Phase 1. The main finding emerging from Phase 1 was that there was a "paucity of techniques to actively improve listening skills" (Wall & Horák, 2006, p. 32). Classroom activity consisted mainly of students listening to a recording of a passage in their coursebook, answering questions about what they had heard, checking their answers in plenary, and listening to the teacher's explanation if they had not arrived at the correct answer. There were some minor variations on this theme—when, for example, the teacher let the students read the tape script to enhance their understanding or when the students were allowed to listen to a recording a second time (p. 36). A few teachers distributed word lists, mostly of synonyms or idioms (p. 35), but on the whole the teaching of listening was as book-bound as the teaching of reading. Only one of the 12 teachers approached listening in a non-test-focused manner, but his

educational philosophy and the characteristics of his group of students differed markedly from the rest of the group of teachers (p. 37).

Several of the teachers were not worried about the listening test, as their students found this the easiest skill to cope with. These students were exposed to English outside their classes through popular culture (Wall & Horak, 2006, p. 38), and the teachers' main advice to them was to try to increase this exposure even further. Not all of the teachers were in this position though, and some dedicated large amounts of class time to practice (p. 36). Overall there was "little consensus on how this skill was perceived or on how best to approach it" (p. 36).

Two points arose from the teachers' discussions of the test itself rather than of their teaching. The first had to do with the relevance of the passages and questions to their own students' target language use situation. The CBT settings were exclusively US-oriented and some of these students would be studying English in other countries, including countries in Europe. For such students, passages that required an understanding of American university culture were problematic. The second point had to do with the role that memory seemed to be playing in the successful answering of test items. Those who took the test could not see the questions in advance of listening, and they had to remember what they had heard until they saw the questions. This was especially difficult when they listened to lecture passages, something they would not be expected to do in real life without being allowed to take notes on what they were hearing.

Findings from Phase 2. The teachers did not seem to register many changes in the listening section, apart from the fact that note-taking would be allowed in the future (this was unanimously welcomed) and that there would be a reduction from three to two types of stimulus material (Wall & Horák, 2008, pp. 44-45). Some details were mentioned by a couple of teachers—for example, that the conversations in the TOEFL iBT might involve more than two people and that questions about the speaker's meaning and attitude would be included. The inclusion of non–North American accents did not register with most of the teachers, although one was worried by rumors her students had heard that they would have to listen to the English of non-natives. The teachers who participated in Phase 3 still seemed not to have picked up on the accent issue, perhaps because the coursebooks they were using did not mention it either.

When asked about how they would teach listening in the future, the teachers commented mainly on the teaching of note-taking. One teacher raised this issue more than the others, perhaps

due to her English for academic purposes (EAP) training and teaching experience. She had searched for suitable materials on the Internet and already had some ideas for teaching, using, for example, staged multiple listenings of passages. The other teachers were also thinking about how to tackle this skill, but had not yet reached any conclusions.

The teachers were aware that they needed to help their students with the listening for both the independent and integrated tasks. One mentioned that since the passages would be longer her students would need to build up their stamina. Another felt that her students would need help comparing listening and reading inputs, but she, like the other teachers, had not come up with any ideas on how to help them in concrete ways.

Findings from Phase 4. The findings from Phase 4 are presented in Table 12 and explained below. All three teachers seemed to be aware of the main changes in the listening test, though they might not be able to list them when asked about them directly. They seemed to be confident in their understanding of the nature of the listening passages when they commented on whether the passages in their coursebooks were similar to or different from those on the test.

The teachers differed in the amount of class time they chose to devote to listening. T1 devoted only 10% of her time to this skill, stating that her students generally had no problems with listening, especially since they could now take notes on what they were hearing. She found it difficult to help students who did have problems, however, as she had a limited stock of techniques for dealing with listening and was not sure they were effective: “Sometimes I’m really desperate. I don’t know. . . how to help them” (T1, 8:540).

T2 spent about 20% of her class time on listening. In contrast to T1, she thought that her students found listening quite difficult. Her main support for this view was the fact that when she offered time for free practice at the end of her classes, students often chose to do extra practice in listening (5:21). This was indeed the case when we observed her lessons, when nearly half the class opted for listening.

The teachers used the same coursebooks for listening as they did for reading, and they believed the materials represented the TOEFL test accurately. We were interested to see how confidently the teachers spoke about the features of the TOEFL iBT exam when in fact none of them had actually taken it. T4 was pleased with his coursebooks (iBT3 and iBT4), not because he thought the listening materials were similar to the TOEFL iBT exam but because he thought

Table 12

Phase 4—The Teaching of Listening

Characteristics of teacher and teaching	T1	T2	T4
Teacher's awareness of new test	Generally aware, though did not mention specific changes	Generally aware, though did not mention many specific changes	Generally aware, though difficult for him to list changes when first asked (2:1659)
Types of passages		Dialogues no longer used (5:1393)	Passages seemed more authentic (1:1715)
Topics			Topics similar to CBT
Subskills			
Other features		Note-taking included in test (4:330)	Note-taking included in test (2:1704)
Teacher's attitude toward new test	Generally positive, but felt students could still select right answer without understanding (8:1397)	Generally positive	Positive (2:1715)
Percentage of class time dedicated to skill	10% (7:51) Felt students did not have problems with this section, especially since they could take notes (8:1262 and 1394) Felt this was the hardest section to teach, because she did not have many techniques (7:335)	20% (4:50) About half the students chose to do more listening practice in their free time near the end of each class (5:21)	25% (1:57)
Materials used	Considered using iBT5 in early stages, but chose iBT2 in the end—similar to TOEFL passages (7:65)	Used iBT5 to introduce topics Used iBT2 for computer practice—similar to TOEFL passages, academic, roughly the same length (4:69)	Used iBT3 and iBT4—authentic passages, covering a wider range of genres than TOEFL, faster than TOEFL, harder than TOEFL (2:1634) Used iBT0 and iBT2 at later stage of course for practice tests (44:143; 50:185)

Characteristics of teacher and teaching	T1	T2	T4
Methods used	Gave name of coursebook when asked about methods (7:22) Gave up idea of teaching note-taking (8:1448)	Gave name of coursebook when asked about methods • Introduced section and explained question types (4:119) • Students did practice exercises on computer and check their answers individually (observed). Gave up idea of teaching note-taking	Gave name of coursebook when asked about methods (1:138) Used iBT3 and iBT4 activities, with prelistening, pair (47:22; 48:33, 76 and 102; 49:15; 48:102) Students discussed all options for listening questions (48:11) Replayed recording if there was disagreement about answers, and encouraged discussion (2:1242) Gave up idea of teaching note-taking
Advice to students			Maximize listening outside class (2:1669)
Has there been any change since Phase 1?	• Change in content. • No change in methods.	• Change in content. • No change in methods.	• Change in content. • Change in methods. More communicative, influenced by choice of coursebook

Note. The numbers in parentheses refer to the transcript and line where information can be found. T 1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4. Coursebooks are identified in Table 3.

they were different. He especially liked the pace of the coursebook passages: "So they speak faster, they speak at a native speaker's pace, and it's hard for the students to, let's say, get what they're saying. But this is good for their preparation because they practice at i+1 as you say. They like it" (T4, 2:1641).[3]

As was often the case in this study, the teachers mentioned the names of their coursebooks when they were asked about the methods they used when teaching. The pattern for listening was similar to what we saw when we looked at the teaching of reading: Both T1 and T2 followed the routine of introducing a section or task type carefully and then asking the students

to do practice exercises. T1's students worked on their own, writing their answers down on paper, and when they were finished she led a discussion of the answers. T2's students worked on practice test material at their individual computers and checked their own responses when they finished each section.

T4 once again differed in his approach, using prelistening tasks, asking the students to work together in pairs or groups, and leading discussions after they finished their activities, including discussing their reasons for discarding distracters as well as their reasons for choosing correct answers. His work was more interactive and seemed to be more motivating than his teaching in Phase 1, which had consisted mainly of students calling out the letters which represented what they thought were the correct answers and T4 indicating whether they were right or wrong.

Findings from Phases 2 and 3 had indicated that we might see the teaching of note-taking in Phase 4 classes, but none of the teachers devoted any time to developing the students' ability in this area. T1 spoke at length about her experience, but summed up the problem in this way: "It's often extremely difficult for them to both listen effectively and put down notes that make sense" (8:1447).

T1 found that teaching note-taking was more complicated than she had first thought. She was not convinced that the systems suggested in the coursebooks, which included abbreviations and symbols, were necessarily helpful, being in effect a whole new language that students had to learn if they were not already familiar with it. (8: 1448)

Summary. There were changes in the content of listening classes insofar as the coursebooks the teachers were using reflected the passages and the question types found on the TOEFL iBT exam.

There were no changes in methodology in the classes given by T1 and T2, but T4's classes included more student-student interaction than the classes we observed in Phase 1. The type of teaching he was doing in Phase 4 seemed shaped by his coursebook, which though directed at practicing for the iBT did not conform to the explanation and practice pattern of other test preparation coursebooks.

The Teaching of Writing

Writing is the section that underwent the biggest change in the switchover from TOEFL CBT to TOEFL iBT.

- The CBT writing task was retained in the same form (essay in 300 words) and with the same topics, but it became known as the "independent writing task." This task would assess the candidate's ability to state a preference or give an opinion.
- A second task was added, which required the candidate to process input from a reading text and a listening passage and write on some aspect of the relationship between them. This was called the *integrated writing task*.
- The scoring rubric for the independent writing task was similar to the CBT scoring rubric but was more detailed and required more detail in the candidates' writing.
- A new scoring rubric was introduced for the integrated task.
- Students were required to type their responses rather than being allowed to choose between writing them by hand and typing them (ETS, 2005a, p. 21).

Rationale for change. The main aim of adding the integrated writing task to the TOEFL writing test was to "move beyond the single independent essay model to a writing model that is more reflective of writing in an academic environment while also dealing with interdependency issues" (Cumming et al., 2000, p. 9).

The main objection to the notion of combining reading and/or listening input with writing output is that if the writing output is poor it is difficult to determine whether the problem lies in the candidate's writing abilities or in the ability to understand the inputs properly and to process them in the required manner. It is clear, however, that candidates taking the TOEFL to enter institutions of higher learning will need to deal with complex subject matter in their writing, explaining what they have heard and/or read (at a minimum) and (probably) transforming it in some way. The decision to include integrated tasks in the new test represented the triumph of authenticity over traditional worries concerning score interpretations.

Intended impact. The framework document for writing does not give details of the sort of impact the new TOEFL should have on classroom practice, stating only that "multiple writing tasks that include both independent and content-dependent tasks" should produce the type of writing that realistically reflects the target language use situation (Cumming et al, 2000, p. 9). The experts we consulted in Phase 1 gave a little more detail about desired impact: that there would be an emphasis on summary and paraphrasing skills and that teachers would work with

their students at a discourse level rather than focusing on decontextualized grammar and vocabulary (Wall & Horák, 2006, p. 15).

Findings from Phase 1. About a third of the teachers considered the writing section of the CBT to be the most difficult for their students (Wall & Horák, 2006, p. 59). This was not because the students lacked the language they needed for writing, but because they found it hard to organize their ideas in a coherent way. The teachers spent much of their class time working on how to structure a piece of writing, focusing on the organization of paragraphs and on notions such as the *five-paragraph essay* (thesis statement to open the first paragraph, two to three supporting paragraphs, and a conclusion at the end; p. 60).

The teachers did not pay as much attention to content as they did to structure. They made use of lists of topics they found in their preparation books and on the Internet, but they did not provide written or aural material on which students could base their ideas. The only teachers who mentioned using written materials used them as models of how to organize writing, rather than as input to what might be seen as integrated writing (Wall & Horák, 2006, p. 64).

Several of the teachers set aside class time for actual writing (as opposed to talking about writing), but for differing reasons (Wall & Horák, 2006, p. 63). One teacher wanted to ensure that students got into the habit of writing and overcame their fear of the blank page. Others felt their students would not do any writing unless they were made to do so in class. Some teachers preferred to assign writing for homework, but this meant that they had to leave it up to the students to time themselves and get used to writing quickly.

The most common way of assessing writing was to write comments on the students' papers, reacting to their individual problems in an ad hoc way (Wall & Horák, 2006, p. 64). The native English-speaking teachers drew on their own experience of how writing was marked in university settings, but the local teachers relied on what their coursebooks suggested was acceptable writing. Most teachers claimed that they and their students were familiar with the TOEFL scoring rubric, but the teachers did not make much use of the rubric in either their teaching or their marking.

Findings from Phase 2. Although one of the aims of Phase 2 was to investigate how teachers learned about the new TOEFL iBT test and whether they understood what they were learning, it was necessary about halfway through Phase 2 to take them through the writing section in detail so that we could see how they reacted to it and what difficulties they might have

preparing students for it in the future. This process gave us confidence in the latter part of this phase, and in Phases 3 and 4, that the teachers were aware of the format of the new writing section and what it demanded.

The teachers were generally positive about the expansion of the writing section (Wall & Horák, 2008, p. 46), believing in the importance of writing in the academic environment that most of their students were aiming to enter. Most of the teachers perceived the independent writing task to be the same as the CBT writing task, but they appreciated the changes in the scoring rubric. They were mainly positive about the integrated task as well, commenting on its authenticity and the fact that it tested skills that were different from the ones required in the independent task. Summarizing and making connections between ideas were mentioned specifically. At least one teacher was worried about the possibility of plagiarism, however. Most of the teachers were happy with the scoring rubric for the integrated task and were confident that they would be able to use it correctly in the future. Their first attempt to apply the rubric was unsuccessful though, as none of them gave the same grade to a piece of writing as a TOEFL writing expert had done. (We did not inform the teachers of these results so as not to contribute more than we were already doing to their awareness of test demands.) This exercise suggested that the teachers would need to become more familiar with the rubric and practice using it before they mastered it. This point was one that we felt we needed to explore further in Phase 4.

Generally speaking, the teachers had not planned changes in how they would teach independent writing. They had more to say about the integrated task and discussed specific aspects on which they planned to focus. One teacher was interested in developing the students' note-taking abilities, another two planned to focus on comparing the reading and listening inputs, and a fourth thought she should sensitize her students to the issue of plagiarism. The teachers were clearly actively wondering what the best approach would be to develop what they saw as new skills, but they had not come up with concrete ideas for teaching at this stage of the study.

Finally, several of the teachers mentioned that they should make sure their students were familiar with the scoring rubrics for writing. Two mentioned the possibility of getting their students to assess each other's writing. There was some feeling, however, that the teachers themselves would need to understand the scoring rubrics better and receive more guidance on how to use them.

Findings from Phase 4. The findings from Phase 4 are presented in Table 13. All three teachers claimed to dedicate a similar proportion of their class time to writing, namely 15-20%. Given the nature of the changes in the writing test, this figure would seem to be low, but what it hides is the amount of time students were expected to devote to writing outside the classroom. T1 indicated that she gave her input (two lectures—one on independent writing and one on integrated writing) early on in her course and expected the students to do writing homework thereafter. She did not feel that asking the students to write during class time would be a good use of the limited time they had together. She also wondered what she would do when they were writing (8:1283). T2, in contrast, did expect her students to write in class as part of the many practice tests they did on the computer. While they were working, she wrote comments on writing they had produced earlier and made sure her records of their results on the practice tests were up to date. The low percentage figure also hides the amount of time that the teachers devoted to reading the students' work and commenting on it. This time commitment is discussed in more detail below.

Table 13

Phase 4—The Teaching of Writing

Characteristics of teacher and teaching	T1	T2	T4
Teacher's awareness of new test	Good awareness (8:45 and 457)	Good awareness	Good awareness (1:348)
Teacher's attitude toward new test	Positive	Positive	Positive
Percentage of class time dedicated to skill	15% + much writing homework (7:138)	20% + some writing homework (4:136)	15% + writing homework (1:75; 2:1175 -1182 and 1734)
Content covered	List of topics from ETS for independent task (8:1520) –tackled easy topics first (8:1499) Focuses on essay structure (8:1288)	List of topics from ETS for independent task (4:136)—tackled harder topics first (5:628) Focuses on essay structure (5:1048)	List of topics from ETS for independent task (2:1754)—no grading of tasks Focuses on essay structure (2:1026) Used formula approach for independent writing (2:995)

Characteristics of teacher and teaching	T1	T2	T4
Methods used	Briefed students about tasks, via interactive lectures Students generally did writing out of class, as homework Teacher gave generous feedback (8:430-438; 14:348)	Briefed student about tasks, via lectures ('the theory') Students did practice tests in class Students also wrote out of class for homework Teacher gave generous feedback (4:119)	Briefed students about tasks Students did some practice tests in class (1:142, 2:1754) Students also did writing as homework Also Used pair and group work to brainstorm ideas for content, as in general language development classes (1:134) Students were asked to write an essay at beginning of course, before receiving any input—so that they could see progress later (2:1739 and 1817)
Use of marking rubrics	Ensured students were familiar with marking rubrics (7:173) Used rubrics to analyze sample responses in class (7:127) But did not use rubrics to mark students' work (7:140)	Ensured students were familiar with marking rubrics (4:167) Had designed own marking sheet incorporating criteria from rubrics (5:1575) Gave students marks based on her understanding of rubrics (4:138)	Ensured students were familiar with marking rubrics (1:192) Gave students marks based on his understanding of rubrics (1:159)

Characteristics of teacher and teaching	T1	T2	T4
Reacting to student writing	Students submitted work via e-mail, even after course was finished Teacher spent many hours marking at home (8:934) Commented on organization, appropriacy of ideas, language. Wrote guiding questions and suggestions for improvement. Sometimes wrote sample paragraphs. (14:176)	Students printed off essays they had written in class (practice tests on computer) Students could also submit work via e-mail, even after course had finished Teacher spent many hours marking at home, and in class while students were doing practice tests on computer. (5:86, 150, 638 -660, 827 and 1624) Took up to 2 hours a day to provide thorough feedback (5:151) Gave advice on common mistakes and how to improve (5:87)	Students submitted work via e-mail, even after course was finished (2:1760) Teacher did not correct everything (2:2036) Commented on structure of essay, and underlined mistakes in vocabulary and grammar (2:1995)
Advice given	Students should write as much as possible—"learning by doing" (5:157)	Students should read as much as possible	Students should read as much as possible. He gave names of magazines he thought would help them most. (2:1174) Do homework in examlike conditions (2:1180)
Other aspects	Typing—not an issue (observed)	Typing—not an issue (observed)	Typing—Some students not fully confident, but teacher referred them to typing software available in institution (2:1839)
Has there been any change since Phase 1?	Change in content Change in methods Change in use of marking rubrics	Change in content No change in methods Change in use of marking rubrics	Change in content Change in methods Change in use of marking rubrics

Note. The numbers in parentheses refer to the transcript and line where information can be found. T1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4.

Our limited observation time (two lessons per teacher) did not allow us to see any teaching of independent writing, but the descriptions that T1 and T2 gave of their approach did not seem to differ greatly from what was described and observed in Phase 1. The teachers provided briefings about the demands of the task and then asked their students to put what they had learned into practice. T4's class seemed much more interactive than in Phase 1, with brainstorming exercises and student-to-student discussions of ideas that could be used in the essays they would write for homework.

T1 displayed a fresh approach when she was working with her students on integrated writing, however. The lecture she prepared was actually an interactive session in which she got students to analyze the demands of a specific writing task, take notes on the written and oral inputs, process the information in the way required and build up an outline together of what their responses should look like. Though she did not ask the students to work in pairs or groups, she managed to generate a lot of communication between them by getting them to listen and react to each other's contributions in plenary. This was quite different from her style of teaching writing in Phase 1. We did not observe T2's initial briefing on integrated writing, but we were present when she gave general feedback to her group on a task that they had worked in the previous session. Her main messages were that they should include more ideas from the listening input as this was the input that really mattered, they should provide more examples for their main points, and they should be careful with their choice of cohesive devices.

What the teachers had most in common in Phase 4 was the use they made of the scoring rubrics, and it was here that the practice of all three teachers differed considerably from their practice in Phase 1. None of the teachers in the Phase 1 sample had made much use of the CBT scoring rubrics, or even displayed much interest in them, but the Phase 4 teachers were familiar with the TOEFL iBT rubrics and reported that they found them useful. All three teachers strove to ensure that their students were aware of the criteria in the scoring rubrics and that they knew how they would be applied. T1 included an analysis of sample responses in one of her sessions as a means of illustrating the standards expected (7:127) and hoped that this process would help students to prioritize what they needed to develop in their own writing (8:1605). T2 had designed a scoring sheet listing "all of the things that are required for a perfect essay," which she used when marking her students' writing (5:1575). T4 also gave marks based on his understanding of the scoring rubrics.

Although T1 made sure that her students were familiar with the rubrics, she reported that she did not feel sufficiently confident to give the students marks based on them (7:140). This was not a problem with the rubrics, however. She felt that giving a mark was too much like "a legal act, something in black and white" (8:1598), and she did not like the responsibility that such an act entailed. She had felt uncomfortable about giving marks in Phase 2 as well (83:26, 84:11, 89:119). What is most interesting, however, is not T1's lack of confidence but rather the confidence the other teachers had in their own judgments. Recall that in Phase 2 none of the teachers had given the same mark to a sample of student writing as a TOEFL writing expert had and that a number of the teachers said they would appreciate more guidance on how to use the rubrics in the future (Wall & Horák, 2008, pp. 32–34).

All three teachers acknowledged that their workload had increased with the new version of TOEFL, now that two types of writing were required (T1, 8:2282; T2, 5:1624; T4, 2:2646). T1 and T2 reported spending many hours of their own time on marking. T1 explained her situation in this way:

> Almost every day of the week I have to do. . . at least five or six of them. . . so I come home from work and basically that's all that I do until I go to work again, because (if) you have a group of six people, if all of them write essays, which I encourage, and they mostly do it, then you get 12 essays before each class, so that's a lot. (8:934)

T2 said that she could spend up to 2 hours a day marking writing (5:151). All three teachers were willing to mark student work even after their courses had finished, accepting this as part of their job.

What we do not know, because the Phase 4 investigation focused on teaching rather than learning, was what the students made of the feedback they received and whether it helped them to develop their writing in the right direction.

The final point to note about the teaching of writing had to do with the students' need to type essays for the TOEFL iBT exam. None of the teachers believed that this was a serious issue, something we found surprising considering that two of them taught students whose first language (L1) was written in a non-Latin script. Observations of the classes where students were doing computer practice tests confirmed that they were able to type quickly, and inspections of some of the students' writing showed that the typing was accurate. Some of the teachers in Phase 1 had considered typing to be a potential problem for their students. It is not known whether in the few years since our study began computer use had reached such levels that the majority of

likely TOEFL candidates in this region were typing fluently, or whether the level of skills (or lack of skills) of some of the teachers in earlier phases of the study had influenced their perceptions of students' typing problems.

Summary. The teaching of independent writing did not seem to have changed from what we observed in Phase 1. The teaching of integrated writing involved the students in careful analysis of reading and listening inputs, and in the case of two teachers' classes, this analysis led to discussions of what the students had understood and how they would use the information in their writing. All three teachers were aware of the scoring rubrics and incorporated them in their teaching in some way.

The Teaching of Speaking

The most notable change in the switchover from CBT to TOEFL iBT exam was the inclusion of a test of speaking. Speaking was not a compulsory part of either the CBT or the PBT; if candidates needed a grade for speaking they had to take the TSE, which was associated with the TOEFL exam but not part of it.

The TOEFL iBT speaking test includes these features:

- There are six separate speaking tasks: two independent tasks and four integrated tasks.
- In the independent tasks (Tasks 1 and 2), the candidates respond to a spoken prompt. They have 15 seconds to prepare their responses and 45 seconds to perform.
- In the integrated tasks, candidates process information they have received either through reading and listening (Tasks 3 and 4) or through listening only (Tasks 5 and 6), and comment on some aspect of this information in speaking. They have 30 seconds to prepare for Tasks 3 and 4 and 20 seconds to prepare for Tasks 5 and 6, and then up to 60 seconds to respond to each task.
- There are different scoring rubrics for independent speaking tasks and integrated speaking tasks.
- The test is computer-mediated. The candidates listen to the prompts and input through headphones and deliver their responses via a microphone.

Rationale for change. One of the main criticisms of the PBT was of the assumption that candidates' results in the reading and listening tests could indicate their abilities in writing and speaking. Traynor (1985) suggested that "one could score well in the TOEFL without being able to say a single word in English or write a single word other than one's name" (p. 44). The addition of a compulsory written test (the formerly optional or TWE®) to the CBT in 1998 addressed the second of these two points, but not the first. The new speaking test was meant to "meet score users' expressed need for information about examinees' English oral language proficiency in an academic context" (Butler et al., 2000, p. 23). The test would "simulate realistic communicative situations" (p. 23), and it would include integrated tasks to reflect as much as possible the candidates' target language use situation (p. 16). It would be possible to build on the thinking that had gone into the TSE regarding the types of functions that should be assessed, and it seemed fruitful to include the sorts of integrated tasks that had been suggested for the new test of writing.

Intended impact. As was the case for the other three skills, we searched the framework documents for statements about intended washback. There was only one statement, which appeared in this form: "By using constructed response items, which are less likely to be coachable, in the TOEFL 2000 speaking component, we will encourage students to learn to communicate orally—not to learn a skill simply to do well on a test" (Butler et al., 2000, p. 23).

We also consulted some of the experts who had served as advisors in the early stages of the design of the new test. Their responses about hoped-for washback were also quite general—that speaking would be taught (two respondents), that there would be more emphasis on productive skills (two respondents), and that students would learn about the pragmatic force of utterances (one respondent). We examine whether those predictions were met later in this section.

Findings from Phase 1. A considerable amount of English was being spoken in all but one of the classrooms we visited. What soon became apparent, though, was that speaking was used as a way of practicing other skills, and little to no attention was paid to develop speaking in its own right. Several reasons were noted for using English as a medium even though it was not tested. There was some feeling that students should practice their speaking since they would need it in their target language use situation. There was also some feeling that using English in the classroom would give the students valuable listening practice. Also important was the fact

that several teachers (the expatriates) did not speak their students' first language, and there was no other option but to communicate in English. In no case did the students speak extensively, however. They mainly responded to their teachers' questions or requested explanations or help with tasks focusing on other skills.

The one thing all the teachers and all but one of the directors of studies knew about the new version of TOEFL (which had not been much marketed at that point) was that it would include a test of speaking. We inferred that this was because the addition of speaking was such a striking change and would, in the teachers' minds, imply the most change in their future classroom routines.

Findings from Phase 2. Our first contact with the teachers at the beginning of Phase 2 revealed that some of them had a faulty understanding of what the speaking tasks would entail (e.g., real-time interaction with a native speaker of English or speaking over a telephone). Everyone's awareness grew as more information appeared on the TOEFL Web site and as the teachers proceeded through the various tasks we set them. For the penultimate task they had to listen to speaking performances at different levels and use the scoring rubric to grade them.

Speaking was the section of the new test that the teachers commented on most, which was not surprising given its novelty. Their attitude toward the idea of testing speaking was generally positive. They mentioned some concerns about the task types, however, once they were more familiar with the format of the test. There were comments about the limited time the students would have to respond to each task and how this would put them under pressure. One of the teachers was concerned that the tasks would not elicit the speaking needed in an academic setting. She would have preferred a human interlocutor and tasks requiring interaction rather than monologic responses.

The teachers were generally satisfied with the scoring rubrics, though they did not find them as easy to work with as the writing criteria. Their difficulties may have been due to the fact that speaking is ephemeral, whereas they could read a piece of writing several times before marking it. There were quite a few questions about the criteria and their weighting. When the teachers were asked to mark the speaking performances, their marks did not match those of the TOEFL expert rater (we did not inform them of these results, however). They gave a range of responses when asked how confident they felt using the rubrics to mark the speaking samples. They were less secure about marking speaking than they were about marking writing. Using

criteria to mark speaking would be a new activity for most of them, whereas they had probably thought about (to some extent at least) whether and how the CBT criteria could be used for marking writing.

The teachers also had some concerns about how they would prepare their students for the speaking tasks. They were concerned about how to provide a testlike practice environment (this was a worry of the teacher called T2 in this phase), how to get students used to talking into a microphone, how to standardize their own marking with that suggested by ETS-marked samples of speaking performances (T1 was one of those who worried about this), and what model of pronunciation to encourage.

Findings from Phase 4. Our findings from the fourth, and final, phase of the study indicated that all three teachers were still positive about the speaking test and had worked out how to teach toward it. The findings from this phase are presented in Table 14.

Table 14

Phase 4—The Teaching of Speaking

Characteristics of teacher and teaching	T1	T2	T4
Teacher's awareness of new test	Very aware	Very aware	Very aware
Teacher's attitude toward new test	Positive	Positive	Positive Finds speaking section hardest to teach (1:364)
Percentage of class time dedicated to skill	35% (7:53) + English was medium of instruction (7:96) Students sometimes used L1 (7:101)	20% (4:52) + English was medium of instruction (4:94) Students often used L1 (4:100)	35% (1:60) + English was medium of instruction (1:177) Students used English (1:121)
Course aim	Build students' confidence Get students used to speaking in front of others (7:107)	Give students as much testlike practice as possible (5:132)	Build students' confidence (2:206 and 2449) Include as many speaking opportunities as possible, when practicing all skills (2:796)

Characteristics of teacher and teaching	T1	T2	T4
			Encourage fluency first, accuracy later
Materials used	iBT2—to introduce tasks and to give test practice (observed)	iBT5—to introduce tasks iBT2—for practice tests Software developed by colleague to simulate experience of speaking exam (4:215)	iBT5—to introduce task and explain characteristics of responses at each level iBT3 and iBT4—to build up confidence through many speaking opportunities iBT0 for test practice (1:145) iBT2—computer-based practice
Content covered	Introduced requirements Organizing thoughts on topic Providing examples Providing support for arguments (8:950)	Introduced requirements Moved quickly into testlike practice (4:128, observed)	Introduced requirements (2:2453) Organizing thoughts on topic (2:1026 and 1201) Time management (1:66, 2:1201) Students did testlike practice (2:404 and 2153)
Methods used	Teacher gave interactive lecture on how to organize input Students were asked to perform tasks in front of other students. Three students were asked to do each task, so each student should show improvement over the former student Teacher gave oral feedback to student and group (observed)	Students did computer practice on own (observed) Students spoke into microphones at same time (5:44, observed) Teacher gave written feedback on speaking performance to each students and to group (observed)	Much student-student interaction Pair/small group brainstorming before tasks to raise ideas (1:134) Students gave short talk in each lesson, to get used to talking to group (2:1211) The teacher worked up to practice tests on computer (2:404)

Characteristics of teacher and teaching	T1	T2	T4
Use of marking rubrics	Teacher felt confident using rubrics (7:168) Introduced rubrics to students (7:173) Used rubrics in feedback, but did not give grade (8:1592)	Teacher felt confident using rubrics (4:163) Introduced rubrics to students (4:156) Used rubrics in feedback, giving grade (5:174)	Teacher felt confident using rubrics (1:188) Introduced rubrics to students, explaining what each descriptor meant (1:192; 2:2-57) Discussed graded responses in iBT5 (1:179) Used rubrics in feedback, giving grade (2.1217)
Reacting to student speaking	Teacher gave individual feedback, orally, immediately after each student performed a task (observed)	Teacher gave individual feedback, in writing, the day after listening to students' recordings (25:145) Gave brief feedback to whole group, the day after listening to recording (observed) Filled in "score sheet" for each student (5:174, observed) Emphasis on organization of ideas (5:1413)	Correction given through feedback (2:115)
Has there been any change since Phase 1?	Content—Speaking was focused on explicitly. Methods—Change. More teacher-student interaction, though limited student-student interaction.	Content—Speaking was focused on explicitly. Methods—No change.	Content—Speaking was focused on explicitly. Methods—Major change. Much higher level of interaction.

Note. The numbers in parentheses refer to the transcript and line where information can be found.

T1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4. Coursebooks are identified in Table 3.

Time spent on speaking. T1 and T4 devoted more than a third of their teaching time to speaking. T4 felt this was necessary because the students needed a great deal of practice (2:1196). T2 devoted about 20% of class time to this skill. All three teachers spoke to their students in English and they all claimed that their students mostly used English as well. T1 added that her students sometimes switched to their mother tongue, however. T2's students did most of their speaking when they were responding to prompts on the computer practice tests—and this was in English—but when they sought clarification about problems they seemed to feel more comfortable speaking in their first language.

Course aims. T1 and T4 held firm beliefs about the need to give their students confidence in speaking. T1 felt she had to get the students to speak in class as this might be the only place where they had the opportunity to practice. She got individual students to perform speaking tasks in front of the group, believing that the more often they did this, the more comfortable they would feel: "The more confident they feel the more confident they will sound and the more fluent they will sound and those people who listen to them will say, 'Well, this person communicates with confidence'" (8:1793).

One of the reasons T4 concentrated on confidence building was that a local "myth" deemed the speaking section very difficult (1:177). He tried to include as many opportunities for speaking in his classes as possible (2:796), using discussion activities to introduce reading texts or listening passages, for example. He called these activities "conversation as a warm-up" (2:28). He saw them as a means of not only activating the students' background knowledge, which would help them to better understand the texts or passages they were about to tackle, but of "warming them up" for later speaking activities (2:744). T4's philosophy was "little and often": after the first few classes each student had to present a short talk in each lesson on the topic offered by the unit of the book on which they were working (2:1211). T4 was sure that if his students got to a point that they felt comfortable talking in front of the whole group, they would feel comfortable talking into a microphone (2:2484).

T4 also held back from correcting his students early in the course, allowing them to develop their fluency and confidence. He increased his attention to accuracy as time went on, eventually building up to awarding the students a grade using the scoring rubrics (2:1217). He did not ask them to do practice tests in early classes, waiting instead until he felt they were more comfortable with the test requirements (2:404).

T2 did not comment on the issue of building students' confidence, even though she had noticed that none of the students wanted to be the only one speaking when they worked their way through practice tests at the computer. They started each day's practice test (which contained sections for all the skills) at approximately the same time, and as the tests were timed, they finished each section within a minute or two of each other. Rather than beginning the speaking section as soon as they could though, "they wait for each other. If somebody finishes earlier they wait for the others to finish so that they start speaking altogether" (5:44).

Materials. T1 mainly used the iBT2 book for speaking, as she did for all skills. T2 used the iBT5 coursebook in order to explain the speaking tasks, and then used iBT2 for practice tests. T4 used four different coursebooks for four different purposes. For building up student confidence he used iBT3 and iBT4, which were the books that stood out as being most similar to coursebooks used for general English as a Foreign Language (EFL) teaching.

T2 also used a software program that had been designed by a colleague to allow students to feel that they were in an authentic test-taking situation. The students responded to speaking test prompts and recorded their responses, which T2 collected and commented on in her own time after the class session.

Content. All three teachers spent time introducing the requirements of the tests. T1 and T4 concentrated on helping their students to organize their ideas before they began speaking. T1 also emphasized the need to use good examples and supporting arguments, while T4 tried to help the students do this in a short amount of time. T2 gave little of this sort of input, preferring instead to give her students practice doing tests at the computer. The students recorded their performances and T2 commented on a variety of features in her written feedback, in response to each student's problems.

Methods used. As we have seen, T2's approach to teaching was to make sure her students did as much computer practice as possible. She spent a few minutes at the start of each session lecturing the students on common problems she had come across when marking their previous day's recordings, but the rest of the time the students worked alone at their computers.

T1's explanations of what was required in each speaking task were interactive, in the sense that she asked the group to respond to questions she had prepared about the ideas they should use and how to arrange them. She proceeded through the speaking test task by task, following the presentation in her coursebook, and asked three students to perform each task in turn. The first

student would perform and she gave immediate and direct feedback, the second student would perform and get feedback, and the third person would perform and get feedback. T1 usually varied the order in which students performed, as she realized that the first student to perform had the hardest job. She made an exception, however, in the case of weak or shy students, asking them to perform last so that they could benefit from the feedback the other students had received.

T4 also got students to perform speaking tasks in front of the whole group, but he gave the students a chance to brainstorm ideas before they began performing. The students started with short talks and gradually built up to being able to give longer talks.

It was in their choice of methods that the teachers differed most, with T2 asking her students to work alone on the computer most of the time, T1 encouraging speaking but producing mainly teacher-student interaction, and T4, who encouraged a great deal of student-to-student interaction.

Use of scoring rubrics. All three teachers stated that they felt confident using the scoring rubrics for speaking, and all of them made a point of introducing the rubrics to their students. T4 felt it was important for his students to know "how high the bar is" (1:186). He used the rubrics in combination with the graded responses in the iBT5 coursebook to show the students how the criteria were applied in practice.

T2 and T4 gave grades when they commented on their students' speaking, basing their judgments on the rubrics. T2 regularly spent 45 minutes to an hour outside class time listening to the recordings the students made when they did their practice tests and preparing written feedback on their performances. She filled in a score sheet for each student for each performance, so that "at least they could have this as a reference, for what they should work on or what they shouldn't do" (5:174). She felt she had to give detailed comments since, "speaking and writing are the sections where students cannot mark themselves. They can't get real information on how they're doing if they don't do it with a teacher." (5:677)

T1 had tried to get her students to give peer feedback (she was experimenting with this practice in Phase 3), but she had not found it very successful. The students were not comfortable commenting on their classmates' performances and she had come to sympathize with them. She had not realized at first how hard it was for some of them to speak up in front of others. She described one case in particular, where a student was "red as a radish." She continued to ask

students to perform but felt it was less painful for them if they only received her feedback (8:1888). Neither of the other two teachers reported using peer feedback.

T1 now reacted to each student's performance immediately after the student spoke, giving details that she hoped would help the individual and the rest of the group as well. She admitted that it was hard to do this, and that she had to concentrate to remember which points to raise once they finished. She had previously tried making notes while the students were talking but she felt this made them uncomfortable. She knew that she did not always target the most important features: "sometimes, somebody says some really, really weird stuff and it's difficult to remember that later" (8:959).

Changes. It was clear from the interviews and from our observations that there were substantial differences in the way speaking was dealt with in Phase 1, before the new test was introduced, and Phase 4, approximately a year after its launch in these countries. The main change was that the teachers were focusing specifically on developing their students' speaking skills rather than using speaking only as a vehicle for communication. It was especially interesting to see the type of preparation the students did for the integrated speaking tasks, which involved taking notes on the ideas emerging from different sources, weighing up the information to see which ideas were most relevant to the specific question, and condensing the information so that it could be transmitted in the limited time available. It was also interesting to see what use the teachers made of the scoring rubrics, and how, as was the case for writing, the rubrics had gained an importance that they did not enjoy in Phase 1.

As far as methods of teaching were concerned, T4 showed the most change in that he encouraged his students to talk not only to him but also to their fellow students. Much of T4's practice seemed related to his choice of coursebook. T1's students did not interact with their classmates as frequently as T4's, but they interacted considerably more with her than students in Phase 1 had done. In fact, in Phase 1 she had not encouraged students to practice speaking at all, believing that "if they wanted to learn how to speak English correctly and fluently, they should take another course" (8:473–486).

T2's methodology had changed the least. Most of her class time was devoted to individual testlike practice on the computer, with students spending little time on speaking. When they did speak it was into a microphone rather than to their teacher or their classmates.

Summary. There was no doubt that there was more of a focus on speaking in Phase 4 than in Phase 1, and in this sense the impact that the TOEFL experts said they hoped would occur was achieved. However, we saw little evidence that the pragmatic force of utterances was being studied. None of our teachers mentioned this feature in relation to developing speaking skills, and it is hard to see how it could be assessed since the monologic responses that students were required to give did not require this sort of sensitivity.

If we return to our earlier definition of a communicative classroom as one with a wide variety of interaction patterns and opportunities for genuine spontaneous, meaningful communication, then we would have to declare T4's classes as the most communicative of the three in our study. They represented a marked change from what we observed in our visit to his Phase 1 classes.

The remaining hope about intended impact was that the constructed response format required in the speaking test would not be coachable. Since the student output is a monologue, not interaction in a (semi) conversational style as in other international exams (e.g., Cambridge exams), teachers could be tempted to get their students to memorize set pieces of language. We witnessed no such practice in Phase 4, however, and were in fact quite impressed with how spontaneous and meaningful the students' responses were, in spite of the fact that they were preparing to talk to a computer.

The Teaching of Grammar and Vocabulary

The original version of TOEFL (1964) contained both grammar and vocabulary sections, made up of multiple-choice items. The vocabulary section was phased out in 1995, but the grammar test (referred to as "Structure") remained until the introduction of the new TOEFL. Read (2000) describes the earlier design as follows: "The inclusion of structure and vocabulary as separate sections reflected the discrete point approach to language testing that prevailed in the US at the time the test was originally designed" (p. 139).

In the TOEFL iBT exam, grammar and vocabulary are no longer tested on their own. Understanding how the language is structured is assessed indirectly as part of the reading and listening sections. It assumes more prominence in the writing and speaking sections as it forms part of the scoring rubrics for both skills; however, it is only one of a number of criteria that are used for marking student performance in both cases. Similarly, vocabulary plays a role in the

context of the assessment of reading and listening and is one of a number of criteria in the scoring rubrics for writing and speaking.

Rationale for changes. The atomistic approach to language that characterized particularly the PBT version of TOEFL no longer matches modern views of language proficiency, and one of the main reasons for revising the TOEFL was to shift the emphasis from language knowledge to "more complex, performance-type assessment tasks" (Chapelle et al., 2008b, p. 3). The TOEFL designers also wished to respond to worries expressed by the language teaching community: "ESL/EFL teachers are concerned that discrete-point test items, and the exclusive use of traditional, multiple-choice items to assess the receptive skills, have a negative impact on instruction" (Jamieson et al., 2000, p. 3).

Read (2000) described a situation in which students were encouraged "to spend time unproductively learning list of words and their synonyms," which were often "uncommon or esoteric" and "not likely to be useful for foreign students in pursuing their academic studies" (p. 140).

Intended impact. No framework document existed for either grammar or vocabulary so there were no explicit statements about how the elimination of the structure section would affect future teaching. This point was not raised by the experts we surveyed either, although one reported that that there had been a hope that new approaches to the testing of writing would encourage work at a discourse level rather than attention to "decontextualized grammar and vocabulary" (Wall & Horák, 2006, p. 15). This statement implied a belief that this was the sort of work being undertaken in preparation for the earlier versions of the TOEFL.

Other members of the language teaching community have since made more explicit predictions about what impact the decision to do away with the Structure section might have. Rogers, for example, the author of several TOEFL preparation coursebooks, wrote that "test prep will no longer focus on memorizing individual vocabulary words and idioms and mastering unrelated grammar points. It will have to focus on understanding and producing larger chunks of language" (Rogers, 2004, p. 39).

Findings from Phase 1. Students who enrolled in TOEFL preparation classes had normally already studied a great deal of grammar, but most teachers had to pay at least some attention to grammar in their classes. For some this was a matter of brushing up only, but several teachers said that grammar was the component on which they spent most of their time. The

techniques that they used included explaining particular grammar points (some which they considered tricky), going through testlike exercises in the coursebooks, occasional drilling, and paying attention to grammar mistakes in the students' writing. There were no inductive or task-based activities of the sort often seen in communication-oriented language coursebooks (Wall & Horák, 2006, pp. 43–46).

Teachers held a clear belief that a rich vocabulary was important for success on the TOEFL, even though there was no separate vocabulary section on the CBT. They used two main ways of teaching vocabulary. The first was by distributing lists of words and phrases to students and asking students to memorize them. The items were generally, but not always, accompanied by some supporting information such as a sample sentence, a definition, or an indication of the pronunciation. The teachers did not have a common approach to selecting these items, however. Some chose synonyms, others cohesive devices for writing, and one chose rare words that his students would enjoy playing with (Wall & Horák, 2006). The second way of dealing with vocabulary was via reading. Several teachers mentioned the importance of seeing vocabulary in context, but while a few seemed to encourage their students to work out meaning for themselves, at least one teacher asked students to look up the meanings of unknown vocabulary before their reading lessons. We have already seen that some teachers asked their students to do their reading exercises in their own time, so it is not clear how these students dealt with vocabulary. Students were generally encouraged to read extensively; some also used CDs and the Internet to build up their stock of words and phrases.

Findings from Phase 2. The teachers who participated in Phase 2 had differing views on whether and how they would teach grammar in the future, but it seemed likely, given the elimination of a separate grammar section and the need to focus on integrated skills and speaking, that they would want to reduce the time they could spend on this aspect of language. Two teachers felt they would still have to devote some time to grammar though, since it was important for the other skills, especially writing. They did not have specific plans in mind, however, mentioning only the need to do some explicit teaching and revision and to pay attention to grammar when marking student writing. None of the teachers talked about their plans for teaching vocabulary in the future.

Findings from Phase 4. The findings from Phase 4 are shown in Table 15. The teachers reported that they spent far less time teaching grammar for the TOEFL iBT exam than they had

spent in their CBT classes. T1 and T2 both estimated that they had previously spent 20% of their time on grammar, and T4 said that he had spent over half his time on this aspect of language. None of them spent more than 5% of their time on grammar for the TOEFL iBT exam. All three teachers spent about the same amount of time teaching vocabulary as they had done before (T1 slightly less time; T2 and T4, slightly more).

Table 15

Phase 4—The Teaching of Grammar and Vocabulary

Characteristics of teacher and teaching	T1	T2	T4
Teacher's awareness that grammar and vocabulary not tested separately	Very aware	Very aware	Very aware
Teacher's attitude toward this change	Positive	Positive	Positive
Percentage of class time dedicated to grammar	2% (down from 20%)	5% (down from 20%)	1% (down from 55%)
Percentage of class time dedicated to vocabulary	3% (down from 5%)	15% (up from 10%)	4% (up from 0%)
General approach	Needs-based—responding to problems and questions	Needs-based—responding to problems and questions	Needs-based—responding to problems and questions
Materials used	Hand-out on working out the meaning of words in context	None	Occasional teacher-made handout on points students asked for at beginning of course
Content covered	Ad hoc	Ad hoc	Ad hoc
Methods used	Grammar—teacher responded to students' queries; noted down common mistakes when marking writing, and	Grammar—teacher responded to students' queries; noted down common mistakes when marking writing and speaking and gave explanation to whole	Grammar—responded to students' queries (1:79); prepared hand-outs in response to concerns expressed at beginning of course and discussed these

Characteristics of teacher and teaching	T1	T2	T4
	gave explanation to whole group (8:1293) Vocabulary—supplied definitions, encouraged students to figure out meaning from context (8:1300)	group (5:470) Vocabulary—encouraged students to figure out meaning from context as they read, but then checked meanings after reading, often using L1. (5:477, observation)	in class (2:1250) Vocabulary—supplied explanations if students asked for clarification (1:77) Wrote word on board, modeled pronunciation, explained or looked word up in dictionary if range of meanings (2:1272, observation) BUT also dealt with vocabulary in prelistening tasks (observed) Vocabulary important in Culture Notes section of iBT3 and iBT4 (2:780)
Feedback to students	When marking writing—paid more attention to organization than to language, but noted common problems in grammar and vocabulary in order to explain to group (8:1293)	When marking writing and speaking—pointed out language mistakes rather than correcting them; asked students questions to get them to think about how to improve; gave grade using scoring rubrics (5:469)	Recommended self-study using iBT0 and *Essential Words for the TOEFL* (2:1280) When marking writing—indicated mistakes in spelling and grammar by using symbols; students were expected to correct their work themselves (2:2004)
Has there been any change since Phase 1?	Content—major change; far less attention paid to grammar and vocabulary; teacher responded to student need rather than preplanning input Methods—Change; more emphasis on working out meaning	Content—major change; far less attention paid to grammar; teacher responded to student need rather than preplanning input; checked meaning of vocabulary after reading and listening activities, however Methods—Change; more emphasis on working out meaning; teacher awarded grade using scoring rubrics	Content—major change; far less attention paid to grammar and vocabulary; teacher responded to student need rather than preplanning input Methods—Change; more emphasis on working out meaning; teacher awarded grade using scoring rubrics

Note. The numbers in parentheses refer to the transcript and line where information can be found. T1, T2, T4 = Teacher 1, Teacher 2, and Teacher 4. Coursebooks are identified in Table 3.

The main change in their approach was that they were now dealing with these aspects of language in a needs-based way as opposed to the preplanned approach that they used in Phase 1. None of the teachers had a fixed list of grammar or vocabulary points that they felt they needed to teach to each group of students. They responded instead to questions the students raised as they were doing other skills work and to what they themselves saw as common problems when they were marking writing (or in the case of T2, writing and speaking). T4 also asked his students at the start of each new course to let him know if they had particular grammar queries. He would then prepare a handout dealing with each query and would discuss the points with the whole class.

When it came to vocabulary, both T1 and T2 mentioned that they encouraged their students to work out the meaning of new words in context. We noticed in T2's class, however, that while she expected students to do this sort of work as they were reading or listening to new material (which was necessary, as they spent most of their class time doing practice tests at the computer), she went over the meaning of new words once they completed these tasks. It was not clear what criteria she had in mind when selecting these words, however. On the day we observed this type of work the semantic field with which she was dealing (names of fish) did not seem important enough to spend any time on, either for the reading text the students had just been through or for the students' general language development. We noted that the teacher had written the L1 translation of some of these terms in her own copy of the coursebook, which suggested that she might have had to look the words up in a dictionary herself before checking whether the students understood them.

T4 did not mention asking students to work out the meaning of words, though this does not mean it was not part of his practice. What he talked about and what we observed was a more directive way of teaching when students asked questions: writing the word on the board, modeling its pronunciation, and when there were questions about multiple meanings, looking the word up in the dictionary (2:1272). It was up to the students themselves though to build up their own stock of vocabulary (2:1266). T4 made suggestions to help them, such as pointing out the section dealing with academic vocabulary in one of the coursebooks they were using (iBT0) and asking them to study these words. If some of the students asked for extra practice he recommended another book, *Essential Words for the TOEFL* (Matthiesen, 1993; T4, 2:1280).

It was not easy for the teachers to estimate how much time they devoted to grammar and vocabulary. We believe that T4 may have been spending more time on the latter than he realized since the prelistening tasks in his coursebooks (iBT3 and iBT4) included familiarization with the vocabulary of the listening passage. This type of work is found in many general English coursebooks but in few TOEFL preparation coursebooks (2:778). In addition there may have been vocabulary teaching related to the Culture Notes sections of the same coursebooks, which present different aspects of student life in the North American academic context (2:778). T4 often asked students to compare the North American situation with what they knew about their own, which required the use of relevant and sometimes new vocabulary (2:780).

All three teachers paid attention to grammar and vocabulary when marking their students' writing. T1 corrected their errors, T2 pointed out mistakes and asked students questions to get them to think about how to improve their language, and T4 used symbols to indicate where there were errors and asked the students to correct themselves.

Summary. There was a major change in the teaching of grammar in that it was dealt with less frequently and less intensively, and teachers mainly responded to their students' queries rather than planning ahead of time what to teach. There was less change in the amount of time devoted to vocabulary but more attention was paid to the idea of guessing the meaning of words and phrases in context.

The Role of Communication

In our report on the Phase 1 study we discussed not only the teaching of the four skills, grammar and vocabulary, but also five themes that had emerged from our analysis of the data. We return to four of these themes in this report: the role of communication, the use of computers, classroom assessment, and teacher training. The fifth theme was the role of the coursebook. We have written about this at length in the Phase 3 section of this report and only reports that the situation described in Phase 3—that of teachers relying strongly on their coursebooks for guidance and material—had not changed in Phase 4. We discuss the role of communication in this section and the other three themes the section to come.

Findings from Phase 1. We saw in Phase 1 (2003) that teachers had learned about the CBT through various sources—amongst them ETS sources (including the ETS Web site, the Bulletin, various books and learning packages such as PowerPrep), other Internet sites, colleagues, and former students. The most important sources, however, were the coursebooks

they used in their preparation classes. The teachers had very little first-hand knowledge of the test (only one teacher had taken it), so it was important that the coursebooks represented the test correctly. Students got information from the Internet (ETS and non-ETS Web sites), education information centers and educational advisory offices such as the Fulbright Commissions, and friends. There was little awareness amongst the teachers that there would be changes in the TOEFL at some point over the next couple of years: Three did not know anything about the changes and the rest knew only that a new test would include a section on speaking. The directors of studies were all aware that a change was in the making, but some of them had found out about this only when we contacted them to ask whether they would be interested in participating in our study. We were interested to see whether this generally low level of awareness would affect the institutions' ability to plan ahead and react appropriately when the new test was introduced in their countries.

Findings from Phase 2. The teachers who participated in Phase 2 were still fairly unaware of what the new test would look like when we collected our first data from them in early 2005. They had learned a little more about its general shape in the 15 months since we had last been in contact with them, but they were not familiar with the details of the test and had not started thinking about how they might change their preparation courses in the future. This lack of awareness was surprising given that the new test was supposed to be being launched in about 9 months' time. Most of the teachers depended on the ETS Web site as their main source of information, but they did not seem to have studied it very carefully. There were, in any case, some gaps in the information provided. The teachers became more familiar with the test requirements as they completed various tasks for our study, but they still had a number of questions even at the end of the data collection period (mid 2005). Their main worries related to the teaching of speaking and to the difficulties they were having obtaining coursebooks to guide them in their planning. Several of the institutions were still trying to find coursebooks as Phase 2 ended. The teachers were feeling less pressured than earlier, however, as they had learned about halfway through the phase that the launch of the test was being delayed until some (unspecified) time in 2006.

What became apparent in Phase 2 was how many sources the teachers used to find out about the new test once they started thinking about it seriously. They used a number of mass media sources (the ETS Web site, other ETS products, non-ETS Web sites, education and

cultural agencies, and commercial coursebooks) as well as a number of interpersonal sources (the school management, the directors of studies, colleagues, students and even ourselves as researchers). The ETS sources seemed to be the most used in this phase; however, this may have been because the teachers could not yet access coursebooks. One of our questions at the end of Phase 2 was whether coursebooks would take over as the most influential source of information after they became more available in the region. Such a development would match findings in other washback studies (e.g., Cheng 1997), which documented the rapid response of publishers to produce test preparation materials when a major examination was changed in Hong Kong. The Phase 3 investigation confirmed the importance of coursebooks in syllabus design and detailed class planning in the countries in our sample.

Findings from Phase 4. Many ETS sources of information were available at the beginning of Phase 4 (April 2007), some of them familiar and some new. The teachers had also inspected a number of coursebooks and all three were using a combination of books that they felt was satisfactory for their purposes. They felt confident that they understood the nature of the new test and that they knew what to do to help their students to prepare for it. Teachers still had some questions about general administrative issues—for example, how long the CBT would continue to be offered—but there were few questions about the test format or the scoring rubrics for the TOEFL iBT exam. The teachers seem to have settled into a routine and they were not as eager to gain new information about the test or about teaching as they had been in earlier phases. We feel that this outcome was due mainly to the confidence that they had in the coursebooks they were using.

We asked the teachers about communication at several points during Phase 4, but our open-ended questions produced responses of different lengths and degrees of completeness. it was thereforet difficult to compare the teachers' experiences and to give a general statement about any particular means of information transmission or teacher support. We therefore decided at the end of the phase to ask the teachers to fill in a table to indicate whether they were or were not aware of specific sources of information and what their reactions were if they had used them.

The results of this survey are presented in Table 16. A list of communication sources is given in the first column of the table, divided into ETS sources for test-takers, ETS sources for institutions and teachers, and non-ETS sources. The teachers' responses are given in the next three columns.

Table 16

Phase 4—Sources of Information

Source of information	Have teachers heard of or used these sources and what comments did they have about them?		
	T1	T2	T4
ETS sources—for test-takers			
TOEFL iBT Overview	Yes. Quite basic. I think I was aware of this information before I saw it on the Web site.	Yes. It gives a general idea of what the new test is like.	Yes It gives the students a general idea of what the new test is like.
TOEFL iBT Tour	Yes, on CD. Good picture of test. I used to use it in Lesson 1. I don't use it now because students can get more useful information from the Longman CD.	Yes. It gives a general idea of what the new test is like.	Yes. I project it on screen in Lesson 1, to give a general idea of what the new test is like.
TOEFL iBT Tips	Yes. Very basic. Good for those who don't have time for anything more.	Yes. Can be useful for those sitting the exam.	Yes. When the TOEFL iBT came out there were not many books available, so the Tips were invaluable. They've lost their importance now, as they've been incorporated by the publishers into their books.
TOEFL Access eNewsletter, including message board	Yes. Not very useful. Mostly contains information about studying abroad. TOEFL information is just news and basic tips. I tell students about it, but I'm not sure it's useful for them.	No—never heard of it. (Had indicated earlier in study however that she had looked at the message board.)	No.

Source of information	Have teachers heard of or used these sources and what comments did they have about them?		
	T1	T2	T4
TOEFL iBT Bulletin	Yes, paper version. Useful information. I used to distribute it to all students. My institution no longer receives the paper version. I don't know why. Web information is always better.	Yes. The administrative information on TOEFL is useful—fees, procedures etc. Also useful are the university codes.	Yes. I distribute it to all our students. Students learn about the administration of the test, and get an application to register by mail.
ETS sources—for institutions and teachers			
TOEFL iBT Frequently Asked Questions	Yes. Good overview for someone who knows nothing about the test.	Yes. Helped me to answer questions students ask.	Yes. We use it in our TOEFL teacher training seminars. The teachers think it's informative.
TOEFL Practice Online Tour (not the same as version for students above)	No.	Yes. It gave me and the students an idea of what the new TOEFL was like.	Yes. We told teachers to do it at home to get familiar with the test. They gave positive feedback.
TOEFL Practice Online Tests	No. (Had been exposed to this in Phase 2, however.) ETS charges for this—as they do for most useful things. Since you have to pay, it's easier and more effective to take a book from the library or bookshops, or to attend a prep course.	No. (Had been exposed to this in Phase 2, however.) ETS charges for this.	Yes. I did it once when TOEFL iBT was first introduced so that I could learn about the test. I've recommended it to my colleagues.
Criterion online writing evaluation	No—never heard of it.	No. ETS charges for this.	No.
Pronunciation in English	No—never heard of it.	No—never heard of it.	No.

Source of information	Have teachers heard of or used these sources and what comments did they have about them?		
	T1	T2	T4
TOEFL Accelerator	Yes. I've seen it on the website but I only have a vague idea of what it is.	No—never heard of it.	No.
TOEFL iBT teacher professional development workshop (face-to-face training, also called Propell Workshop for the TOEFL iBT exam)	Yes—I would like to attend. I have e-mailed for information about a workshop in Istanbul but got no reply.	No—never heard of it.	No.
TOEFL workshop manual (also called Propell Workshop Kit for TOEFL iBT)	Yes—I know it's available. A manual is only useful if there's no way to attend the workshop, which is what I want to do.	No—never heard of it.	Yes. I've used it in my training seminars. It's useful for older teachers especially, those who are not so familiar with the latest technology.
Official Guide to the New TOEFL iBT (ETS, 2006) Copyright—Educational Testing Service, and McGraw-Hill[a]	Yes. Useful if you just want to see what's on the test and don't need much preparation. It's official and unquestionable.	Yes. We use this book to cover the most important things at the beginning of the course (before beginning to practice).	Yes. It's an excellent practice tool which provides students with authentic materials. The material is developed by the test-makers.
NorthStar—Building Skills for the TOEFL iBT *High Intermediate* (Solórzano, 2005) *Advanced* (Fellag, 2006) (Copyright Pearson Education, but cover states "in cooperation with ETS")[b]	Yes. It's good if you have a long detailed course, but other books are better for my needs. I like different levels of difficulty, but the exercises cannot be used independently of the whole unit, and the CD needs to be bought separately.	No—never heard of it.	Yes. An excellent book. We use it with our students to build up skills for the test. Their performance usually improves, measured by the progress tests they take.

Source of information	Have teachers heard of or used these sources and what comments did they have about them?		
	T1	T2	T4
ETS—direct contact (with local representative, over telephone, etc)	No. Other people at the institute are in charge of such communication.	No. I've never had to contact a representative.	No.
Any other ETS source	No.	No. I have not needed anything else.	Yes. I attended one seminar on TOEFL iBT offered by ETS people, a year before the changes. This was the first time I heard about the new format.
Non-ETS sources			
Non-ETS Web sites	Yes. Only when looking for preparation books.	Yes. www.free-english.com The students can find a free test there, so it's useful for them.	No. No need for extra material.
Non-ETS coursebooks	Yes. Longman, Barron's, Princeton Review, Delta	Yes. Kaplan, Cambridge We use the books to practice in class.	Yes. Barron's, Longman. We use the books for further practice in our labs. Students like them and consider them to be at the right level.
Seminars or conferences	Yes. Open Society held a seminar about studying abroad. TOEFL iBT information was included.	No. I've never had the chance to go.	No.

Source of information	Have teachers heard of or used these sources and what comments did they have about them?		
	T1	T2	T4
Director of studies at institution, or other member of management team	No. I'm in charge of the prep course, and they know nothing about it. (Earlier in study had talked about cooperation with a different director of studies.)	No. It's mainly me who deals with TOEFL.	Yes. The director of studies has provided us with all possible training opportunities to improve ourselves as educators.
Fellow teachers	No. I'm the only teacher.	No. I'm the only teacher.	No. (but see explanation below)
Former students	Yes. They give me feedback after the test. Its usefulness to me is variable. Management is interested, for assessing my work and for marketing.	No. When students finish the course it's very difficult to get hold of them.	Yes. We always get feedback from students who have taken the test. They tell us that the reading section is the most difficult.
Current students	No. They don't know anything unless I tell them.	No. They don't have any information.	No.
Any other non-ETS source	No.	No. I don't know of any other sources.	No.

Note. The numbers in parentheses refer to the transcript and line where information can be found. T1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4.

[a] "The TOEFL is created by ETS… ETS also created this book as the official guide to the test" (p. 5). [b] "Pearson Longman and ETS combine their expertise in language learning and test development to create an innovative approach to developing the skills assessed in the new TOEFL Internet-based test (iBT)"—(Solórzano, 2005, p. iv; Fellag, 2006, p. iv).

ETS sources for test-takers. The teachers were aware of all of the sources for test-takers and they had used most of them with their students. They were positive about four of the sources: the TOEFL iBT Overview, the TOEFL iBT Tour, the TOEFL iBT Tips and the TOEFL iBT Bulletin. T1 felt these were quite basic, however. She indicated earlier in the study that she had used the TOEFL iBT Tour with her students on the 1st day of her course, but she now felt that the students could get more useful information from the CD that accompanied her coursebook. T4 still used the TOEFL iBT Tour as a general introduction to the test on the 1st day of hiscourse. He indicated that the TOEFL iBT Tips had lost their importance though, as the information they contained was now available in commercial coursebooks. It was interesting to note how T1 and T4 compared what was available from the official sources with what was available in coursebooks and seemed to perceive the latter information as being more useful.

The TOEFL Access eNewsletter was known by name to T1 only. She was not sure it was useful to students as it focused on giving information about studying abroad rather than about the test itself. She did not comment on the usefulness of the message board for her students, although in an earlier phase she had registered disappointment that that the only message board she had found on the ETS Web site was for students rather than for teachers. T2 said that she had never heard of the Access eNewsletter; however, she too had commented on the message board earlier in the study. Like T2, she had been disappointed that nothing there was helpful to teachers. T4 was not aware of the eNewsletter.

ETS sources for institutions and teachers. There were five different types of information for institutions and teachers. The first type was information that could be obtained from the Web site for free. This included the TOEFL iBT Frequently Asked Questions and the TOEFL Practice Online Tour (different from the tour for students). All three teachers knew about the Frequently Asked Questions and considered them useful, and two of them knew about the Practice Online Tour and considered it useful. T4 mentioned that he had used both these sources in training seminars for teachers. Although he had not mentioned it in earlier stages of the study, he seemed to have become part of a training team preparing other teachers to teach TOEFL.

The second type of information related to ETS products that had to be paid for. The most relevant of these was the TOEFL Practice Online Tests, which T1 and T2 had used during Phase 2 of our study (ETS granted us free access for a period of time so that we could use the writing

test and speaking samples with the teachers), and which T4, who had not participated in Phase 2, had used before the TOEFL iBT exam was introduced in his country. T1 and T2 had not used the Practice Online Tests since Phase 2. Both mentioned that they had to be paid for, and both teachers had indicated in earlier phases of the research that their institutions were not willing to pay for this kind of training. T4 was from a larger and better-resourced institution and he had not only used the package himself but also recommended it to colleagues.

None of the teachers had used the other three products in this category: the Criterion Online Writing Evaluation, Pronunciation in English, and the TOEFL Accelerator. T1 had not heard of the first two products and had only a vague impression of the third. T2 knew that she would have to pay for the first product but had not heard of the other two. T4 was not familiar with any of them. We are not sure why the teachers were not familiar with the products, but we suspect that the fact they had to be paid for might have discouraged them from trying to learn more about them.

The third type of information had to do with teacher training workshops or materials. Only T1 knew about the TOEFL iBT Teacher Professional Development Workshops. Earlier in the study she had expressed interest in attending such a workshop but she knew that her institution would not support her financially. By the end of Phase 4 she had e-mailed for information about a workshop in Istanbul (quite a long way from her own city, but presumably near enough so that she could envisage getting there) but she had not received a reply.

T1 and T4 both knew about the manual that was available to teachers who were not able to attend the workshops. One or two teachers had heard about another manual (*Helping Students Communicate With Confidence* [ETS, 2004]) in an earlier phase of the study, but they knew that it cost $50 and they could not afford to purchase it. T1 did not mention the cost of the TOEFL iBT Teacher Professional Development Workshops manual ($60) though, saying instead that she would rather attend the workshop. T4 had used the manual in his training seminars and felt that it was especially useful for teachers ("older teachers") who were not comfortable with technology and presumably could not get the information off the Internet.

The fourth type of information included coursebooks that had been endorsed by ETS in some way. All three teachers were familiar with the McGraw Hill *Official Guide to the New TOEFL iBT* (ETS, 2006). T4 felt it was an excellent coursebook; T1 and T2 felt it useful for informing students about the test rather than for doing practice activities. Two teachers

mentioned the official status of the coursebook, with T1 in particular commenting on the authority she saw it representing (8:1978). We also saw in the Phase 3 investigation that the McGraw Hill book was highly regarded because of its association with ETS.

The Pearson Education *NorthStar* series (Fellag, 2006; Solózano, 2005) was known to T1 and T4, and the latter used it extensively in his courses. T1's comments matched comments she had made in Phase 3. She liked the fact that the books in the series catered to learners at different levels, but she had mixed ability groups and could not use different books with the same group. She also liked the way the explanations and practice exercises were interwoven but felt it was difficult to use the exercises independently of the rest of the material. It was interesting to discover that T2 did not know that the series existed.

The final ETS source we asked about was direct contact with ETS representatives. All three teachers stated that they had not had contact with a representative. We later wondered whether they had understood the term *representative*, however. T1 indicated in an earlier section of the questionnaire that she had e-mailed for information about a teacher training workshop and had not received a reply. Who would she have sent the e-mail to, if not an ETS representative? T4 also indicated later in the questionnaire that he had attended a seminar "offered by ETS people," which seems to indicate that he must have had some contact with an ETS representative of some kind.

Non-ETS sources. The non-ETS sources included both mass-media channels of communication (other Web sites, other coursebooks, and seminars and conferences) and interpersonal channels (discussions with their directors of studies, fellow teachers, former students, or current students).

Perhaps the most interesting point about mass-media channels of communication was that these teachers were not making much use of non-ETS Web sites in contrast to Phase 1, when they used them regularly. T1 had used them when she was searching for TOEFL iBT coursebooks, but she only consulted the ETS Web site now. T2 mentioned only one Web site, where her students could get a free sample test to try out at home. T4 had previously incorporated material from non-ETS Web sites into his CBT classes, but he felt he had too much material to fit into his classes now.

Also interesting was the fact that one of the teachers had attended a seminar where she had received information about the TOEFL iBT. T4 had earlier indicated that he had attended an

ETS-organized seminar. This situation was in contrast to the one we found in Phases 1 and 2, when teachers lamented not being able to participate in such activities because they could not afford to pay for them themselves and their institutions could not or would not pay for them. Unfortunately, this situation was still the case for T2.

The teachers did not make as much use of interpersonal channels of communication as they had done in Phase 2. T2 did not receive information from any of the sources listed; she only passed on information to her students. T1 wrote that she was the only person in her institution who knew about the TOEFL. She had indicated earlier in the study that she had worked closely with her director of studies, so this later comment seems to have been due to recent personnel changes at her workplace. T1 did make use of feedback from former students who contacted her after the test, and we learned in Phase 4 that it was as a result of feedback about the difficulty of the TOEFL iBT reading section that she had decided to devote more class time to this skill. T4 was complimentary about his director of studies and the support she had given to all the teachers, and he also talked about feedback he got from former students. We knew that T4 was only one of several TOEFL teachers in his institution and that he had the possibility of consulting with others, but he indicated that he did not use fellow teachers as a source of information. It was clear, however, that they would have benefited from his understanding of the test, as he was by the end of Phase 4 involved in some teacher training. Unfortunately, he did not inform us of when he began working as a trainer.

Summary. The TOEFL teachers in Phase 4 were fairly confident about what they should be doing to prepare students for the TOEFL iBT and did not seem to need as much outside information as they needed in Phase 2 and even in the early stages of Phase 3. They had studied the free material that ETS made available to students and other free material for teachers. All three had tried out the Online Practice Tests, though two of them did this as part of our project and might not have tried them if they or their institutions had had to pay for access. The teachers did not know much about the ETS products that had to be paid for. They made less use of non-ETS Web sites than they had done in CBT days. Two of the teachers had been able to attend seminars and had received some information about the test themselves, which was an improvement over earlier phases when several teachers depended on their directors of studies to bring back information from seminars or conferences and to pass it on accurately. The two teachers who worked in smaller institutions did not have the possibility of discussing TOEFL

with fellow teachers, while the third teacher had become a disseminator of information in his institution.

The Use of Computers, Classroom Assessment, and Teacher Training

The analysis of data in Phase 1 revealed many comments about the use of computers in TOEFL classrooms, the use of classroom assessment, and the role of teacher training in the development of teacher abilities. As part of the Phase 4 study we tried to establish whether there had been any changes in any of these areas, and if so, whether these could be linked to the appearance of the new test.

Computer use: Findings from Phase 1. Considerable variation was observed in the way computers were used in CBT teaching. Some institutions had no computers available for use during TOEFL classes, but two had computer labs where some classes were held in order to allow students to work on practice tests that mimicked the CBT. Some institutions had computer facilities that students could use outside class hours, but most students had access to computers elsewhere. The students we interviewed seemed confident in their computer skills, though some students said that they were not able to type quickly. This was not a problem for them as they had the option of handwriting their work for the CBT, but we marked this area as one that might be problematic for students taking the TOEFL iBT, especially if their first language was written in non-Latin script. Not all of the teachers felt confident in their own computer skills, and some would not have wanted to use computers in the classroom even if their institutions had been able to provide them. We wondered whether changes in the TOEFL would influence institutions to provide more computers and to require computer skills in their teachers. We did not see this as a necessary consequence of the new test, however, as the switchover from the PBT to the CBT had not resulted in a need for TOEFL teaching to be via computer.

Computer use: Findings from Phase 4. Table 17 shows what we found regarding the use of computers in Phase 4.

What we saw in Phase 4 was the same sort of variety we saw in Phase 1—one institution that did not use computers for teaching, one that used them for part of the course, and one where most of the classes depended on the use of computers. T1 had no computer facilities available for teaching and had designed her course so that students could benefit from her input and do practice tests in their own time. T4's course began with lots of interaction and introduced computer practice gradually. T2 believed strongly that what students needed most of all was

Table 17

Phase 4—The Use of Computers

Focus	T1	T2	T4
Role of computers	Not used for TOEFL teaching One session only—to project general information about the test in Lesson 1, using a laptop and CD (8:859)	All classes held in computer lab	Some classes held in computer lab to allow students to do practice tests + Teacher projected general information about the test in Lesson 1; not known if he used CD or Internet connection (1:295)
Facilities	No computers in the classroom (observed) Laptop and data projector brought in especially for Lesson 1 (8:859) Problems with electricity (8:889)	Seven computers available to students—one student per computer (observed) Most classes consisted of computer practice (observed)	Well resourced computer lab (observed) Writing Centre (a special kind of lab) available for teaching. (D3:62 and D3: 224)
Student access outside class	Students had access but did use it (8:902)	Students had access to institution's computers outside class time (4:234, 5:496 and 795)	Students had access to institution's computers outside class time (1:250, 2:2208)
Students' computer competence	Students told teacher they were competent (8:2013) Teacher expected them to be competent (8:2009)	Students observed to be competent at typing in Latin script Students could use Internet to find extra materials (4:234, 5:497) Teacher expected them to be competent (4:244) DOS expected them to be competent (D6:255)	Students observed to be competent at typing in Latin script Students could use Internet to find extra materials (2:1350)

Focus	T1	T2	T4
Teacher's computer competence	Competent, though not very confident Felt her computer ability did not affect her ability to teach TOEFL (7:277) Preferred searching bookshops to searching Internet (7:1979) DOS felts teacher had run a good course without computers (D9:480 and 706)	Competent and confident Felt her computer ability affected her ability to teach TOEFL (4:479) No training or support available in institution (4:1328)	Competent and confident Felt his computer ability affected his ability to teach in general, not only his ability to teach TOEFL (1:303) All teachers computer competent (D3:215) Training available in institution (1:279) Technical back-up available (3:674)
Change from Phase 1?	Basic provision the same Institution had acquired computers to set up as TOEFL center, not to use them in TOEFL teaching.	Basic provision the same Computers recently upgraded, but this was due to happen anyway and TOEFL iBT just speeded up the process	Basic provision the same Writing Centre was now available, but not as a result of changes in TOEFL

Note. The numbers in parentheses refer to the transcript and line where information can be found. D3 and D6 = Director of Studies 3, and Director of Studies 6.T1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4.

computer practice, and the majority of her sessions consisted of students working individually at their consoles practicing all four skills under test conditions.

The students in all three institutions were computer competent, or reported to their teacher (T1, who could not observe them at a computer) that they were. T2's and T4's students had had to learn to type in Latin script but they seemed to have few problems with this and were able to type quickly. T2 said that her students had learned to type in Latin script in school, by typing their own language using Latin keys. T4 said that if students had a problem with typing he would refer them to typing software available in his institution. The teachers expected their students to have good computer skills: T1 felt this skill was a given for modern educated people, and T2 said this was natural in a "generation who grew up with computers and the Internet" (5:245).

All three teachers were computer competent as well, though T1 was not as confident as the other two teachers. She was not a technophobe, however, and stated that she could not afford to be one as no one could "run away from computers today" (8:2035). She did not feel that her ability or her attitude toward computers (she did not like using them in her private life) affected her ability to teach TOEFL. T2 was comfortable teaching a course that was almost totally computer-dependent, though she admitted that she had had to teach herself and learn from her mistakes (4:1338). She had no technical support when she was teaching, so she needed to be able to perform a range of tasks and to help her students when their machines malfunctioned. T4 was also comfortable with the computer work he supervised, stating that he did not need advanced skills to do the teaching: "Teaching the [TOEFL] iBT has nothing to do with, let's say, being able to create computer programs or things like that, or fixing computers…That's a wrong impression not only my colleagues, but teachers in general, have. I'm trying to kind of demystify this myth" (2:991).

However, he did say that he felt "computer-literate teachers can understand the mechanics of the test faster than teachers who are not" (T4, 2:2256). He had also noted that a printed training manual might be useful for some teachers ("older teachers") who were not so comfortable using technology (see Table 16, entry under TOEFL Workshop Manual).

Although there had been some changes in computer provision in the three institutions, none of these were because of changes in the TOEFL. The fact that students were computer competent seemed to be a function of the times rather than a result of their having to work toward the iBT. The teachers had developed their own competence for their own reasons and not in order to teach the TOEFL. The courses they ran resembled the courses they were running in Phase 1 in terms of computer usage. It seems reasonable to conclude that the changes in TOEFL did not affect the use of computers in the classroom for these teachers and learners.

Assessment in the classroom: Findings from Phase 1. The teachers in Phase 1 used a fair number of tests in their courses, but the purposes for which they used them were limited. Some teachers used tests for screening purposes to try to make sure that the students who entered their courses had a high enough level of English to benefit from the course. Other teachers gave what they called diagnostic tests early in their courses to get an idea of the types of problems their students had. These tests were found in CBT coursebooks. We doubted that the tests were diagnostic in any way other than the broadest sense: They might be able to reveal a student's

ability in a particular skill area but not be sensitive enough to indicate specific problems. In any case all of the teachers had a syllabus to follow and it was unlikely they would have changed their teaching priorities as a consequence of seeing the results of one of these tests. The teachers often asked the students to take practice tests so that they could familiarize themselves with the demands of the CBT and also gain a sense of their own level of knowledge and skill.

Assessment in the classroom: Findings from Phase 4. The findings from Phase 4 are presented in Table 18. T4 was the only teacher who gave a screening test. The other teachers expected their students to be aware of their own level and the level of the course and register for it if they felt it was appropriate. This practice led to mixed ability groupings. T1 had teaching skills that allowed her to cope with students at different levels, while in T2's course, students worked at computers individually so it did not matter whether they were at the same level or not.

Table 18

Phase 4—Assessment in the Classroom

Focus	T1	T2	T4
Screening purposes	No. Students were told what the level of the course was and that it would be difficult if they were not at that level (7:33)	No. Students were told what the level of the course was and that it would be difficult if they were not at that level (4:28)	Yes. Students needed to be at FCE level to enter TOEFL course (1:37)
Diagnostic purposes	No	No	Yes Test taken from the iBT0 coursebook (2:2124)
Test familiarization and practice	Yes.	Yes.	Yes.
Time spent doing practice tests	20% (7:185)	80% (4:177)	40% (1:202)
Timing of tests	End of course	Middle and end of course	Beginning and end of course
Change in assessment practices?	No	No	No

Note. The numbers in parentheses refer to the transcript and line where information can be found. T1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4.

The amount of time teachers devoted to practice tests varied from 20% for T1 to 80% for T2. T1 used practice tests only at the end of her course, presumably purely for test familiarization. T2 used the tests throughout the course with the aim of giving the students lots of practice in controlled conditions. T4 gave a test at the beginning of the course to give the students a point of comparison for the several tests they would do near the end of the course.

As in Phase 1, we need to comment on the assumptions behind the use of any of these tests. The teachers clearly trusted that the tests were a true reflection of the TOEFL iBT exam and that their students were practicing the right skills, at the right level, in the right way. T4, in fact, stated that he "trusted each and every book on the market" (2:900). The iBT3, iBT4, and iBT5 coursebooks stated on their covers that they used authentic test material from ETS, but it is not known whether the other TOEFL iBT courses were accurate in their representation of the TOEFL iBT exam in their practice tests. Neither is it known whether the tests that were used in any of the coursebooks were of appropriate difficulty or were reliable. There appeared to be no change in the way tests were used in the classroom between Phase 1 and Phase 4.

Teacher training: Findings from Phase 1. Not much training was available to help teachers develop their approach to teaching preparing students for the earlier version of the TOEFL. Two of the larger institutions in the sample (including T4's institution) offered teachers the opportunity to study for professional qualifications, and some of the other institutions offered in-house training of a general kind, but most teachers who wanted to teach TOEFL had to figure out how to do so on their own. T1 was fortunate in that she was able to start her TOEFL career by observing the classes of a more experienced teacher. T2, however, had to create her course from new, with no help from others. The training that teachers could access was usually oriented toward general language teaching, which they did not see as relevant to TOEFL preparation. Even when there was more specific training available (via ETS workshops, for example), some institutions were not willing to invest in training for their staff. The TOEFL courses did not bring in a great deal of income so the returns on the investment would not, in their eyes, justify the outlay.

Teacher training: Findings from Phase 4. The findings from Phase 4 are presented in Table 19. The situation in Phase 4 was not very different from the situation we saw in Phase 1. Two of the teachers, T1 and T2, had gone without training in the intervening years and had had to create their new preparation courses on their own. T1 wanted to attend a TOEFL workshop and

Table 19

Phase 4—Teacher Training

Focus	T1	T2	T3
Size of institution	Small, adequately but not well resourced	Small, adequately but not well resourced	Large and well-resourced
Support available for outside training	Funding very unlikely Teacher would have to train in own time (8:685)	No funding available	Support available (funding and time) to undertake work toward qualifications, including PhD (D3:685)
In-house training	No training available in the country (8:386)	Some training available but not for teaching test preparation classes Training not compulsory Teachers were assessed each year to keep job, so most wanted to attend training (D6:176)	Available and encouraged Training in methodology and technology (1:2679, 2:2233) T4 had made a presentation re TOEFL iBT to colleagues (2:873)
Conferences	Open Society seminar	None mentioned	T4 has made a presentation at a conference (1:277)
Other	T1 had received no training since Phase 1 (7:255) She had never received any TOEFL training T1 was the only TOEFL teacher in her school, so she had no colleagues to share ideas with (8:1013) Had enquired about training, but school not eager to fund (8:700) and ETS had not replied	T2 had received no training since Phase 1 (4:250) She had never received any TOEFL training (5:839) T2 was the only TOEFL teacher in her school, so she had no colleagues to share ideas with (6:151) Designed her TOEFL iBT course on her own. (5:844)	T4 received TOEFL-specific training, and is now giving this kind of training to others (2:872) Was part of team of four, so could discuss teaching with other teachers (2:828)

Note. The numbers in parentheses refer to the transcript and line where information can be found. D3 and D6 = Director of Studies 3 and Director of Studies 6. T1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4.

had written away for information, but she had not yet received a response from ETS. She was realistic about how likely it was that she would get funding from her employer:

> I would really have to provide my boss with some very very very specific reasons—really make a presentation for him like this is going to be good for us in different ways. I guess in that way he might cover the costs. I'm not really sure though as… TOEFL is just one little thing that we do and he doesn't actually know anything about it. That's not his domain. (8:695)

T2 did not give the impression that she was interested in further training, even if her institution were able to support it. Although she had had a difficult time putting her new preparation course together, she seemed satisfied that it was providing the sort of practice that students needed.

T4 was fortunate in that his institution provided not only general training opportunities but also TOEFL-specific training. He mentioned a presentation that he had given to his colleagues about the TOEFL iBT exam, and other comments he made suggested that he was now a trainer himself. He had developed his approach to teaching TOEFL using the input that was available to him through his institution's training program and in collaboration with fellow teachers.

Summary. Although the changes in the TOEFL meant that teachers and institutions needed to consider new course designs and ways of teaching (to accommodate the teaching of speaking at the minimum), not all the institutions were able or willing to provide the support teachers needed as they were making these changes. One teacher was able to benefit from the resources available in his institution and the ethos for developing teachers' understanding and capabilities, but the other two teachers had to find their own way, in their own time, to develop and maintain their new courses.

Discussion and Implications

What the sections above have indicated is that there were indeed changes in the classroom practices of the three teachers with whom we worked from Phase 1 of our project (2003 for T1 and T2, and 2004 for T4) to Phase 4 (2007). These changes are summarized in Table 20.

Table 20

Phase 4—Presence or Absence of Change in the Teaching of Reading, Listening, Writing, Speaking, and Grammar and Vocabulary

Focus	T1	T2	T4
		Teaching of reading	
Content	Change in content—governed by content in TOEFL iBT coursebook, which resembles TOEFL iBT	Change in content—governed by content in TOEFL iBT coursebook, which resembles TOEFL iBT	Change in content—governed by content in TOEFL iBT coursebook, which resembles TOEFL iBT
Method	No great change in method, though more teacher-student interaction	No change in method—Mostly students working on practice tests at computer	Major change in method—more communicative, student-to-student interaction, aided by choice of coursebook
		Teaching of listening	
Content	Change in content—governed by content in TOEFL iBT coursebook, which resembles TOEFL iBT	Change in content—governed by content in TOEFL iBT coursebook, which resembles TOEFL iBT	Change in content—governed by content in TOEFL iBT coursebook, which resembles TOEFL iBT
Method	No great change in method, though more teacher-student interaction	No change in method—mostly students working on practice tests at computer	Major change in method—more communicative, student-to-student interaction, aided by choice of coursebook
		Teaching of writing	
Content	Change in content—governed by content in TOEFL iBT coursebook, including work on integrated writing	Change in content—governed by content in TOEFL iBT coursebook, including work on integrated writing	Change in content—governed by content in TOEFL iBT coursebook, including work on integrated writing
Method	Change in method—input sessions have more teacher-student interaction	No change in method—mostly students working on practice tests at computer	Major change in method—more communicative, student-to-student interaction, aided by choice of coursebook

Focus	T1	T2	T4
Use of scoring rubrics	Change—teacher more aware of criteria, helps her students to understand them, though does not give grade to students	Change—teacher more aware of criteria, helps her students to understand them, uses them while marking and to give grade to students	Change—teacher more aware of criteria, helps his students to understand them, uses them while marking and to give grade to students
		Teaching of speaking	
Content	Complete change—speaking is taught now, not just used as means of communicating 35% of class time devoted to speaking	Complete change—speaking is taught now, not just used as means of communicating 20% of class time devoted to speaking	Complete change—speaking is taught now, not just used as means of communicating 35% of class time devoted to speaking
Method	Students do tasks in front of group and teacher gives immediate feedback to individuals	Students work on practice tests at computer, recording their responses; teacher listens at home and gives written feedback to individuals and common feedback to group	Students do tasks in front of group and teacher gives immediate feedback to individuals
Use of scoring rubrics	Change—teacher aware of criteria, helps her students to understand them, though does not give grade to students	Change—teacher aware of criteria, helps her students to understand them, uses them while marking and to give grade to students	Change—teacher aware of criteria, helps his students to understand them, uses them while marking and to give grade to students
		Teaching of grammar and vocabulary	
Content	Change—Very little grammar or vocabulary teaching takes place	Change—Very little grammar teaching takes place	Change—Very little grammar or vocabulary teaching takes place (2:1250)

Focus	T1	T2	T4
Method	Change—Teacher responds to student queries in class, and corrects grammar and vocabulary when marking writing Teacher encourages students to guess meaning of words in context	Change—Teacher responds to student queries in class, and checks grammar and vocabulary when marking writing Teacher encourages students to guess meaning of words in context	Change—Teacher responds to student queries in class, and checks grammar and vocabulary when marking writing. Indicates where there are errors but students must correct selves Teacher encourages students to guess meaning of words in context

Note. The numbers in parentheses refer to the transcript and line where information can be found. T1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4.

Recall that the authors of the original framework made only general statements about the sorts of impact they envisaged as a result of the introduction of a new TOEFL.

There were just two specific comments:

- that there would be "a move beyond the single independent essay model to a writing model that is more reflective of writing in an academic environment"
- "Students will learn to communicate orally –not to learn a skill simply to do well on a test" (Wall & Horák, 2006, p. 12)

It is clear that there had been a change in the teaching of writing since Phase 1. The most notable change in terms of content was the inclusion of integrated writing tasks, for which students had to process reading and listening inputs before producing output that in some way synthesised the ideas they had been exposed to. Two of the teachers changed their methods as well: T1 now elicited more ideas from her students in the sessions where she prepared them to write, and T4 encouraged more student-to-student interaction than he had done in any of his CBT teaching. All three teachers were also more aware of the writing rubrics for both independent and integrated writing, and made sure their students were aware of and understood them. T2 and T4 used the rubrics when marking their students' writing, and gave grades based on them.

It is harder to comment on the second change envisaged by the framework authors, as they seemed to assume that students were not communicating orally before the introduction of TOEFL iBT. In fact the medium of classroom instruction in all but one of the classes we observed in Phase 1 was English, so students were communicating orally even if their opportunities for communication were limited. What we saw in Phase 4, though, was that the teachers were making efforts to build up their students' confidence so that they would be able to speak for an extended amount of time (up to a minute) per task, expressing their own views and responding spontaneously to written and spoken input. This was not always easy for the teachers to do, and it required a great deal of patience and prodding of students, but we observed in the two courses where students spoke in front of their classmates (T1 and T4) that they were gaining in confidence and competence.

Because the framework statements about change were so general, we also surveyed a number of experts who had served as advisors to TOEFL during early stages of test design, asking them to tell us whether they had discussed the sorts of impact that the new test might have on classroom practices. The types of impact they mentioned were listed in the report on Phase 1 (Wall & Horák, 2006; pp. 15–16). We reproduce them in Table 21, along with an indication of whether we consider these features to have been present in Phase 4. (The experts also mentioned other types of impact—for example, more meaningful results, more differentiation amongst test-takers, and so forth—but we have not listed these in Table 21 as our focus from the beginning of the study has been on classroom impact, or washback.)

Table 21 shows that the features of language and language learning that the experts saw as desirable effects of the new test were, in our view, present in the teaching we saw in Phase 4. Before these can be labelled as *impact*, however, it is necessary to establish an *evidential link* (Messick, 1996) between the introduction of the new test and the features that we found in the classroom. We believe that that link has been established through the detailed work we have carried out with the teachers we have been working with since Phase 1—during Phase 2, when the teachers were becoming familiar with the new test and beginning to plan how they would cope with its new requirements in their future preparation courses; during Phase 3, when they had chosen the coursebooks they felt would serve them best in this endeavor, using their understanding of the test demands as one of their main selection criteria; and during Phase 4, when they told us repeatedly that their choice of content was fully determined by the contents of the test.

Table 21

Impacts Mentioned by Experts in Phase 1 and Whether They Were Present in Phase 4

Possible impact	Present in Phase 4?	Comment
General positive impact		
Changes in test preparation exercises	√	See Tables 10–13
Improved academic language and skills	?	The focus of this project was "processes" rather than "products" (Hughes 1993), so we did not collect test scores that could indicate this
Students rethink what they need to study	√	No student views could be gathered in Phase 4, but the students were following the new coursebooks, which explained and illustrated new skills
Reduction in organization and test-taking techniques as a preparation method	√	Organization no longer necessary and students could take notes while reading and listening The test-taking techniques teachers told us about were sensible strategies rather than tricks
General: Authenticity		
More authentic language input	√	Authentic texts included in coursebooks
More authentic (academically relevant) tasks	√	Integrated tasks included in coursebooks
Integrated skills	√	Integrated tasks included in coursebooks
Reading		
Complex reading texts	√	Longer texts allowed the possibility of more complexity
Study of more complex rhetorical structure	√	Longer texts allowed the possibility of more complexity
Longer texts and making connections between different parts	√	Longer texts allowed the possibility of more complexity

Possible impact	Present in Phase 4?	Comment
	Writing	
Emphasis on summary and paraphrase skills	√	This was taking place via integrated tasks
Working at discourse level rather than dealing with decontextualized grammar and vocabulary	√	Little grammar or vocabulary work done now; discourse level work being done in reading and writing
	Speaking	
Speaking will be taught	√	*Taught* = practiced
More emphasis on productive skills	√	Clear increase in amount of attention given to speaking
Study of pragmatic force of utterance	√/?	Observed in listening exercises, but not observed in speaking exercises or mentioned by teachers

Note. √ = yes, X = no, ? = cannot say.

It is important to stress, however, that the teachers' claims about the contents of the test were based not on their own experience as test-takers but on their understanding of the information and some of the sample material on the ETS Web site and their study of the coursebooks they accepted as representative of the test. We believe it only logical that the ETS Web site would provide an accurate reflection of the test, and we saw in Phase 3 that the coursebooks that the teachers depended on, which included books that had been endorsed by ETS, offered a good representation of the contents of the test. We conclude then that the introduction of the new test was the prime mover in a chain of activities—including dissemination of the contents and format of the test by ETS and the research undertaken by coursebook authors that resulted in teaching material—that led to the content aspect of the type of teaching we learned about and observed in Phase 4. (See Chapman & Snyder, 2000, for a discussion of the notion of linkages.)

What the framework authors and expert advisors did not comment on were the specific teaching methods they thought might or should be used in future TOEFL preparation courses. There were, however, some general statements in the framework documents that referred to a communicative approach to teaching. The authors of the listening framework stated:

> We anticipate that this [test] will encourage language teachers and materials developers to focus more on communicative language use in academic contexts, and that so-called "TOEFL preparation courses" will more closely resemble communicatively oriented academic English courses. (Bejar et al., 2000, p. 36)

The authors of the reading framework wrote:

> Research can be designed to investigate washback effects on what examinees study and to determine whether the emphasis on communicative learning increases once the new test is operational. (Enright et al., 2000, p. 49)

More recently, Wang et al. (2008) stated, "The revision of the TOEFL was motivated in part by language teachers' desires for a test that would reinforce a communicative language curriculum" (p. 298)

We have been guided throughout our study by the belief that terms like *communicatively oriented academic English courses* and *communicative learning* referred not only to what was being taught in the classroom but how it was being taught as well. We did not adhere to a specific definition of *communicativ*e in Phases 1 and 2 because no definition was offered in the framework documents; however, we did ask the teachers in Phase 2 to study a list of task types that we considered to be representative of communication-oriented classrooms (including, for example, information gap, problem-solving, and other cognitively challenging and interaction-based activities) and to say whether they could envisage using any of them in their test preparation classrooms. We needed to be more specific when we analyzed TOEFL iBT coursebooks in Phase 3. We decided to focus on a fairly limited set of features that related mainly to developing the students' strategic competence—prereading and prelistening activities to activate schema, questions that encouraged purposeful reading, questions that allowed students to exercise their creativity (even if only in a limited way) rather than being constrained by multiple-choice and other objective formats, and activities that encouraged the negotiation of meaning through interaction in pairs and groups.

We had seen very little of these sorts of activities in Phase 1. The teaching in almost all of the classes was teacher or coursebook centered, with few instances of students expressing anything but what they felt the correct answers were to the many practice exercises they were asked to complete. There was almost no student-to-student interaction. Most teachers told us in

Phase 1 and in the early part of Phase 2 that they had chosen their approach because it was what their students expected and/or needed in a test preparation course (which is similar to teachers' views recorded in Alderson & Hamp-Lyons, 1996). It was with some surprise then that we found at the end of Phase 2 that some teachers said they would consider more cognitively challenging and interactive activities as part of their TOEFL iBT courses. We were also surprised in Phase 3 to see that the iBT3 and iBT4 coursebooks stood out from the rest of the materials we analyzed by their inclusion of features, which offered room for the exchange of ideas through interaction. Unfortunately, the task we sent to the teachers to ask them about their planning of two classes and how they taught the lessons (Phase 3, Task 2) did not yield responses that were detailed enough for us to see whether the teachers were actually employing any new techniques in their classes. It was not until we were able to visit their institutions again, during Phase 4, that we were able to understand whether or not the approach to teaching that they had shown in Phase 1 had changed in any way.

We found that T2's classes had changed very little, if at all, in terms of methodology. She still practiced an *input and copious practice approach* to teaching, spending several sessions at the beginning of her course going over the requirements of the test in detail and then getting the students to work on practice tests at the computer most of the rest of the time. We saw in the section on the use of computers that she devoted 80% of her class time to computer practice. Her students worked individually, responding to test items and tasks under test conditions. It was only if they had problems that they communicated with the teacher, and they did not communicate with their classmates at all. Apart from 15 minutes or so at the beginning of every lesson, when the teacher lectured to them about the problems she had found while reviewing their homework or asked them to give her the translation of new words in their latest reading exercise, they worked alone. Such practice was not due to T2's inability to teach in another way. She reported that she used other techniques in her regular teaching, but "in general, in the TOEFL classes you can't see a lot of methodology. . . . It's simply a course where we are aiming to prepare the students for the TOEFL and improve their scores and skills with whatever we can" (5:581).

The most apparent change in T1's classes was that she interacted more with her students than she had done in Phase 1, eliciting not only responses to exercises and reasons for choosing certain answers, but also their experiences, their opinions, and in the case of speaking, testlike

performances on which she gave immediate feedback. She still retained control of the class and probably did half or more of the speaking, but the integrated tasks in particular provided opportunities for asking students to explain things like the main ideas of what they had read or the extra information they could take from listening to the oral input. T1 reported that she did not feel the students needed to interact amongst themselves in order to achieve their (and her) goals, and she felt they did not want to. She had tried earlier to get students to do peer assessment of each other's oral performances but they were uncomfortable with this; she also believed they were interested in their own progress only, not that of others. Finally, she felt her 36-hour course was simply did not allow enough time for pair or group work. Nevertheless her classes were more active and stimulating than in Phase 1.

T4's teaching showed the most change from the type of teaching we had witnessed in Phase 1. He used prereading and prelistening exercises to activate background knowledge and vocabulary, encouraged students to work in pairs and groups, encouraged whole-class discussion, discussed cultural points, and so on, and he was pleased to say he "encouraged learning through humanistic methods" (2:1160). He did not see the use of such activities as contradicting the goal of the course: "The whole course is test-oriented, right. But there is room for language teaching. There is room for interaction. For creative production of the language" (2:76).

T4 was happy with the iBT3 and iBT4 coursebooks (2:706; 3:562), feeling that they helped students to prepare for the TOEFL while allowing him to use skills that he had learned when he did his initial teacher training.

It was interesting to see how these three teachers, who were all well trained, experienced, and reflective (this could not be said of all the teachers we worked with in Phase 1 of the study), and who were responding to the same test and dealing with more or less the same content, used three different approaches to conducting their classes—one was characterized by controlled practice on computer, with little room for spontaneity or interaction; one displayed some self-expression and exchange of ideas; and one was similar to the type of teaching that might be seen in any normal (non-test-preparation) classroom. If it wasn't the test that determined how they taught, what was it?

Recall that in previous phases of this research we made several references to the Henrichsen (1989) hybrid model of the diffusion/implementation process. This model served as our basic framework as we tried to determine whether a particular educational innovation (the

TOEFL iBT) would have the consequences intended by its creators (positive impact in the classroom) after it had been introduced into several different user systems (countries in Europe). Space limitations prevent a detailed discussion of all the factors that might have influenced the consequences, but Table 22 gives an indication of some of the main ones we saw, along with an example of the type of influence the factor might have had on the outcome.

Table 22 presents only some of the factors that, in the words of Henrichsen, can "hinder/facilitate the implementation of change" (1989, p. 81). Although the Henrichsen (1989) framework and the ideas of others who have done research into innovation in education (Chapman & Snyder, 2000, and Fullan, 2001, *inter alia*) have been of great value to us as we shaped our investigation, the purpose of this final report is to declare whether the changes in the test itself have had an effect on teaching in the educational establishments we have studied for the last 5 years.

We believe that the new test has indeed had impact on the teaching taking place in the test preparation classrooms studied in Phase 4. The major impact has been in the content of teaching, with considerable change in the areas of writing (the inclusion of multiple inputs to integrated writing tasks, and the raised level of awareness of the writing rubrics), speaking (the focus on developing and practicing speaking, whereas formerly speaking was only used as a language for managing the classroom), and grammar (which occupies a much reduced percentage of class time and is focused on when the students need it rather than as a matter of course). There have also been some changes in teaching methods, though these changes are by no means uniform and seem to have been mediated by teacher characteristics such as beliefs and personal teaching styles as well as by the coursebooks that the teachers or their institutions chose to use as the core of their courses. The new test was received favorably by all three teachers, although it took some time for them to understand the requirements and to decide what approach to use to prepare their students for it. We saw in Phase 2 of the study that their main worries had to do with how they would cope with the teaching of speaking, not so much because of the complexity of the testing tasks themselves, but because they did not at that time have enough models of adequate performance or much material to guide them in how to develop the skills their students might need. Their confidence increased and their questions about how to deal with this skill and others decreased in Phase 3, once they had had an opportunity to inspect and work with their new test preparation coursebooks.

Table 22

Factors Facilitating or Hindering Change

Factor	Example	Outcome
	Characteristics of the test	
The test format	The test contained six speaking tests, which would contribute significantly to the students' overall result.	T1 and T4 devoted a third of their class time to speaking.
	Characteristics of communication	
The way the test was presented in the most used channels of communication	There were not many scored samples of speaking on the ETS website.	T1 was concerned that she might be expecting too high a standard from her students.
The way the test was presented in the coursebook ("form")	Some coursebooks presented TOEFL iBT in the same way they presented CBT, while others were more innovative.	T4's institution chose a coursebook that included many communicative activities.
The teachers' understanding of the nature of the test	The teachers might not have understood the scoring rubrics as well as they thought they did.	T1 and T2 spent many hours (of their own time) marking multiple aspects of their students' writing, without necessarily focusing on what ETS would see as most important.
	Characteristics of the teachers	
The teachers' beliefs about the best way to prepare students for a test	One teacher believed that practice was more important than detailed explanation.	T2 devoted 80% of her class time to computer practice tests.
The teachers' training and preferences for teaching	The teachers had different personal styles.	T1's class was more teacher-centered; T4 was happy that he could now use more communicative techniques.
The teachers' language ability	The teachers were very proficient in English.	T1 could give detailed on-the-spot feedback to her students' oral performances without having taken notes.

Factor	Example	Outcome
The teachers' finances	At least one teacher was paid by the hour.	T1 could not afford to go to conferences, not only because they were expensive but because she would miss classes. Her exposure to new input was therefore limited.
	Characteristics of the institution	
Management priorities	Desire that investment result in returns	T1 reported that her director would need to be convinced that paying for TOEFL training would result in a gain in income. Her exposure to new input was therefore limited.
The ethos of the institution	Possibilities for and encouragement of collaboration	T4 felt part of a team. He had worked with several other teachers to decide on shape of the TOEFL course, the materials, etc, and was involved in training other teachers. T1 and T2 worked on their own and therefore had to rely on their own ideas.
The resourcing of the institution	Provision of computers	T1 would not have been able to do computer practice in the classroom even if she had wanted to, as there were no computers available for teaching.
Classroom considerations	Class size	T1 and T2 were only able to cope with the amount of feedback they gave students because their student numbers were small.
	Characteristics of the students	
The students' occupations	Full-time students or workers, and study time limited	T1's course was very short (36 hours) and she felt she could not spend time on student interaction.

Factor	Example	Outcome
The students' finances	Cost of courses and books	T4's institution chose its coursebook not only on the basis of its approach but because it was half the cost of the other book they were considering.

Note. T1, T2, and T4 = Teacher 1, Teacher 2, and Teacher 4.

The findings regarding the effect of the new test on the content of teaching generally correspond to findings in other studies in the literature; however, this study is unique because of the context in which washback is being measured. A number of studies analyze how tests affect teaching in the state-supported education sector, comparing, for example, the ordinary teaching that takes place in the lower years of a curriculum with the more focused teaching that takes place later, as the time approaches when the students have to take a high-stakes test for matriculation or university entrance purposes, or comparing the teaching that takes place early in the last year before the high-stakes test, with teaching taking place in the last months before its administration (e.g. Lam, 1994; Wall & Alderson, 1993)It is common in such contexts to see what Madaus (1988) would call a "narrowing of the curriculum," which "concentrates attention on those skills most amenable to testing, constrains the creativity and spontaneity of teachers and students, and finally demeans the professional judgment of teachers" (p. 85). Other studies compare what happens in test preparation classes with what happens in general English classes (Alderson & Hamp-Lyons, 1996; Hayes & Read, 2004), where the focus on test-related activities in the former could be seen as being less enrichening for the students involved. What makes the present study different is that the issue of narrowing the curriculum does not apply. There is no logical reason to be dismayed about a focus on testlike practice since this is expected and indeed required by the students, who are customers who will take their business elsewhere if do not feel satisfied with the content they are given. The challenge for the teachers in this study was coping with a test that had expanded in its demands (more and different work on the skills that were already tested, especially writing, and the addition of a completely new skill), when the students who came to them would not be interested in courses that also expanded, with the time and financial implications that such an expansion would involve. Paying substantial attention to the

newest element of the test (an increase in attention to speaking from 0 to 35% for two of the teachers, and from 5 to 20% in the case of the other) would seem to make sense, even though this might mean devoting less time to other equally important skills (and next to no time to grammar and vocabulary). The fact that speaking was receiving such increased attention could only be seen as "beneficial washback" (Bailey, 1996) in the eyes of the test designers.

The findings regarding teaching methods indicate more change than is evident in some other studies. This is certainly the case with the Wall & Alderson study (1993, expanded in Wall 2005) carried out in Sri Lanka, where many teachers did not understand the concepts underlying the new examination (for example, the idea of selective reading) or the curriculum it supposedly represented, and lacked the technical expertise to help their students to develop these skills. Similar results emerge from studies in other developing countries (e.g., Eisemon, 1990), where teachers have not received the necessary training to understand the changes that are desired, have not received sufficient support, and have lacked materials and time to figure out how to teach toward new tests in a productive way. This is not only a problem in underresourced settings, however. Cheng (1997), reporting on the introduction of a new test in Hong Kong, also reported changes in content but lack of change in methods. Although the teachers increased their attention to role plays (a desired change), they actually dealt with it through drilling (following their former teaching patterns). Shohamy et al. (1996) described a contrasting situation, though, where teachers reported using a variety of teaching activities such as brainstorming, jigsaw work, debates, discussions, and speeches to develop their students' abilities to respond to tests containing other types of speaking tasks. This study presents one of the most optimistic accounts of how tests can affect teaching methods in a positive way.

It is becoming more common, however, to find differing amounts and types of change in teaching methods, depending on teacher factors such as beliefs, knowledge, or perceptions of what will be acceptable or rewarded in a given context (Beretta, 1990; Burrows, 2004; Huang, 2009; Watanabe, 1996, 2004), and other factors relating to the test itself; the messages being communicated about the test and the channels these are communicated through; and the educational setting (Fullan 1991 and 2001; Henrichsen, 1989). The findings from this study fit within this set of studies, as can be inferred from Table 22 above.

Finally, it is important to stress the power of the coursebooks in mediating teaching behavior, which is also a common finding in studies of test washback and impact (Andrews et

al., 2002; Cheng, 1997 and 1998; Read & Hayes, 2003; Wall & Alderson, 1993; *inter alia*). One of the strongest images in this regard is of the swiftness with which Hong Kong publishers provided materials for teachers who had to prepare their students for a new high-stakes school examination (Cheng, 1997). Phase 2 of the present study revealed how uncertain teachers were of what the TOEFL iBT would require of them and their students until the time that international coursebooks began appearing in their settings. Their appearance was quite late in most cases, although it was not as problematic as it might have been had TOEFL management not announced a phased roll-out that gave the teachers more time to find resources. The Phase 3 teachers were more confident about their plans for their new courses, and by Phase 4 they were, with the help of their coursebooks, no longer asking questions about what was required by the TOEFL iBT. Spratt (2005) questioned whether the appearance and heavy reliance on coursebooks was a "fruit of uncertainty" (p. 11). [[in times of change and whether teachers would begin to produce their own materials once they got used to changes in their educational systems. It remains to be seen what will happen with the teachers who participated in Phase 4, but our impression at the time of the investigation was that they would not have the desire, the need, or the time to stop depending on published materials in the future.

Implications. Two of the main implications of this study relate to the type of communication that is needed between test designers and the teachers and students preparing for high-stakes tests, and the communication that is desirable between testing agencies and the publishers and authors who design preparation coursebooks.

If, as this study suggests, the main means the testing agency has for communicating its messages to teachers and students is its Web site, then it is important for these users to be able to find information about the test quickly, efficiently, and free of charge. One of the challenges faced by the teachers, and by the researchers, in the early phases of the Impact Study was accessing clear information about the structure of the new TOEFL and how it would differ from the PBT and CBT versions of the test. We published a table comparing the PBT, CBT, and the new TOEFL in our report on Phase 1 (submitted in mid-2004, published as Appendix F in Wall & Horák, 2006), but in order to do so we needed to piece together information from a variety of sources, including the *LanguEdge* practice materials (ETS, 2002) and conference presentations and personal communications with staff at ETS. More information was available on the Web site during the time we gathered our Phase 2 data (2005), but the only way we could provide the

teachers in the project with a variety of sample writing and speaking performances was to ask ETS to give them free access to practice test material that they would otherwise have to pay for. Some of the teachers were still not confident at the end of Phase 2 that they understood the levels represented on the scoring rubrics for writing and speaking, as they had not seen enough scored samples with explanations of why particular scores had been given. It was clear even in the closing days of the project (early 2008) that the teachers who worked in small institutions, who did not have colleagues to exchange ideas with, would have benefitted from the opportunity to participate in online discussions with colleagues teaching TOEFL preparation courses in other places. It is for these reasons that we recommend the following to any agency (organization, institution, ministry, etc.) hoping to create positive washback via the introduction of a new test. We feel that they should, at a minimum:

- Rationalize the number and type of documents that users need to look through to get a good idea of the test design
- Provide free access to sample materials and to practice materials so that teachers and institutions with limited resources can enjoy the same opportunities to see officially approved materials as better-resources users
- Provide as many samples as possible of written and spoken performances at all levels of ability, again free of charge to all users
- Set up and monitor online discussion lists for teachers to allow them to voice questions they have about the test constructs or design and to exchange ideas about appropriate materials and methods for teaching

The TOEFL Web site has developed considerably since the end of the Impact Study and now contains not only descriptions of the test and practice materials, but also information about useful publications, links to teaching tips on YouTube, recordings of Webinars with suggestions for lesson planning, hints about where to find helpful materials online, and more. There are, of course, practical and economic considerations that will affect how much work agencies can do in each of these areas, but these are beyond the scope of this report.

If, as this study has concluded, the impact of high-stakes tests is mediated by the test preparation coursebooks that teachers select, then it is also important that testing agencies pass on very clear messages to coursebook designers about the type of impact they wish to generate,

both in terms of the knowledge and skills that are to be developed in the classroom and in terms of the processes or activities teachers should use to help their learners become competent and confident. It is also important for the test designers themselves to review all the main coursebooks that are available for preparing students for their tests to see whether the content they present and the teaching activities they include match what the original and current test designers desired. While the work of independent researchers may provide some useful insights, it is sometimes difficult for those who have not been present at the original discussions about test design and impact to be able to retrieve and appreciate the original designers' intentions. If test designers carry out this review themselves, it adds force to judgments about whether coursebooks have represented the test demands and intentions correctly and in full. It is, of course, important for the test designers' intentions and some form of the test specifications to be available to all users, but given the dependence of teachers on coursebooks, it is crucial that these should be accurate in their interpretation of test demands.

The immediate implications of this study have to do with communication between testing agencies and the teachers that prepare students for their tests, and the communication between the agencies and the publishers who have such influence over the teachers. There are other implications, however, having to do with the desire to create positive washback in the first place and the research that is needed to determine whether the attempts to create washback have been successful.

We have written here and elsewhere (Wall & Horák, 2006, 2007) about the work that was necessary in the earliest stages of the Impact Study to identify the sorts of impact/washback desired by the experts behind the design of the new TOEFL. Little was recorded at that time about the type of teaching that would appear if the effort to create positive washback proved successful. The TOEFL 2000 framework documents mentioned impact only in the most general terms. We wrote to a number of advisors to the new test and asked them whether they had been involved in discussions of washback and, if so, what types of washback had been mentioned. Their responses were also very general and therefore not very illuminating. It would have been unreasonable to expect otherwise, given that they were being asked to recall discussions that may have taken place years before we were asking them to state what they remembered. The difficulty we had in recovering intentions led us to the conviction that those who wish to influence teaching by introducing new tests should 1) be clear about whether it is realistic to try

to change current teaching practice (this, of course, implies that an adequate description exists of what that practice is, which implies that baseline studies should be undertaken before test design commences), 2) be specific about the kinds of washback they hope to create, and 3) document their intentions in a form that is easily accessible both by test users and by the researchers who may one day be asked to investigate whether the test has had the washback that was desired.

This leads us to the final implication, which relates to the challenges of carrying out impact studies that can provide insights not only into whether the desired impact has occurred, but also into the processes by which it occurred. The point of investigating processes is to learn how they may be made more efficient in the future, thus leading to fuller and more fruitful outcomes. Achieving these insights requires a long-term investment, however, not just a visit before and after the launch date of the test in question. We were fortunate to have four consecutive grants from ETS, which enabled us to keep our small team together for 5 years, purchase some equipment and access other resources, pay for transportation and subsistence during our Phase 1 and Phase 4 visits, and maintain contact with our participants over the long term (the very long term in the cases of the teachers who stayed on through Phase 4). However, given inflation, changes in exchange rates, institutional overhead demands, and the fact that we were following up participants in so many countries, it still was not possible for us to carry out some of the work we would have like to have done. It was not, for example, possible to visit the teachers during Phases 2 and 3. It was not possible to video record them in Phases 1 and 4 (even assuming they would have allowed this, which is doubtful, at least in Phase 1). We have mentioned practical constraints at several points in this report, which led one of the reviewers of our first draft to question whether ETS had not been generous enough in their funding. We did not mean to imply anything like this. It is important to indicate, however, that undertaking a longitudinal impact study requires substantial investment, and it is important for all testing bodies to factor this investment in to the cost of developing their new means of assessment.

Strengths and Limitations of the Impact Study

The TOEFL Impact Study has provided a unique opportunity to investigate whether the introduction of changes in a high-stakes test will cause meaningful changes in classroom practices. We know of no other study that has followed the same teachers for 5 years, from before the time they learned about the characteristics of the new test to a time when they felt familiar with the test demands and had had the opportunity to try out their teaching ideas with

several different groups of students. We feel that this longitudinal study has provided a contribution to the construction of a validity argument for TOEFL (Chapelle et al., 2008a), by providing evidence of the changes that have come about in at least the content of the teaching in a small sample of TOEFL preparation courses. It has also shown the difficulties that ordinary teachers can face as they try to understand the demands that new tests place upon them, as well as the challenges that testers face as they try to figure out how best to inform and support teachers.

We feel that the research questions and overall design of the Impact Study were appropriate, and although this outcome was not planned ahead of time, that they were made more effective by the decision in 2005 to launch the TOEFL in stages. This decision gave us the opportunity to gather more data in the transition period, when teachers were still finding out about the test and working out how to deal with its new elements in their future classes. We believe that test designers can benefit from seeing the sorts of questions the teachers were asking during the process of learning about the test and from understanding how difficult it was for them and their institutions to come up with plans when there were delays and gaps in the information they received about the new test requirements.

Every phase of the Impact Study presented challenges, however. The main challenge in Phase 1 was trying to determine years after test revision work started whether any explicit statements had been made about desirable classroom impact. It was also difficult to piece together how the new TOEFL would differ from the PBT and the CBT. The information available on the TOEFL Web site took some time to reach its final form, and this delay made it hard to predict the type of impact that might occur (as opposed to what was intended by the designers). We needed this information in order to incorporate it into our instrument design.

A second challenge was building up a sample of institutions to visit and teachers to interview and observe. When we were invited to carry out the research we were asked to focus on countries in Central and Eastern Europe. It is difficult to imagine at present how hard it was to get information about institutions that offered TOEFL preparation courses. We scoured Web sites and lists of contacts, but not much TOEFL preparation was taking place in 2003. It was also hard to get access to institutions once we found out they existed. This difficulty was not because of their geographical location, but because we were asking a great deal of people we did not know personally to visit them for several days, look at how their teaching was organized,

interview the directors of studies, interview and observe the teachers, take copies of the teaching material, interview the students, and correspond with them afterward for clarification. We are not sure any institutions would have cooperated had they known the study would go on for 5 years.

The main challenge in Phase 2 was deciding how to probe teachers' awareness of the new test and their concerns about the future without influencing them through our questioning. There were, to our knowledge, no previous studies focusing on a transition period and therefore no methodological models could follow. We saw in later phases that the questions and tasks we set for the teachers had indeed raised their awareness of the basic shape of the test earlier than might have occurred otherwise. However, we also realized that (a) there was no other way of collecting the data we needed and (b) the fact that the teachers may have learned about the test more quickly than if we had not been present did not invalidate our findings regarding the influence the test would have on their teaching.

The challenge in Phase 3 was how to find out what the teachers' earliest attempts at teaching for the new TOEFL looked like when it was not possible for us to observe them. We found that none of the four teachers we were working with, even the two who in earlier phases of the study had recorded their reflections at length and in detail, were able to provide the depth of detail we thought we needed in order to judge how (as opposed to how much) the coursebooks might have been influencing their teaching. We were therefore eager to observe their classes with our own eyes in Phase 4. The challenges of this final phase were to redesign our instruments and procedures so that we could make the most of the brief time we had to visit the teaching institutions and to try to pull the most important strands of the research together without drowning our readers in too much detail.

Although we believe the long-term nature of the project to have been one of its main strengths, the fact that it spanned 5 years meant that there was, very naturally, some attrition amongst the participants. We started the project with 12 teachers in 7 countries. It was not possible to work with all 12 throughout the 5 years, as some of them relocated to other places, some stopped teaching TOEFL, and some were not able to spare the (considerable) time we asked of them as the study progressed. Of the three participants who stayed on until the end, two were from Central European countries and one was from Western Europe. We did not find any differences between the teachers that could be attributed to their countries' former political

orientations or economic policies, so the fact that not all three countries were from the original region did not disturb us—apart from making it difficult to decide on a title for this final report!

We explained in the previous section that we were unfortunately not able to visit our participants in Phases 2 and 3, and we are well aware that this could be seen as a methodological weakness. A common criticism of studies in which self-report plays an important role is that participants may not report on their activities reliably. We have acknowledged that the depth of description was not always as helpful as we would have desired; however, we would also like to stress that we got to know our participants very well through our communications with them over the years, so we are confident that by checking and cross-checking with them during so many tracking sessions and tasks we did get information we could believe in.

We were disappointed not to be able to pursue our original interest in the views of the students studying for the TOEFL. In Phase 1 we managed to interview a number of students at each research site, record and transcribe the interviews, and add their information and opinions to those of their teachers. The long-distance nature of the research in Phases 2 and 3 made the inclusion of further students difficult, and budget and time limits in Phase 4 meant that it was not possible to talk to students in any depth or to record or transcribe what they told us.

In the end though, we are grateful for the opportunity to carry out this research during such an important time of TOEFL's development, and we hope that the analyses we have presented and our reflections on the results and the processes we engaged in will provide at least a small contribution to ETS's attempts to develop the test in the future.

References

Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing, 13*(3), 280–297.

Alderson, J. D., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115–129.

Allwright, R. L. (2000). *Interaction and negotiation in the language classroom: Their role in learner development* (CRILE Working Paper 50). Lancaster, United Kingdom: Centre for Research in Language Education.

Andrews, S. (1995). Washback or washout? The relationship between examination reform and curriculum innovation. In D. Nunan, R. Berry, & V. Berry (Eds.), *Bringing about change in language education: Proceedings of the International Language in Education Conference 1994* (pp. 67–81). Hong Kong, China: University of Hong Kong.

Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19–36). Mahwah, NJ: Lawrence Erlbaum Associates.

Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback: A case study. *System, 30*, 207–223.

Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing, 13*(3), 241–256.

Bailey, K. (1999). *Washback in language testing* (TOEFL Monograph Series, MS-15). Princeton, NJ: ETS.

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.

Beretta, A. (1990). Implementation of the Bangalore Project. *Applied Linguistics 11*(4), 321–337.

Biber, D., Conrad, S. M., Reppen, R., Byrd, P., Helt, M., Clark, V., . . . Urzua, A. (2004). *Representing language use in the university: Analysis of the TOEFL® 2000 spoken and written academic language corpus* (TOEFL Monograph No. MS-25). Princeton, NJ: ETS.

Bonkowski, F. (1996). *IELTS preparation textbook analysis instrument.* Unpublished manuscript.

Breen, M. P. (1987). Contemporary paradigms in syllabus design, part I. *Language Teaching, 20*(2), 81–91.

Breen, M. P., & Candlin, C. N. (1987). Which materials? A consumer's and designer's guide. In L. E. Sheldon (Ed.), *ELT textbooks and materials: Problems in evaluation and development* (ELT Documents 126; pp. 13–28). London, United Kingdom: Modern English Publications in association with the British Council.

Brown, J. D. (1998). An investigation into approaches to IELTS preparation, with particular focus on the academic writing component of the test. In S. Wood (Ed.), *IELTS Research Reports* (Vol. 1, pp. 20–37). Sydney, Australia: ELICOS/ IELTS.

Buck, G. (2001). *Assessing listening*. Cambridge, United Kingdom: Cambridge University Press.

Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian Adult Migrant English Program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 113–128). Mahwah, NJ: Lawrence Erlbaum Associates.

Butler, F., Eignor, D., Jones, S., McNamara, T., & Suomi, B. (2000). *TOEFL 2000 speaking framework: A working paper* (TOEFL Monograph Series, MS-20). Princeton, NJ: ETS.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics 1*(1), 1–47.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008a). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008b). Test score interpretation and use. In C.A. Chapelle, M. K. Enright., & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 1–26). New York, NY: Routledge.

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. (1997). How does washback influence teaching? Implications for Hong Kong. *Language Education, 11*(1), 38–54.

Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation, 24*(3), 279–301.

Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 147–170). Mahwah, NJ: Lawrence Erlbaum Associates.

Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.): *Vol. 7. Language testing and assessment* (pp. 349–364). New York, NY: Springer Science+Business Media LLC.

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3–18). Mahwah, NJ: Lawrence Erlbaum Associates.

Clarke, D. (1989). Communicative theory and its influence on materials production: State-of-the-art article. *Language Teaching, 22*(2), 73–86.

Cohen, L., & Upton, T. A. (2007). 'I want to go back to the text': Response strategies on the reading subtest of the new TOEFL. *Language Testing, 24*(2), 209–250.

Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper* (TOEFL Monograph Series, MS-18). Princeton, NJ: ETS.

Cunningsworth, A. (1984). *Evaluating and selecting EFL teaching materials*. London, United Kingdom: Heinemann Educational Books.

Dudley-Evans, A., & Bates, M. (1987). The evaluation of an ESP textbook. In L. E. Sheldon (Ed.), *ELT textbooks and materials: problems in evaluation and development* (ELT Documents 126; pp. 13–28). London, United Kingdom: Modern English Publications in association with the British Council.

ETS. (2002). *LanguEdge courseware—Handbook for scoring speaking, and writing*. Princeton, NJ: Author.

ETS. (2004). *Helping your students communicate with confidence*. Princeton, NJ: Author.

ETS. (2005a). *TOEFL iBT at a glance.* Princeton, NJ: Author.

ETS. (2005b). *TOEFL iBT tips—How to prepare for the next generation TOEFL test*. Princeton, NJ: Author.

ETS. (2006). *The official guide to the new TOEFL iBT*. New York, NY: McGraw Hill.

Eisemon, T. O. (1990). Examinations policies to strengthen primary schooling in African countries. *International Journal of Educational Development 10*(1), 69–82.

Ellis, R. (1997). The empirical evaluation of language teaching materials. *English Language Teaching Journal, 51*(1), 36–42.

Enright, M. K. (2004). Research issues in high-stakes communicative language testing: Reflections on TOEFL's new directions. *TESOL Quarterly, 38*(1), 147–151.

Enright, M. K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework: A working paper* (TOEFL Monograph Series, MS-17). Princeton, NJ: ETS.

Fellag, L. R. (2006). *NorthStar: Building skills for the TOEFL iBT—Advanced*. White Plains, NY: Pearson Education.

Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 191–210). Mahwah, NJ: Lawrence Erlbaum Associates.

Fullan, M. (1991). *The new meaning of educational change* (2nd ed.). London, United Kingdom: Cassell PLC.

Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). London, United Kingdom: Cassell.

Garinger, D. (2001). Textbook evaluation. *TEFL Web Journal, 1*(1). Retrieved from .http://www.teflweb-j.org/v1n1/garinger.html.

Gear, J., & Gear, R. (2002). *Cambridge preparation for the TOEFL test* (3rd ed.). Cambridge, United Kingdom: Cambridge University Press.

Green, A. (2003). *Test impact and English for academic purposes: A comparativ*e *study in backwash between IELTS preparation and university pre-sessionalc ourses.* Unpublished doctoral dissertation. Roehampton, United Kingdom: University of Surrey, Roehampton.

Green, A. (2006). Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly, 3*(4), 333–368.

Hamp-Lyons, L. (1998). Ethical test preparation practice: The case of the TOEFL. *TESOL Quarterly, 32*(2), 329–337.

Hamp-Lyons, L. (1999). Comments on Liz Hamp-Lyons' "Ethical test preparation practice: The case of the TOEFL." Polemic gone astray: A corrective to recent criticism of TOEFL preparation. The author responds. *TESOL Quarterly 33*(2), 270–274.

Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge, United Kingdom: Cambridge University Press.

Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 97–112). Mahwah, NJ: Lawrence Erlbaum Associates.

Henrichsen, L. E. (1989). *Diffusion of innovations In English language teaching: The ELEC effort in Japan, 1956–1968*. New York, NY: Greenwood Press.

Hilke, R., & Wadden, P. (1997). The TOEFL and its imitators: Analyzing the TOEFL and evaluating TOEFL-prep texts. *RELC Journal, 28*(1), 28–53.

Huang, L. (2009). *Washback on teacher beliefs and behaviour: Investigating the process from a social psychology perspective.* Unpublished doctoral dissertation. Lancaster, United Kingdom Lancaster University, Lancaster.

Hudon, E., Clayton, I., Weissgerber, K., & Allen, P. (2005). *TOEFL iBT with CD-ROM*. New York, NY: Kaplan Publishing.

Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript.

Hughes, A. (2002). *Testing for language teachers*. Cambridge, United Kingdom: Cambridge University Press.

Hutchinson, T. (1987). What's underneath?: An interactive view of materials evaluation. In L. E. Sheldon (Ed.), *ELT textbooks and materials: Problems in evaluation and development* (ELT Documents 126; pp. 37-44). London, United Kingdom: Modern English Publications in association with the British Council.

Hutchinson, T., & Torres, E. (1994). The textbook as agent of change. *ELT Journal, 48*(4), 315–328.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centered approach*. Cambridge, Englane: Cambridge University Press.

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 speaking framework: A working paper* (TOEFL Monograph Series, MS-16). Princeton, NJ: ETS.

Johnson, K. E., Jordan, S. R., & Poehner, M. E. (2005). The TOEFL trump card: An investigation of test impact in an ESL classroom. *Critical Inquiry in Language Studies 2*(2), 71–94.

Krashen, S. D. (1981). *Principles and practice in second language acquisition.* London, United Kingdom: Prentice-Hall International (UK) Ltd.

Lam, H. P. (1994). Methodology washback—An insider's view. In D. Nunan, R. Berry, & V. Berry (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.

Littlejohn, A. (1992). *Why are ELT materials the way they are?* Unpublished doctoral dissertation.Lancaster, United Kingdom: Lancaster University, Lancaster.

Littlejohn, A. (1998). The analysis of language teaching materials: Inside the Trojan Horse. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 191–213). Cambridge, United Kingdom: Cambridge University Press.

Long, M. H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly, 26*(1), 27–56.

Lumley, T., & Stoneman, B. (2000). Conflicting perspectives on the role of test preparation in relation to learning. *Hong Kong Journal of Applied Linguistics, 5*(1), 50–80.

McNamara, T. (1996). *Measuring second language performance.* Harlow, Essex, UK: Addison Wesley Longman.

McNamara, T. (2001). The challenge of speaking: Research on the testing of speaking for the new TOEFL. *Shiken: JALT Testing & Evaluation SIG Newsletter 5*(1), 2–3.

Madaus, G. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *Critical issues in curriculum: Eighty-seventh yearbook of the National Society for the Study of Education* (pp. 83–121). Chicago, IL: University of Chicago Press.

Madsen, H. (1976). New alternatives in EFL exams or "How to avoid selling English short." *English Language Teaching Journal 30*(2), 135–144.

Mahnke, K. M., & Duffy, C. B. (1996). *Heinemann ELT TOEFL preparation course*. Oxford: Macmillan.

Markee, N. (1997). *Managing curricular innovation*. Cambridge, United Kingdom: Cambridge University Press.

Matthiesen, S. J. (1993). *Essential words for the TOEFL.* Woodbury, NJ: Barron's.

Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless or fraudulent? *Educational Measurement: Issues and Practices, 8*(1), 14–22.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). (pp. 13–103). New York, NY: American Council on Education.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241–256.

Miekely, J. (2005). ESL textbook evaluation checklist. *The Reading Matrix, 5*(2). Retrieved from http://www.readingmatrix.com/reading_projects/miekley/project.pdf

Miller, G. S. (2002). *Cracking the TOEFL*. New York,NY: Princeton Review Publishing.

Pearson, I. (1988). Test as levers for change. In D. Chamberlain & R.J. Baumgardner (Eds.). *ESP in the classroom: Practice and evaluation* (pp. 98–107). London, United Kingdom: British Council and Modern English Publications.

Philips, D. (2001). *Longman complete course for the TOEFL test—Preparation for the computer and paper tests*. White Plains, NY: Longman.

Philips, D. (2006). *Longman preparation course for the TOEFL test: iBT*. White Plains, NY: Pearson Education.

Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappa*, 68, 679–682.

Popham, D. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice, 10*(4), 12–15.

Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 171–190). Mahwah, NJ: Lawrence Erlbaum Associates.

Read, J. (2000). *Assessing vocabulary*. Cambridge, United Kingdom: Cambridge University Press.

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.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching*. Cambridge, United Kingdom: Cambridge University Press.

Roberts, M. (2002). TOEFL preparation: What is our Korean students doing and why? *The Korea TESOL Journal, 5*(1), 81–106.

Rogers, B. (2003). *TOEFL CBT success 2004*. Lawrenceville, NJ: Peterson's.

Rogers, B. (2004). Next generation TOEFL: New test, new prep. *The Language Teacher, 28*(7), 37–39.

Rogers, B. (2007). *The complete guide to the TOEFL test: iBT edition*. Boston. MA: Thomson Heinle.

Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). New York, NY: The Free Press.

Rossner, R. (1988). Materials for communicative language teaching and learning. *Annual Review of Applied Linguistics, 8*, 140–163.

Shanks, J. (2004). *TOEFL CBT exam* (3rd ed.). New York, NY: Kaplan, Inc.

Sharpe, P. J. (2001). *How to prepare for the TOEFL* (10th ed.). Woodbury, NJ: Barron's.

Shohamy, E., Donitsa-Schmidt, A., & Ferman, I. (1996). Test impact revisited: Washback effect over time? *Language Testing, 13*(3), 298–317.

Skierso, A. (1991). Textbook selection and evaluation. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 432-450). Boston, MA: Heinle and Heinle.

Solórzano, H. (2005). *NorthStar: Building skills for the TOEFL iBT—High Intermediate*. White Plains, NY: Pearson Education.

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.

Sullivan, P. N., Brenner, G. A., & Zhong, G. Y. Q. (2003). *Master the TOEFL CBT 2004*. Lawrenceville, NJ: Arco.

Swain, M. (1985). Large-scale communicative language testing: A case study. In Y. Lee, A.C.Y.Y. Fok, R. Lord, & G. Low (Eds.), *Initiatives in communicative language teaching* (pp. 35–46). Oxford, United Kingdom: Pergamon.

Swan, M. (1992). The textbook: Bridge or wall? *Applied Linguistics and Language Teaching, 2*(1), 32–35.

Taylor, C. A., & Angelis, P. (2008). The evolution of the TOEFL. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 27–54). New York, NY: Routledge.

Thornbury, S. (2000). A dogma for EFL. *IATEFL Issues, 153,* 2.

Traynor, R. (1985). The TOEFL: An appraisal. *English Language Teaching Journal, 39*(1), 43–47.

Tsagari, C. (2006). *Investigating the washback effect of a high-stakes EFL exam in the Greek context: Participants' perceptions, materials design and classroom applications*. Unpublished doctoral dissertation. Lancaster, United Kingdom: Lancaster University, Lancaster.

Vernon, P. (1956). *The measurement of abilities*. London, United Kingdom: University of London Press.

Wadden, P., & Hilke, R. (1999). Polemic gone astray: A corrective to recent criticism of TOEFL preparation—Comments on Liz Hamp-Lyons' "Ethical test preparation practice: The case for TOEFL." *TESOL Quarterly, 33*(2), 263–270.

Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from Innovation Theory. *Language Testing 13*(3), 334–354.

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. Lancaster, United Kingdom: Lancaster University, Lancaster.

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System, 28*, 499–509.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge, United Kingdom: Cambridge University Press.

Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan Impact Study. *Language Testing 10*(1), 41–69.

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 Series, MS-34). Princeton, NJ: ETS.

Wall, D., & Horák, T. (2007). Using baseline studies in the investigation of test impact. *Assessment in Education 14*(1), 99–116.

Wall, D., & Horák, T. (2008). *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe—Phase 2, coping with change*. Princeton, NJ: ETS.

Wang, L., Eignor, D., & Enright, M. K. (2008). A final analysis. In C.A. Chapelle, M. K. Enright., & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 259–318). New York, NY: Routledge.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13*(3), 318–333.

Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 129–146). Mahwah, NJ: Lawrence Erlbaum Associates.

Waters, A. (2009). Advances in materials design. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 311–326). Basingstoke, United Kingdom: Blackwell.

Williams, D. (1983). Developing criteria for textbook evaluation. *English Language Teaching Journal, 37*(2), 251–255.

Willis, D. (2008). *Form focus and recycling: Getting grammar.* Retrieved from http://www.teachingenglish.org.uk/think/articles/form-focus-recycling-getting-grammar.

Zacharias, N. T. (2005). Teachers' beliefs about internationally-published materials: A survey of tertiary English teachers in Indonesia. *RELC Journal, 36*(1), 23–37.

Notes

[1] We learned after submitting this report that although ETS originally discussed including different native English varieties (indeed, the 2005 edition of *TOEFL iBT at a Glance* [ETS, 2005a] mentions the inclusion of additional native English accents], it was decided not to include this feature until further research findings were available in the areas of accent, intelligibility, and world Englishes.

[2] Some of the review appeared in Wall & Horák (2006). The review has been expanded and updated for this report.

[3] *i+1* is a term coined by Krashen (1981), which represents the idea that language acquisition is facilitated if learners are exposed to input that is slightly more difficult than the language they can already understand.

List of Appendices

Page

Appendix A

Timeline for the TOEFL Impact Study

Year	Month	TOEFL developments	TOEFL Impact Study activity
2000		Publication of TOEFL 2000 framework documents	
2002	Autumn		TOEFL research subcommittee commissions TOEFL Impact Study
2003	January		Phase 1 begins—The Baseline Study Purpose—To gather data about teaching before teachers become aware of the characteristics of the new TOEFL Analysis of framework documents, to find out what sort of impact the new TOEFL was meant to have
	February to March		Survey of TOEFL advisors, to find out what sort of impact the new TOEFL was meant to have
	September to December		Interviews and observations at teaching institutions in 6 countries in Central and Eastern Europe (10 teachers, 9 directors of studies, 10 students—at 10 institutions)
2004	June		Phase 1 ends
	October		Phase 2 begins—Coping with Change Purpose—to track a subset of Phase 1 teachers, and analyze their attitudes and challenges as they learned about the requirements of the new TOEFL At this time it was assumed that the new TOEFL would be launched in their countries some time in 2005
	November		Interview and observations at teaching institution in a country in Western Europe (this was an extension of the Phase 1 study, requested by the TOEFL research subcommittee) 2 teachers, 1 director of studies, 2 students—at 1 institution

Year	Month	TOEFL developments	TOEFL Impact Study activity
2005	Start of year	ETS announces phased roll-out of iBT	
	January to May		Data gathering with 6 teachers in 5 countries; tasks 1 to 5, monthly multiphase tasks followed by computer-mediated interviews Teachers not sure when TOEFL would be launched; waiting for iBT preparation coursebooks to appear on the market
	September	iBT launched in United States	
	October	iBT launched in Canada, France, Germany and Italy	
2006	March		Phase 2 ends
	April		Phase 3 begins—The Role of the Coursebook Purposes—to analyze the coursebooks teachers were using for CBT and iBT preparation, to see whether the iBT coursebooks represented change in content and teaching methods and to investigate how teachers were using coursebooks in the planning and delivery of their courses Data gathering begins, with 4 teachers in 4 countries Tracking questions—Set 1 sent to teachers iBT coursebooks have appeared on the market, and teachers making choices about which coursebooks to use
	May and June	iBT launched in the countries studied	
	June and July		Teachers complete Task 1 and follow-up computer-mediated interviews
	September		Tracking questions—Set 2 sent to teachers.
	October and November		Teachers complete Task and follow-up computer-mediated interviews

Year	Month	TOEFL developments	TOEFL Impact Study activity
2007	March		Phase 3 ends
	April		Phase 4 begins—Describing Change Purposes • to gather data about teaching in the TOEFL preparation courses of a subset of the original baseline teachers, to describe teaching 1 year after the launch of the iBT in their countries • to compare teaching in these courses with the teaching taking place before the launch of the iBT, and to comment on whether any changes observed could be linked with changes in the TOEFL
	May to July		Interviews and observations in 3 countries (3 teachers, 4 directors of studies, at 3 institutions)
	Autumn		Further data gathering, until early 2008
2008	March		Phase 4 ends

Appendix B

Codes Used in TOEFL Impact Study

Code	Meaning	Phase introduced			
		1	2	3	4
	Antecedents				
	Characteristics of the User System				
EdAd	Education administration (above school level)	x			
Sch	School factors	x			
SchN	School factors (where a teacher has moved school and is now discussing their new employer)		x		
SchRes	School's resources				x
SchT	Technology in school	x			
SchTr	School-based training	x			
Crm	Classroom factors	x			
Cult	Cultural factors	x			
Econ	Economic factors	x			
Geo	Geographical factors	x			
Man	Managers of the school		x		
Pol	Political factors	x			
TLU	Target language usage		x		
TSupp	Teacher support		x		
TT	Teacher training		x		
	Characteristics of the Users				
Dab	DOS's abilities	x			
DAbT	DOS's technical abilities	x			
Mot	S's motivation		x		
SAb	S's abilities	x			
SAbT	S's technological abilities	x			
SBLs /Rd /Sp/ Wr	S's beliefs about listening ing construct etc			x	
SClG	S's goals for class	x			
SClGN	S's goals for iBT classes			x	
SDescr	Student description =what are they like			x	
SEcon	S's economic situation	x			
SInt	S's interests	x			
SLEd	S's level of education	x			
SOOC	S's out-of-class preparation activities	x			
SPL	S's personal life	x			
SPsG	S's personal goals	x			
T Ab	T's abilities	x			
TAbT	T's technological abilities	x			
TACrmT	T's attitude toward classroom teaching	x			
TAEd	T's attitude toward education	x			
TAEng	T's attitude toward English	x			
TAEx	T's attitude toward exams	x			

Code	Meaning	Phase introduced			
		1	2	3	4
TAIds	T's attitude toward new ideas	x			
TALT	T's attitude toward language teaching	x			
TBGr	T's beliefs about construct of grammar				
TBInt	T's beliefs about integrated skills		x		
TBLang	T's beliefs about language in general		x		
TBLs	T's beliefs about construct of listening		x		
TBRd	T's beliefs about construct of reading		x		
TBSp	T's beliefs about construct of speaking		x		
TBVo	T's beliefs about construct of vocabulary		x		
TBWr	T's beliefs about construct of writing		x		
TClG	T's goals for class [cf aim]	x			
TConf	T's confidence		x		
TEcon	T's economic situation	x			
TExper	T's experience to date (e.g., in testing, teacher training)				x
Tint	T's interests	x			
TLEd	T's level of education	x			
TPL	T's personal life	x			
TPsG	T's personal goals	x			
TTExp	T's length of experience teaching TOEFL			x	
	Traditional Pedagogic Practices				
Aim	Course aims	x			
AimN	iBT course aims			x	
Typ	Course type [no longer relevant—our 3 Ts run TOEFL prep only]	x			
	Class Content				
Cass	Content re classroom assessment	x			
CAssN	Content re classroom assessment for iBT				x
Crit	Marking criteria/ scales for iBT		x		
CritSp	Speaking criteria/ marking scales/ rubric		x		
CritWr	Writing criteria/ marking scales/ rubric		x		
CtGr	Content re grammar	x			
CtGrN	Content of TOEFL classes re iBT grammar		x		
CtIntN	Content re integrated skills in iBT classes				x
CtLang	Content re language areas general	x			
CtLs	Content re listening	x			
CtLsN	Content of TOEFL classes re iBT listening		x		
CtMat	Content re materials (see later section for detailed codes)	x			
CtMatN (MatN)	Content re materials for teaching iBT TOEFL		x		
CtRd	Content re reading	x			
CtRdN	Content of TOEFL classes re iBT reading		x		
CtSp	Content re Speaking	x			
CtSpIndN	Content of TOEFL classes re iBT speaking—independent			x	
CtSpIntN	Content of TOEFL classes re iBT speaking—integrated			x	

Code	Meaning	Phase introduced			
		1	2	3	4
CtSpN	Content of TOEFL classes re iBT speaking		x		
Ct	Content—general	x			
CtN	Content—iBT general		x		
CTTT	Content re test taking techniques	x			
CTTTN	Content of TOEFL classes re iBT test taking techniques		x		
CtVo	Content re Vocabulary	x			
CtVoN	Content of TOEFL classes re iBT vocab			x	
CtWr	Content re writing	x			
CtWrIndN	Content of TOEFL classes re iBT writing—independent			x	
CtWrIntN	Content of TOEFL classes re iBT writing—integrated			x	
CtWrN	Content of TOEFL classes re iBT writing—general		x		
CtNon-Lang	Content of TOEFL classes other than language			x	
CtNon-LangN	Content of TOEFL classes re iBT other than language		x		
FBack	Content re feedback		x		
FBackSp	Content re feedback to students on their speaking		x		
FBackWri	Content re feedback to students on their writing		x		
HW	Content re homework			x	
Mark	Marks given eg for essays (NB previous use re Task 1—see below)				x
Mis	Mistakes			x	
Notes	Refs to note-taking (cf skills development) in TOEFL classes		x		
EvalProc	Means by which Ts judge the success of their courses			x	
	Methodology				
MthGr	Methodology re grammar	x			
MthGrN	Methodology in iBT classes re grammar		x	-	
MthInt	Methodology re integrated skills	x		-	
MthIntN	Methodology re iBT integrated skills		x	-	
MthLang	Methodology re language areas general	x		-	
MthLangN	Methodology re language areas general in iBT classes			-	x
MthLs	Methodology re listening	x			
MthLsN	Methodology in iBT classes re listening		x		
MthMan	Methodology re classroom management (includes choice of classroom language)	x			
MthManN	Methodology re classroom management of iBT classes				
MthMat	Methodology re materials	x			
MthMatN	Methodology re iBT materials		x		
MthRd	Methodology re reading	x			
MthRdN	Methodology in iBT classes re reading		x		
MthSp	Methodology re speaking	x		-	
MthSpN	Methodology in iBT classes re speaking		x		
Mth	Methodology—general	x			
MthN	Methodology—iBT general		x		

Code	Meaning	Phase introduced			
		1	2	3	4
MTTT	Methodology re test taking techniques	x			
MTTTN	Methodology in iBT classes re TTT		x		
MthVo	Methodology re vocabulary	x			
MthVoN	Methodology in iBT classes re vocab [new code?]			x	
MthWr	Methodology re writing	x			
MthWrN	Methodology in iBT classes re writing		x		
Timing	Timing of certain section of class—to give sense of importance placed on the different sections			x	
	Materials				
Barron	Barron's (= publisher)			x	
BarronN	Barron's iBT (= publisher)			x	
Building	Building Skills for TOEFL by Longman (= title)				
Camb	Cambridge (Author: Gear & Gear)			x	
Crack	Cracking the TOEFL (= title)			x	
CrackN	Cracking the TOEFL iBT (= title)			x	
Essential	Essential Words for the TOEFL			x	
Flash Gr	TOEFL Grammar Flash (= title)			x	
Flash Rdg	TOEFL Reading Flash (= title)			x	
Heineman	Heinemann (= publisher)			x	
Helping	Helping Your Students to Communicate With Confidence—ETS (= title)			x	
Kaplan	Kaplan (= publisher)			x	
KaplanN	Kaplan iBT (= publisher)			x	
Long	Longman (author: Philips)			x	
LongN	Longman iBT (author: Philips)			x	
McGraw	McGrawHill—ETS "official" textbook (= publisher)			x	
North	Northstar (= title)			x	
Prince	Princeton Review (= publisher)			x	
Rogers	Rogers (publisher: Peterson's)			x	
RogersN	Rogers iBT (publisher: Thomson's)			x	
Sampler	ETS Sampler (CBT)			x	
Sullivan	Sullivan (=Author)			x	
ETSMats	Materials from ETS—no specific titles				x
Tests	TOEFL Tests—practice materials			x	
TGs	Teachers' guides			x	
WebMats	Web-based support materials for New TOEFL		x		
Bk	TOEFL iBT textbook—mention of			x	
BkN	Used for when a specific iBT title is being described/discussed				
BkAtt	attitude/opinion of TOEFL books			x	
BkAttN (BkNAtt)	attitude/opinion of TOEFL iBT books			x	
BkChoice	reasons for selection			x	

Code	Meaning	Phase introduced			
		1	2	3	4
BkChoiceN (BkNChoice)	reasons for selection—iBT books			x	
BkInfluence	Influence of the coursebooks on the teachers			x	
BkOther	Any other TOEFL prep title without a specific code				
BkRej	Reasons for rejection			x	
BkRejN (BkNRej)	Reasons for rejection—iBT books			x	
BkRole	What role does the book play in teaching			x	
BkRoleN (BkNRole)	What role does the iBT book play in teaching			x	
BkUse	How are books actually being used in class			x	
BkUseN (BkNUse)	How are iBT books actually being used in class			x	
Famty	Reason for choice—familiarity with author, publishers etc.			x	
MatProd (Mat Prod)	Material production—things Ts—or colleagues—produce			x	
Fam	Familiarization of Ss with test in general as part of exam prep process		x		
Item	Familiarization with item types to be found on TOEFL		x		
Rol	Re role of teacher		x		
	Process[a]				
	Characteristics of Communication				
Comm	Communication	x			
CommAgcy	Communication via agencies such as Fullbright, British Council, etc.		x		
CommConf	Communication via conferences		x		
CommETS	Communication via ETS Web sites or other materials		x		
CommInt	Communication via Internet sites excluding the ETS Web site		x		
CommMan	Communication via management		x		
CommMats	Communication via non-ETS TOEFL materials—usually coursebooks		x		
CommMouth	Communication via word of mouth (not necessary re iBT)				x
CommRes	Communication about TOEFL via our research project		x		
CommSch	Communication about TOEFL within a school/ institution		x		
CommSs	Communication about TOEFL from students (not to Ss)				x
CommT	Communication to others about TOEFL from our teachers		x		
Delay	Delayed launch of new TOEFL		x		
Misap (was MIS)	Misapprehensions		x		
~~SFdbk~~ -> CommSs	Feedback from Ss to Ts re courses, iBT			x	
TQs	Teacher queries re new TOEFL		x		

Code	Meaning	Phase introduced			
		1	2	3	4
TSpec	Teacher speculation re new TOEFL		x		
	Receiver				
	Awareness/Interest				
DAw	DOS's awareness of TOEFL	x			
DAwN	DOS's awareness of new TOEFL	x			
DIntN	DOS's interests/concerns about new exam				x
Saw	S's awareness of current exam	x			
SAwN	S's awareness of new TOEFL	x			
SIntN	S's interests/ concerns etc about new exam		x		
Taw	T's awareness of current exam	x			
TAwN	T's awareness of new TOEFL	x			
TAwNInt	T's awareness of new TOEFL—Integrated Tasks		x	-	
TAwNLs	T's awareness of new TOEFL—Listening section		x		
TAwNRd	T's awareness of new TOEFL—Reading section		x		
TAwNSp	T's awareness of new TOEFL—Speaking section		x		
TAwNWr	T's awareness of new TOEFL—Writing section		x		
TAwLS /Rd/ Sp/ Wr	T's awareness of PBT/ CBT Listening etc.			x	
TIntN	Teacher's interest in new TOEFL (iBT)		x	-	
	Evaluation				
DACrmT	DOS's attitude toward classroom teaching	x			
DAEx	DOS's attitude toward exams	x			
DAIds	DOS's attitude toward new ideas	x			
DALT	DOS's attitude toward language teaching	x			
DATC	DOS's attitude toward TOEFL classrooms	x			
DAtt	DOS's attitude toward TOEFL	x			
DAttN	DOS's attitude toward new TOEFL	x			
SACrmT	S's attitude toward classroom teaching	x			
SAEd	S's attitude toward education	x			
SAEng	S's attitude toward English	x			
SAEx	S's attitude toward exams	x			
SAIds	S's attitude toward new ideas	x			
SALT	S's attitude toward language teaching	x			
SATC	S's attitude toward TOEFL classrooms	x			
SAtt	S's attitude toward TOEFL	x			
SAttN	S's attitude toward new TOEFL	x			
TATC	T's attitude toward TOEFL classrooms	x			
TATCN	T's attitude toward iBT TOEFL classrooms			x	
TAtt	T's attitude toward TOEFL	x			
TAttN	T's attitude toward new TOEFL	x			
TAttNInt	T's attitude toward new TOEFL—Integrated tasks		x		
TAttNIntNeg	T's negative attitudes toward new TOEFL—Integrated tasks		x		
TAttNIntPos	T's positive attitudes toward new TOEFL—Integrated tasks		x		

Code	Meaning	Phase introduced			
		1	2	3	4
TAttNLs	T's attitude toward new TOEFL—Listening section		x		
TAttNLsNeg	T's negative attitude toward new TOEFL—Listening section		x		
TAttNLsPos	T's positive attitude toward new TOEFL—Listening section		x		
TAttNneg	T's negative attitudes toward new TOEFL		x		
TAttNpos	T's positive attitudes toward new TOEFL		x		
TAttNRd	T's attitude toward new TOEFL—Reading section		x		
TAttNRdNeg	T's negative attitude toward new TOEFL—Reading section		x		
TAttNRdPos	T's positive attitude toward new TOEFL—Reading section		x		
TAttNSp	T's attitude toward new TOEFL—Speaking section		x		
TAttNSpNeg	T's negative attitude toward new TOEFL—Speaking section		x		
TAttNSPos	T's positive attitude toward new TOEFL—Speaking section		x		
TAttNWr	T's attitude toward new TOEFL—Writing section		x		
TAttNWrNeg	T's negative attitude toward new TOEFL—Writing section		x		
TAttNWrPos	T's positive attitude toward new TOEFL—Writing section		x		
TExp	T's expectations (contrast with teacher speculation TSpec)		x		
TRepS	T's representation/reporting of students' views		x		
Trk	Tricks—the perceived methods to gain extra points on TOEFL without requisite language ability		x		
~~SEvN~~	S's evaluation of new TOEFL (use SAttN)		x		
~~TEvN~~	Teacher evaluation of new TOEFL [not used in Phase 3—use TAttN]		x		
Perc	Perceptions of TOEFL (contrast attitudes and awareness)		x		
SReac(t)	Student reaction to news of new TOEFL		x		
Plns	Plans re introduction of new TOEFL courses		x		
Worries	S's, T's, DOS's, institutions' re iBT			x	
	Factors That Facilitate/Hinder[bc]				
	Characteristics of the Innovation				
Comps	Comparisons with other exams		x		
Comx	Complexity	x			
Expl	Explicitness	x			
Flex	Flexibility	x			
Fm	Form	x			
Orig	Originality	x			
Obs	Observability	x			
Pra	Practicality	x			
Prim	Primacy	x			
RelAd	Relative advantage	x			
Stat	Status	x			
Tri	Trialability	x			

Code	Meaning	Phase introduced			
		1	2	3	4
	Characteristics of the Resource System				
Cap	Capacity	x			
Hy	Harmony	x			
Op	Openness	x			
St	Structure	x			
Tech	Technological features (of the testing system)		x		
	Extra Codes				
	Background Data				
Avail	Availability of iBT			x	
	Course Data				
CrseData	Info on how many courses ran (and any other info not covered by 2 codes below)			x	
CrseDate	Course dates		x		
CrseLgth	Course length		x		
	Student Data				
SData	Info about students—numeric [adaption]		x		
	Teacher Data				
TData	Info about Ts since Phase 1		x		
	Task-Specific Codes [d]				
Mark	Mark awarded for the essay in the March task—see new usage		~~x~~		
MarkSp	Mark Ts might give to S's spoken work		~~x~~		
MarkWr	Mark Ts might give to S's written work		~~x~~		
Score	Score given for task (same as Mark?)		~~x~~		
QInfo	Tracker question: Any new sources of info on iBT?		~~x~~		
QInst	Tracker question: Is new TOEFL being discussed in your institution?		~~x~~		
QMonth	Tracker question: Has anything of interest re TOEFL happened this month?		~~x~~		
QNew	Tracker question: Have you learnt anything new since last month/ last chat?		~~x~~		
QSs	Tracker question: Have students asked anything?		~~x~~		
QWorries	Tracker question: Do you have any worries/ concerns?		~~x~~		
Challenges	Challenges faced in preparing iBT courses			x	
AdvNov	Ts' advice to novice TOEFL Ts			x	
AdvWrter	T's advice to textbooks writers			x	
	Lesson Descriptions				
CrseDiv	Course division—proportion of time spent on 4 skills, vocabulary, grammar				x
Crse Design	Course design			x	
Ftr EAP	Features of EAP/advanced general English classes		x	-	

Code	Meaning	Phase introduced			
		1	2	3	4
FtrNon-TOEFL	Features of non-TOEFL exam classes		x		
FtrTOEFL	Features of TOEFL classes		x	-	
LessDescr	Lesson description			x	
LessPlns	Lesson plans (esp. for Task 1 Pt 1)			x	
Metaphor	Metaphor for the TOEFL textbooks			x	
TSE	TSE exam	x			
TWE	TWE exam	x			
Vers	Version of TOEFL taken	x			
ExamsOther	Any other (non-TOEFL) exams				x
ApprChange	Change in approach between teaching PBT/CBT and iBT			x	
Influence	Factors influencing the nature of the TOEFL courses			x	
Impl	Implications		x		
WB	Washback		x		
TstMthEff	Test method effect		x		
Res (Influence of Research—Ph3)	Research—any reference to our or T's own research		x		
Soundbites (SndBite)	Quotable snippets		x		

Note. Codes in brackets are permutations of same concept used in previous phases. DOS = director of studies, S = student, T = teacher.

[a] Source/message/plans & strategies = no codes. [b] Interelemental factors = no codes. [c] The Process codes for the characteristics of the user system are the same as the Antecedent codes for characteristics of the user system. [d] Not loaded onto Atlas for Phase 4 data.

Appendix C

Teacher Interview Schedule (Phase 4)

SECTION 1—Factual Questions

Questions about the observation

1. Was that a typical lesson?
2. Do you feel you reached your objectives for that lesson?
3. If not, why not?
4. To what extent do you think the TOEFL influenced your content in that lesson?
5. To what extent do you think the TOEFL influenced your methodology in that lesson?
6. How?

Questions about the current course (from previous version of observation sheet)

7. What stage of course was this class (beginning/middle/end)?
8. How frequent are the lessons (times per week)?
9. How long is this course (# of hours)?

Questions about the students

10. How many students are registered in this class?
11. How many students in the class will take TOEFL?
12. Are all students planning to take TOEFL or are they just there to get a high-level course?
13. How many students in the class have taken TOEFL before?
14. How long before the students take the exam (days/months/not known)?
15. Which version are they sitting?
16. Why do they take TOEFL and not other similar exams?

SECTION 2—Nature of Changes

Step 1: Show teacher the attached diagram (Topic Sheet). Tell them you will ask two general questions and they can select which topic they would like to start talking about. Get them to deal with as many topics as possible.

General questions:

Has the change from CBT to iBT affected any of the following areas?
Has this been for better or for worse?

Your institution
Staffing
Teacher training
Resources
The content of classes
Methodology
Class size
Communication re iBT

Step 2: Cover these points if the teacher does not mention them:

Your institution

Is your job easier/more difficult as a result of iBT?
Has enrolment changed? More students? Different type of students?
Has the administration related to the test changed?
Has your competition changed? More or fewer rivals?
What is the atmosphere like at the school? Teachers worried? Students worried?
Have any management issues arisen?
Publicity?

Staffing

What criteria are used to select teachers for TOEFL classes?

Are these criteria different to those required for teachers for general English or EAP classes? If so how?

Are computer skills seen as important for the TOEFL course? Why/why not?

Do you think teaching TOEFL has any effect on the teaching of other classes? Why/why not?

Teacher Training

Is any training offered by your institution on teaching general English classes or EAP? What? By whom?

Is any training offered by your institution on teaching TOEFL classes? What? By whom?

Is any training in how to teach computer-based classes offered to teachers? What? By whom?

Do teachers take up the training offered? Why/Why not?

Resources

What resources are available at your institution for students and teachers (e.g., library/ computers)?

Are they heavily used?

What resources do you have that are specific to TOEFL preparation?

Are any computers available for students to use in class? How many?

Are any computers made available for students to use outside class hours?

The content of TOEFL classes

Are TOEFL classes more academic than they used to be?

Are all four skills taught?

Are integrated activities practiced?

Is grammar taught?

Methodology of TOEFL classes

Are TOEFL classes more interactive than they used to be?

Are they more communicative?

What determines this?

Class size

How big are the classes on average? (number of students)

What decides this?

In your opinion, does class size affect the teaching and learning in TOEFL classes?

Check also:

What is the director of studies' relationship to the actual classes

How much control do they have?

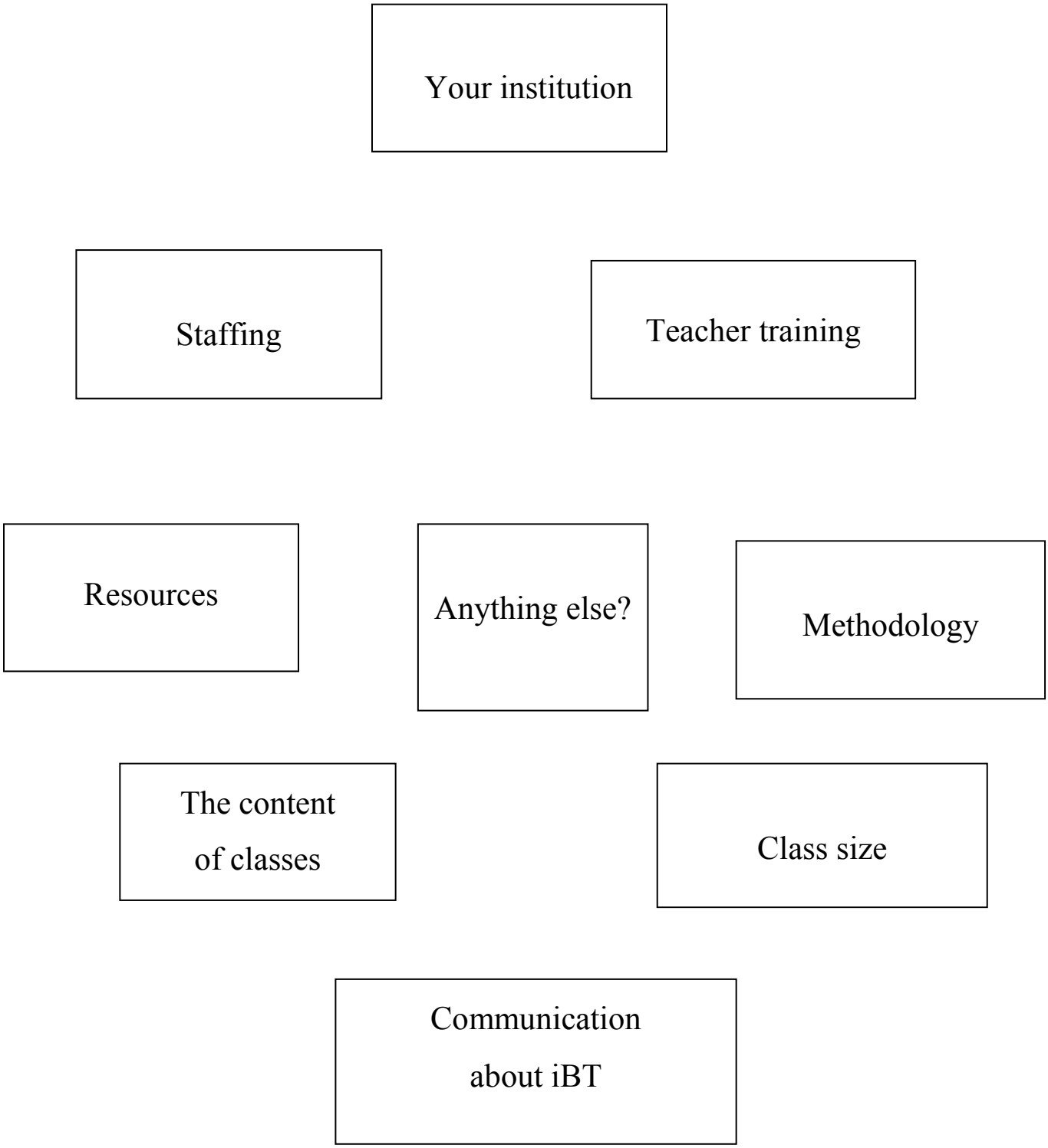
Who makes which decisions?

Re content

Re methodology

Re assessment

SECTION2—Topic Areas



SECTION 3—Follow-up to tracking questions, May 2007

Experience

1. Have you done any work with any exam board or exam bodies since our first contact with you at the beginning of the project? Yes/ No
2. In what role?

What effect has this had?

3. Have you taken the TOEFL exam yourself? Yes/No
4. If so, when?
5. Which version?
6. If so, do you think this experience has influenced how you teach TOEFL preparation classes? Yes/No

How?

TOEFL preparation courses—aims

7. What is the main aim of this class

to work on the right things for passing TOEFL	
to improve the students' general English	
to prepare students for working in an academic environment	
something else—please specify	

8. How do you feel about that?

TOEFL preparation courses—selection

9. Are students **screened** (preselected) in any way before they can join the TOEFL preparation class? Yes/No
10. If so, how?

TOEFL preparation courses—course content

11. Is there a **course outline** / **course description** for the TOEFL courses in this institution?
12. Who produced it?
13. What's it based on?

14. Over the length of a whole course **what percentage** of that time do you spend in class on these **language elements**?

	%
Listening	
Reading	
Writing	
Speaking	
Grammar	
Vocabulary	
other (what?)	

15. What is your **rationale** for this division of time?

Expand

16. What kinds of texts do you usually give students to **read** in the lessons?
17. How long are they on average?
18. What topics do they generally cover?

Why do they use the **reading** texts they use?

Source? What materials do they use?

What is being tested?

What are students' problems with reading—if any?

Teacher's attitude toward how best to practice reading?

Favorite/ any test-taking techniques?

Cf Question 34—Why do they do the activities for developing **reading** they indicated?

19. What kinds of passages do you usually give students to **listen** to in the lessons?
20. How long are they on average?
21. What topics do they generally cover?

Why do they use the **listening** passages they use?

Source? What materials do they use?

What is being tested?

What are students' problems with listening—if any?

Teacher's attitude toward how best to practice listening?

Favorite/ any test-taking techniques?

Demands on memory? Comparison between CBT and iBT.

Authenticity? More authentic than in PBT/CBT?

Cf Question 33—Why do they do the activities for developing **listening** they indicated?

Cf Question 35—Why do they do the activities for developing **writing** they indicated?

What's the source of topics students write essays on? (if they do)

What is being tested?

Ss' ability on arrival?

Issues with typing essays?

Favorite/any test-taking techniques?

Cf Question 36—Why do they do the activities for developing **speaking** they indicated?

What is being tested?

Any test-taking techniques?

Level of Ss' knowledge of **grammar** on arrival?

How do they feel about absence of grammar in iBT?

Consequences?

Role of grammar for TOEFL success?

If they teach **vocabulary—**why?

How?

What?

Role of vocabulary for TOEFL success?

22. Do you do any activities working on **two or more skills at once**, for example, reading a text and then speaking about the content of that text or listening to a passage and then doing writing based on the passage?

Yes/No

Expand

TOEFL preparation courses—methodology

23. Have you ever taught or do you now teach **high-level English** or other **EAP** (English for Academic Purposes) classes?
24. If so, how are your TOEFL classes different or similar to them?

Expand

25. Have you ever taught or do you teach any other **(non-TOEFL) exam preparation** classes?
26. If so, how are your TOEFL classes different or similar to them?

Expand

27. Which **language** (your mother tongue or English) do you use most in class?
28. How do you decide which language to use in class?
29. Which language do your students use most in class?
30. How do you feel about that?
31. Which of these different **working arrangements** do you use in your TOEFL course?

	% of total course
individual work	
pair work	
group work	
whole class	
something else—please specify	

Why?

32. What do you tend to use most in a typical class?
33. What activities do you do in class to develop **listening**?
34. What activities do you do in class to develop **reading**?
35. What activities do you do in class to develop **writing**?
36. What activities do you do in class to develop **speaking**?

NB 33–36 should have been covered in questions above—see Question 18 onwards

TOEFL preparation courses—assessment

37. Do you give your students writing tasks to do?
38. If so, what types of tasks?
39. Do you give the students marks for their writing?
40. What system of marking (grading) do you use?

Does this include feedback? Format?

41. Do you use the iBT **writing "rubric"** (also called "scoring scales" or "rating scales")?
42. Do you use both the independent tasks rubric and the integrated tasks rubric? Yes/No
43. If so, for what?
44. Do you feel comfortable using these rubrics? Yes/No
45. Why/Why not?
46. Do you refer to the rubrics in class? (i.e. are the students familiar with them?) Yes/No

Expand

47. Do you use the iBT **speaking "rubric"** (also called "scoring scales" or "rating scales")?
48. Do you use both the independent tasks rubric and the integrated tasks rubric? Yes/No
49. If so, for what?
50. Do you feel comfortable using these rubrics? Yes/No
51. Why/Why not?
52. Do you refer to the rubrics in class? (i.e., are the students familiar with them?)

Expand

53. Do you give **tests** to check other skills (reading, listening)? Yes/No

Expand Do they do any: Screening? Diagnostic testing? Practice tests? Self-assessment on computers in class? Practice tests taken under test conditions?

TOEFL preparation courses—test-taking techniques

54. Do you cover test-taking techniques in your lessons (e.g., analyzing questions, etc.)?
55. Do you use practice tests in class?
56. What proportion of the whole course is spent on students taking practice tests? ……%
57. At which stage of the course do you use practice tests most in class?

beginning	
middle	
End	
Throughout	

TOEFL preparation courses—teaching materials and resources

58. Which of these **materials** are available at your institution for teaching TOEFL preparation?

	Q.58	Q.59
Materials produced by ETS		
Practice materials downloaded from the TOEFL Web site		
Other (non ETS) commercial publications published locally		
Other (non ETS) commercial publications published abroad		
Unpublished materials produced by your institution		
Materials produced by yourself		
Past exam papers		
Something else		

Expand

59. Which of these materials do you use most in class? (please indicate in the table above)
60. Why?
61. If you produce your own material, what resources do you draw on to help you?
62. Do you use or make reference to the ETS Web sites in class?
63. Do you spend time in class on **computer-based** tasks? Yes/No
64. If not, why not?

If not covered:
What is provision of computers at the school like?

65. If so, how much of the course as a whole is typically spent on computer-based tasks/practice? ……%
66. Do students do computer practice outside of TOEFL classes, as far as you know?

Are Ss used to computers?
Ss' confidence using computers?

Has iBT affected provision/ resourcing?
Has iBT affected teacher training in use of computers?

Students' independent language development strategies

67. Do students do any studying outside class to help prepare for TOEFL, as far as you know?
68. What do they do?
69. Is this prompted by you (e.g. by giving tips or ideas for what to do)? Yes/No

Teacher support

70. Do you refer to official TOEFL materials (e.g. booklets such as TOEFL Tips or the website) to give you guidance on how to teach these courses? Yes/No

Expand

71. Have you recently (since November 2006) had any training to teach high-level English or EAP? Yes/No
72. If so, from where?
73. Have you recently (since November 2006) had any training to teach TOEFL preparation? Yes/No
74. If so, from where?
75. Is any training on how to teach computer-based courses available? Yes/No
76. If so, from where?
77. Do teachers tend to take up training opportunities offered? Yes/No
78. If not, why not?
79. If you have had any training: How far has this training influenced the **content** of your TOEFL preparation course lessons (what you teach)?
80. How far has this training influenced the **methodology** of your TOEFL preparation course lessons (how you teach it, the activities you use, how you manage the class, etc.)?

Expand on training in general

Computer skills

81. How confident do you feel in your computer skills?
82. Do you feel your confidence in using computers has an effect on your ability to teach TOEFL preparation? Yes/No

Expand

TOEFL Awareness

83. What are your **sources of information** about **iBT** TOEFL?
84. Which sources do you find most helpful?
85. Why?

Expand on sources of information. Is the public generally familiar with iBT TOEFL yet?

86. As far as you know, what preparation materials are available from ETS, the producers of the TOEFL exam?
87. In your opinion does **iBT** TOEFL test the following things?

Ability to…	Yes	No	Unsure
use grammar correctly			
use a wide range of vocabulary appropriately			
use idioms correctly			
understand a wide range of texts			
express original ideas in writing			
translate from your native language to English and vice versa			
take an active part in an academic discussion or seminar			
understand lectures			
infer someone's opinion, when it is not stated clearly			
understand the organization of a text			
write formal letters			
analyze information from several texts			
make inferences from information in a text			
give a presentation			
understand unfamiliar vocabulary from context clues			
state your opinion on a given topic and support it			
understand language used in everyday situations and conversations			
speak for an extended period on a familiar topic			
write an academic style article			

Attitudes about iBT TOEFL

88. From your experience, do iBT TOEFL scores reflect students' real language ability?

What experience is that? What is the response based on?

89. What language skills and sub-skills does a candidate need to do well on **iBT** TOEFL, in your opinion?
90. What knowledge or skills <u>apart from language</u> does a candidate need to get good TOEFL scores, in your opinion?
91. What do you think are the good features of the **iBT** TOEFL exam (if any)?
92. What do you think are the bad features of the **iBT** TOEFL exam (if any)?

Expand

93. Which section/aspect is hardest to teach?
94. Which section/aspect is easiest to teach?

Expand

Any views on the length of iBT? (cf stamina issue)

Attitudes about teaching TOEFL

95. Do you personally like teaching TOEFL? Yes/No
96. Why/ Why not?
97. Are you yourself learning anything by teaching TOEFL courses? Yes/No
98. If so, what?

Attitudes about tests

99. Do you agree or disagree with the following statements?

	Agree	Disagree
Tests promote good learning		
Tests encourage students to study		
Tests encourage good teaching		
Students can improve their language skills by doing practice tests		
Tests make students study how to take tests not how to develop your language skills		

100. What has most influenced your teaching of TOEFL?

********** **Ask for a copy of the course outline and any publicity.**

****

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