ETS

TOEFL®

Monograph Series

MS - 34
June 2006

The Impact of Changes in the TOEFL Examination on Teaching and Learning in Central and Eastern Europe: Phase 1, The Baseline Study

Dianne Wall

Tania Horák

The Impact of Changes in the TOEFL Examination on Teaching and Learning in Central and Eastern Europe: Phase 1, The Baseline Study

Dianne Wall and Tania Horák

Lancaster University, UK

Abstract

The purpose of this report is to present the findings of the first phase in a longitudinal study of the impact of changes in the TOEFL® test on teaching and learning in test preparation classrooms. Observations were carried out and interviews conducted with teachers, students, and directors of studies at 10 institutions in six countries in Central and Eastern Europe. The findings constitute a baseline against which finding from future phases can be compared. The report includes an analysis of the types of impact originally envisaged by advisers to the new TOEFL and a discussion of factors in the educational context that may influence the types of impact that may appear once the new test is operational.

Key words: TOEFL, Test of English as a Foreign Language™, impact, washback, backwash, baseline study, innovation, test preparation, English for academic purposes, Central Europe, Eastern Europe

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

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

❖   ❖   ❖

A continuing program of research related to the TOEFL test 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, 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.

Because the studies are specific to the TOEFL test and the testing program, most of the actual research is conducted by ETS staff rather than by outside researchers. Many projects require the cooperation of other institutions, however, particularly those with programs in the teaching of English as a foreign or second language and applied linguistics. Representatives of such programs who are interested in participating in or conducting TOEFL-related research are invited to contact the TOEFL program office. All TOEFL research projects must undergo appropriate ETS review to ascertain that data confidentiality will be protected.

Current (2005-2006) members of the TOEFL Committee of Examiners are:

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

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

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

Table of Contents

List of Tables

List of Figures

Introduction

The purpose of this study is to provide a baseline description of the teaching and learning that currently takes place in commercial TOEFL® preparation classes in Central and Eastern Europe. This is the first phase in a series of studies that have as their aim the measurement of the impact of the new TOEFL test on classroom practices. The TOEFL 2000 framework documents (hereafter the framework documents—Bejar, Douglas, Jamieson, Nissan, & Turner, 2000; Butler, Eignor, Jones, McNamara, & Suomi, 2000; Cumming, Kantor, Powers, Santos, & Taylor, 2000; Enright et al., 2000; Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 2000), which set out the thinking behind the new test, declare an interest not only in incorporating modern views of language and measurement into the examination, but, also, in improving its impact on teaching and learning. The assumption in these documents is that the current test exerts a problematic influence on the classroom (see, for example, references to Hamp-Lyons & Kroll, 1997; Peirce 1992; and Raimes, 1990 in the writing framework [Cumming et al., 2000, p. 28]). The hope is that, with the introduction of new components such as integrated skills tasks and a speaking test, the emphasis in classroom teaching will shift and "courses will more closely resemble communicatively oriented academic English courses" (Bejar et al., 2000, p. 36).

In order to determine whether the desired impact has occurred, it is first necessary to have a clear description of teaching and learning before the introduction of the new TOEFL test. The purpose of a baseline study is to establish what the situation is like before the introduction of an innovation, in order to be able to evaluate whether the desired effects have appeared later. In the words of Weir and Roberts (1994): "We need to establish what conditions are before 'treatment' . . . which will help us to monitor any effects that occur during or after 'treatment'" (p. 46).

Vital to this notion is the need to analyze the situation before the participants are exposed to the change that is intended to influence their practices. The study should be carried out before the participants are even aware that a change is imminent, and certainly before they have understood what the change will involve and what the intended effects are supposed to be. This study was begun and completed before the subjects developed a full awareness of the content and format of the new TOEFL test. We therefore believe that the results will provide a useful point of reference for the future.

Test Impact

It is now accepted that agencies involved in the design of examinations with high-stakes need to be conscious of the impact that their exams have on teaching and learning. This issue has long been discussed in the literature of general education (e.g., Vernon, 1956; Kirkland, 1971; Kellaghan, Madaus, & Airaisian, 1982) and has been looked at from a variety of angles (Popham, 1987; Frederiksen & Collins, 1989; Airaisian, 1988; Madaus, 1988; see Wall, 1999 for a detailed review), but it was not until the 1990s that it attracted attention in language education. Much of what was published earlier lacked an empirical foundation. The few studies based on data relied on self-report techniques or on test results rather than on direct contact with teachers and learners (e.g., Wesdorp, 1983; Hughes, 1988; Khaniya, 1990; Li, 1990). Alderson and Wall (1993) were the first to look critically at the notion of test *washback* (the impact of tests on teaching and learning—also known as *backwash* in general education), arguing that test designers should be more explicit about the type of impact they hoped their tests would produce (the Washback hypotheses), that researchers into impact should use more varied research techniques (including classroom observation), and that more use should be made of research findings in other fields such as educational innovation. Hughes (1994) distinguished between washback on the participants, the processes and the products of an educational system, and Bailey (1996) made a further distinction between washback to the learners and washback to the program. Bailey also discussed the difficulties of designing washback investigations, including the challenge of determining how much of what takes place in a classroom can be "evidentially linked to the introduction and use of the test" (Messick, 1996, p. 24; Alderson & Wall, 1993, p. 117). The importance of studying test impact was reinforced by the inclusion of impact in Bachman and Palmer's notion of *usefulness* (1996) and by the prominence given to the consequential aspect of validity in Messick's (1989)]unified view of validity.

More empirical work has appeared in the last decade, investigating the impact of different kinds of tests in different settings: national school examinations in Sri Lanka (Wall & Alderson, 1993; Wall, 1996), Israel (Shohamy, Donitsa-Schmidt, & Ferman, 1996) and Hong Kong (Cheng, 1997, 1998); university entrance examinations in Japan (Watanabe, 1996) and the TOEFL exam in North America (Alderson & Hamp-Lyons, 1996). A recent volume edited by Cheng and Watanabe (2004) presents a series of studies from various parts of the world into the impact of tests in specific educational contexts. Most of the studies we have referred to focused

mainly on what Hughes (1994) called *processes* (the content and methodology of teaching) and *participants* (the attitudes of teachers and learners). Less attention has been paid to the effect of examinations on *products* (e.g., the quality of students' learning).

Some studies have attempted to identify the factors influencing the form that impact takes. These factors include not only the nature of the examination and related training, but also characteristics of the educational context (e.g., management practices within institutions, classroom conditions, resourcing, feedback mechanisms between the testing agency and schools) and of the teachers and the learners (e.g., teacher understanding of the test and the approach it is based on, teacher ability, willingness to innovate). Wall (1999) applied insights from innovation theory (i.e., Henrichsen, 1989; Fullan, 1991; Rogers, 1995) in an attempt to help test designers to gauge the amount of risk involved in trying to innovate through testing. Wall's framework took as its starting point Henrichsen's (1989) hybrid model of the diffusion/implementation process, discussed further in this report. Among Wall's conclusions was that impact should be monitored from the earliest stages of a test reform project and that the results should be disseminated widely "so that decisions can be made about the match between the test and the curriculum, the appropriateness of the design and procedures, the adequacy of communication, the allocation of resources, the planning of further training, and changes in overall and specific policy" (Wall, 2000, p. 507). Similar conclusions were put forward in general education by Chapman and Snyder (2000), who blamed the failure of many attempts to use tests with high-stakes decisions to improve classroom instruction and student achievement on a lack of understanding of "the intermediate conditions that had to be met for changes in test content, format, or use to have the desired impact on teachers' classroom practice" (p. 457). In their model of the links between testing and classroom practice they specified that the most difficult condition to meet was getting teachers to understand what changes were needed to raise student performance.

Impact of the New TOEFL Tests

It is evident in the series of framework documents that the impact of the new test is of prime concern to its designers. Each of the monographs declares that the new examination should have more positive impact on the classroom. Some mention is made of desirable outcomes (e.g., the use of less coachable question types, including tasks that require the integration of two-or-more skills) (Butler et al., 2000), but there is no composite statement of what impact is desired of the examination as a whole. Few suggestions are given about how

impact might be investigated or how research results might be fed back into the design and dissemination processes to improve classroom teaching and learning. The commissioning of a TOEFL Impact Study provided an opportunity to apply what has been learned about impact in language education and other fields and to investigate the consequences of the new TOEFL test as early as possible. This was done in the hope of promoting beneficial washback whenever possible and identifying and eliminating unintended negative effects before they become ingrained in the TOEFL preparation system.

What we, the designers of the impact study, proposed was a longitudinal investigation that would be divided into several phases. The study would not only provide an accurate description of the teaching and learning taking place at different times, but, also, allow the results to be used in a *formative* way. Phase 1 (the focus of this report), a baseline study, would describe the situation existing before the test became operational and, ideally, before detailed information was released that might tempt teachers to alter their approach to teaching in anticipation of changes in the test. Phase 2, a transition study, would investigate how teachers, students, and teaching institutions found out about the new test, what they understood of its construct and form, and how they prepared to cope with its new demands. Phase 3 would provide detailed descriptions of teaching and learning after the introduction of the new test and provide explanations for why change might or might not have occurred in the way that was projected. The intention throughout the study would be to feed results back into the test design and dissemination process. The new TOEFL test could in this way benefit from contact with its users and respond to their needs.

Theoretical Framework

The theoretical framework for the impact study is Henrichsen's hybrid model of the diffusion/implementation process (Henrichsen, 1989; see Figure 1). The new TOEFL test is meant to be an innovation in the language teaching and testing community, and the Henrichsen model illustrates and explains the process by which innovations are either accepted or rejected by their intended receivers.

Henrichsen divides the process of innovation into three stages: antecedents, process, and consequences. Henrichsen's main message is that the consequences (impact) of an innovation are determined not only by its own characteristics (in this case, the characteristics of the new TOEFL) but by the interaction of features in the antecedent situation (the context into which the

innovation is being introduced) and a number of factors that work together (or against one another) during the process period (the time that the innovation is introduced and being tried out by the users).
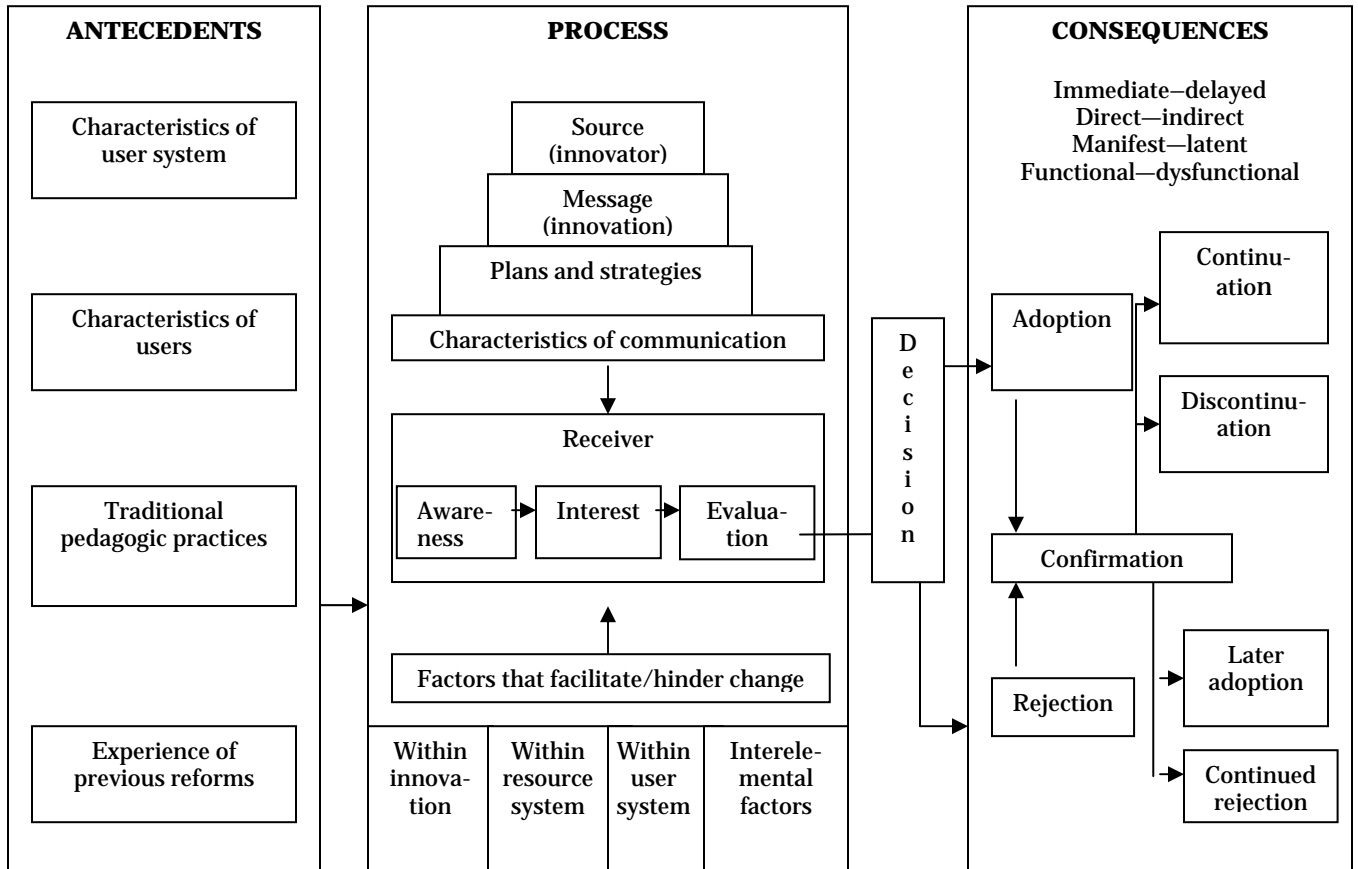
Figure 1. **Henrichsen hybrid model of the diffusion/implementation process.**

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

The purpose of Phase 1 of the TOEFL Impact Study (the baseline study) was to investigate the antecedent situation and the source and message segments of the process element of the model. In Phase 2 (the transition study), we will investigate factors working to facilitate or hinder change, and, in Phase 3 (after the introduction of the new TOEFL), we will investigate whether the consequences of the innovation process matched the expectations or desires of designers of the innovation would be investigated.

Phase 1 Research Questions

In light of all the above, we set out to answer three questions during this phase of the research. The first question had to do with what the impact of the new TOEFL test was meant to be, from the point of view of the advisers who helped to shape the framework documents and other experts who worked with ETS to translate theory into test design. This question related to the source and message sections of the Henrichsen framework, and the findings would contribute to the construction of the data-gathering instruments and procedures used not only in this phase but also in later phases of the research.

The second question related to what Henrichsen called *antecedents*—that is, the type of teaching taking place in the period before the introduction of the new test. The focus was on the content of current TOEFL preparation lessons, the materials in use, the methodology teachers employ, and their criteria for judging students' performance and progress. We also gathered information about the user system (the teaching institutions and their policies and practices) and the users (the teachers and students) so that we could begin thinking about how their characteristics might influence the amount and type of impact the new test might have in the future. Knowing more about the user system and the users would also help to build, in Phases 2 and 3, on Chapman and Snyder's notion of "intermediate conditions" (2000) and Messick's idea of an "evidential link" (1996).

The third question related to the most appropriate mechanism for feeding the results of the baseline study and later phases of the impact study into a TOEFL test strategy for examining dissemination practices, to create the best possible fit between the intentions of the examination designers and the understanding and abilities of the teachers and teaching institutions. We hoped to identify any factors that might hinder the implementation of the new TOEFL test and to flag these for attention.

General Description of Phase 1

Phase 1 consisted of three stages. The first stage was a groundwork phase that involved

- a survey of the framework documents and the experts who contributed to design of the new TOEFL, to see whether it was possible to narrow down the notion of positive impact that appeared in the documents and identify specific features of classroom practice that it was hoped would occur as a result of the introduction of the new examination;

- the construction of a detailed description of the new TOEFL, as deduced from our study of information from the ETS Web site, conference presentations by ETS personnel, and publications such as LanguEdge, which were designed to help users to understand the general direction in which the new test was heading;

- the design of instruments for use in interviews with teachers and students, and in observations of classrooms; and

- the selection of a sample of teaching institutions in Central and Eastern Europe, where teachers were preparing students to take the current TOEFL.

The second stage consisted of visits to 10 institutions in six countries (Bulgaria, Croatia, Lithuania, Poland, Romania, and Slovakia). The goal was to conduct in-depth individual interviews with teachers and directors of studies and group interviews with students, as well as to observe classes being taught to TOEFL groups and to other groups of advanced learners.

The third stage consisted of a detailed analysis of the data to produce descriptions of what teaching looked like in each of the four skills areas, structure and vocabulary, and to explore several themes that had emerged as important as the study was unfolding: assessment in the classroom, the role of the coursebook, the role of the computer, teacher training, and communication.

This study was small scale and qualitative in nature. It was important to analyze several situations in some depth in order to find out not only what teachers and students were doing but what their beliefs and attitudes were and what other factors might contribute to the way they pursued their goal of success on TOEFL. This provided a valuable contribution in its own right, but it would also, we hoped, complement the large-scale quantitative approach taken in a parallel investigation into TOEFL impact that was being carried out by Hamp-Lyons and Shohamy.

7

Organization of This Report

The following two sections discuss the steps followed when deciding on the research focus on and how to collect the data. The second section reports on the survey we carried out of TOEFL advisers to find out what they intended the impact of the new test to be. The third section gives details of the data collection and analysis procedures. Information from interviews and classroom observations found in visits to teaching sites in the Central and Eastern Europe region (hereafter CEE, the CEE region, or simply the region) is presented in sections 4 through 14. A summary of findings and a discussion of follow-up points for the future conclude the report.

Intended Impact of the New TOEFL

The following are excerpts from correspondence with TOEFL advisers:

Adviser 6: We talked about the importance of constructing test tasks that would positively influence what teachers would teach in ESL/EFL

Adviser 3: I don't recall discussing how this would happen. There was an implicit assumption that if we made a "good" test the impact would be automatic.

Adviser 1: While we all agreed that the consequences and impact were ultimately empirical questions, we wanted T2K (the TOEFL 2000 project) to change test preparation practices worldwide.

The first step in a baseline study is to try to understand the vision of the creators of the innovation: What do they intend the effects of their work to be and what conditions do they think need to be in place in order to achieve their vision? The baseline study should describe the current situation accurately, and, in so doing, find out whether the features that are supposed to be effects of the innovation are or are not already present in the environment. If they are not present at the baseline stage but are present later, there is some possibility that they are a consequence of the innovation. Is it then necessary to show an evidential link (Messick, 1996) before claims can be made that it is the innovation (in this case, the new TOEFL) is causing changes in the target situation (in this case, in TOEFL preparation classrooms), rather than other factors in the environment (see also Wall & Alderson, 1993, p. 67–68).

Before beginning the baseline study, we needed to uncover what the original vision was behind the new TOEFL test. Only then could we be certain that we were asking the right questions in the baseline interviews and observations. Our search for this vision took place in two steps:

> We read the framework documents to find any statements on the type of impact that was intended. These documents reported what those at ETS thought about what was first called TOEFL 2000 (the term *new TOEFL* was not used until much later).

> We asked experts who were involved in the early stages of TOEFL 2000 development work whether they had discussed test impact during that time. If they had, we asked what they had predicted the impact of the new test might be.

The Framework Documents

Discussions about the new TOEFL test seem to have started in the early 1990s and led to the publication of the framework documents in 2000. These documents represent the first full public statement about what was envisaged for the future. They contain discussions concerning the constructs that might underlie the various sections of the test (*might* because not all of the final decisions had been made when the documents were published), possible task types, and the impact the new test might have on classrooms and more widely. Our first impression on reading the documents was that impact was mentioned frequently, but when we read them more carefully we realized that most of the statements referred to impact in a general sense rather than to specific changes in the classroom. The statements we found are presented below, with the assumed cause of an impact in italics and the possible impact itself underlined.

General framework. Jamieson, Jones, Kirsch, Mosenthal, and Taylor (2000) stated:

> 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. (p. 3)

Presumably if the use of these types of items was reduced, the negative impact would also be reduced.

Listening framework. Bejar, Douglas, Jamieson, Nissan, and Turner (2000) stated:

By being aware during the research and development process of threats to validity that affect computer-based testing (CBT), and by being mindful of the potential positive and negative consequences the test could have, we expect to produce *an assessment that satisfies the demands of several constituencies without sacrificing construct representation*. For example, we expect to have <u>a positive effect on language instruction and applied linguistics research. . . .</u> We anticipate that this will encourage language <u>teachers and materials developers to focus more on communicative language use in academic contexts</u>, and that so-called "<u>TOEFL preparation courses" will</u> <u>more closely resemble communicatively-oriented academic English courses.</u> (p. 36)

Reading framework. Enright, Grabe, Koda, Mosenthal, Mulcahy-Ernt, and Schedl (2000) stated:

The consequential aspects of validity concern how test scores are used and the intended and unintended consequences of these uses. One of the motivations for revising the TOEFL test is the potential for positive washback; that is, a redesign of the test may have a positive effect on teaching practice. However, a potential unintended negative consequence is reduced access to higher education because of an increase in the cost of the test. Consequences such as these need to be anticipated and their impact evaluated as part of the research process. (p. 43)

It is our belief that improved communication about what is being tested and improved interpretability of examinee performance will have <u>positive washback effect</u> as well. Assessing reading with various types of prose and nonprose documents, and with writing, listening and speaking in the integrated skills tasks, broadens the constructs being measured to more realistically represent real-world language needs and language use. Research can be designed to investigate washback effects on what examinees study and to determine whether the <u>emphasis on communicative learning</u> increases once the new test is operational. (p. 49)

Writing framework. Cumming, Kantor, Powers, Santos, and Taylor (2000) stated:

Thus with *multiple writing tasks that include both independent and content-dependent tasks*, we hope to move <u>beyond the single independent essay model to a writing model that is more reflective of writing in an academic environment</u> while also dealing with interdependency issues. (p. 9)

Speaking framework. Butler, Eignor, Jones, McNamara, and Suomi (2000) stated:

We expect that the introduction of an oral component will have a positive washback effect on the ESL teaching community. 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. (p. 23)

There is certainly a sense of cause and effect in these statements. It is important to be clear about what the causes were presumed to be and what effects they were supposed to lead to in the future. This helps in the search for Messick's evidential link in Phases 2 and 3 of this study. Table 1 lays out the presumed causes and desired positive effects unambiguously.

This breakdown gives some insight into what the framework designers thought the positive impact of the test might be but the third column shows that most of the effects were general (e.g., a positive effect on language instruction and applied linguistics and focus more on communicative language in use in academic contexts). The only specific effects mentioned were that there would be more than one type of writing model in the classroom and that students would learn to communicate orally (which is still, in fact, fairly general). The only negative effect that was mentioned (reduced access to higher education) was a social consequence rather than an impact on the teaching and learning in preparation classes, so it is beyond the brief of this study.

Inspection of the framework documents gave us less information than we had hoped for, but it suggested that we needed to pay attention to certain features when carrying out the baseline study—in particular, whether classrooms had a communicative or academic orientation, whether students were being asked to deal with more than one model of writing, and whether they were being taught speaking. We needed to turn to other sources to find more specific examples of intended impact. This formed the second step in the investigation.

Table 1

Test Impact: Presumed Causes and Possible Effects

Presumed cause	Possible effect/s	General or specific effect	Positive or negative effect
General framework:			
A reduction in discrete-point test items and multiple-choice items	Eliminate a negative impact on instruction	General	Positive
Listening framework:			
An assessment that satisfies the demand of several constituencies without sacrificing construct representation	A positive effect on language instruction and applied linguistics	General	Positive
	Teachers and materials developers will focus more on communicative language in use in academic contexts	General	Positive
	TOEFL test preparation courses will more closely resemble communicatively oriented academic English courses.	General	Positive
Reading framework:			
A redesign of the test	A positive effect on teaching practice	General	Positive
An increase in the cost of the test	Reduced access to higher education	General	Negative
Improved communication about what is being tested and improved interpretability of examinee performance	A positive washback effect	General	Positive
Assessing reading with various types of prose and nonprose documents, and with writing, listening and speaking in the integrated skills tasks (broadening the construct)	The emphasis on communicative learning increases	General	Positive
Writing framework:			
Multiple writing tasks that include both independent and content-dependent tasks,	Move beyond the single independent essay model to a writing model that is more reflective of writing in an academic environment	Specific	Positive
Speaking framework:			
Using constructed response items, which are less likely to be coachable	Students will learn to communicate orally—not to learn a skill simply to do well on a test.	Specific	Positive

Survey of TOEFL Advisers

We asked ETS for the names of experts they had commissioned to help in the conceptualization of the new TOEFL, and ETS staff sent us the names of 19 advisers. We wrote to the advisers with a set of 10 questions, presented in Table 2.

Table 2

Questions to TOEFL Advisers

1	What part did you play in the development work that resulted in the new TOEFL examination? (e.g., Did you work on a particular committee? Did you contribute to a particular document?) What year(s) did you do this work? Are you still involved?
2	What was the remit you were given when you were invited to participate in the development work?
3	Were you (or you and your colleagues) asked to think about the impact of changes in TOEFL on teaching, learning or any other aspect of education?
4	Did you explicitly discuss the types of impact you wanted the new TOEFL to have (positive impact)?
5	If you did, what types of impact did you mention? (Please give as much detail as you can.)
6	How did you think this impact would occur? (e.g., Was there anything in the design of the test that would encourage changes in the classroom?)
7	Did you explicitly discuss the types of impact you didn't want the new TOEFL to have (negative impact)?
8	If you did, what types did you mention? (Please give as much detail as you can.)
9	Why did you think these types of impact might occur? (e.g., Was there something in the design of the exam that teachers might not understand or be able to cope with?)
10	Has your thinking regarding the potential impact of New TOEFL changed? If so, what thoughts do you have now about the potential positive or negative impact of the new examination?

We received 10 replies and entered into further correspondence with several respondents when we felt it was necessary to clarify some points. All of the advisers were well-placed to talk about the thinking that had gone into the impact of the new test, having served as special consultants to the TOEFL 2000 development effort, on the Committee of Examiners, or on committees devoted to research and design for all of the skills areas. Over one half of the advisers replied that they had been involved in discussions regarding the potential positive impact of the new TOEFL test. Figure 2 reproduces a sample of their responses. These answers indicate how optimistic they seemed at the time the new test was taking shape.

13

"It was a regular topic at all team meetings and at larger meetings and ReDOC meetings."

"Yes—impact was an implicit part of the development agenda and we discussed it explicitly on a number of occasions!"

"If teachers are going to teach to the test, then the new test ought ideally to encourage them to teach worthwhile stuff—and the learners to acquire worthwhile abilities."

"We talked about the importance of constructing test tasks that would positively influence what teachers would teach in ESL/EFL classrooms in TOEFL prep courses and that would positively influence what learners would see as essential for studying English."

"As we develop any new assessment measure here at ETS, we are always considering the impact that it will have on the examinees, the users, and the field."

"Our thinking was that by increasing the authenticity of the test tasks, we would reduce the likelihood of negative impact."

"Specifically, we thought that interactive reading-writing and listening-writing tasks requiring summarizing and paraphrasing were not only essential academic skills but were also ones that students all over the world could practice in their English classes and in TOEFL cram schools. Thus, the washback effect would only be positive and beneficial to international students for success on both TOEFL and in North American colleges and universities."

"Since TOEFL prep courses will probably always be in existence, they might as well help students become as proficient as possible at something academic and authentic."

Figure 2. **Selected quotes from the advisers' responses.**

Table 3 presents a summary of the information the advisers gave in response to the question about what sorts of positive impact they had envisaged.

Table 3

TOEFL Advisers' Views of Positive Impact

Questions and responses	Number of responses
Question 4: Was positive impact discussed?	Yes = 6 No = 2
General positive impact	
Changes in test preparation practices	4
Improved academic language and skills	3
Students rethink what they need to study	2
Reduction in memorization and test-taking techniques as a preparation method	2
General: Authenticity	
More authentic language input	1
More authentic (academically relevant) tasks	1
General: Register	
Study of formal and informal registers	1
Integrated Skills	5
Reading	
Complex reading texts	1
Study of more complex rhetorical structure	1
Longer texts and making connections between different parts	1
Writing	
Emphasis on summary and paraphrase skills	1
Working at discourse level rather than dealing with decontextualized grammar and vocabulary	1
Speaking	
Speaking will be taught	2
More emphasis on productive skills	2
Study of more pragmatic force of utterances	1
Testing impact	
More meaningful results	1
More differentiation amongst test takers (especially at higher end of scale)	1
Greater understanding of variables that contribute to successful performance (can be used to produce proficiency scales)	1
Social impact	
Increased fairness/access due to new technology	1

Note. n = 10.

As with the framework documents, most of the references to positive impact were still quite general: changes in test preparation practices, improved academic language and skills, and so on. There were, however, a few examples of specific changes in the teaching of reading, writing, and speaking. Some of the advisers expressed hope that integrated tasks and speaking would be emphasized more in the future than they (presumably) were when the new test was first discussed. These specific examples were indications of what we should be looking for when we visited teaching institutions for the baseline study, and we made notes to include them in the observation schedule (see the Methodology section in this report for more details).

We also asked whether the advisers had been involved in discussions of negative impact. Table 4 presents their replies.

Table 4

TOEFL Advisers' Views of Negative Impact

Questions and responses	Number of responses
Question 7: Was negative impact discussed?	Yes = 5 No = 1
Testing issues 　Construct irrelevant variance such as testing background knowledge (e.g., of the United States higher education system), which might disadvantage some candidates	1
Teaching issues 　Teachers who are nonnative speakers of English might encounter language problems since higher skills required by new test	2
Social issues 　Equity issues due to probable increase in exam fees and varying levels of familiarity with computers	1

Note. n = 10.

It is clear that there was some concern about possible negative effects, but it is notable that more specific examples were given of positive effects (see Table 3) than negative ones. The mention of the possibility that teachers might not be able to handle the higher skills envisaged for the new examination (at that stage at least—we do not know yet whether this will materialize) was a reminder to pay special attention to the language ability of the teachers interviewed and observed. The point about construct-irrelevant variance reminded us to ask the teachers and

students they felt the current TOEFL test (and future versions) might be testing apart from the four skills and language proficiency.

We also asked the advisers whether they had discussed *how* positive impact would occur. Only half of them answered this question, and only four of these said they had been involved in discussions of this sort. The responses they gave showed a trust in the notion that if certain task types were included in the test (e.g., speaking, integrated skills, and reading to learn tasks), and (according to one informant) if these were weighted appropriately, the same tasks would be practiced in the classroom. Only one respondent mentioned that more than good task design might be needed to create positive impact, specifying the need for test preparation materials (specifically, model tests), workshops for teachers at conferences, and transparency in the test design process:

> Adviser 2: Most of the work that was undertaken for development of the test would be made available to the public.

From the study of the literature on test impact and educational innovation, it appears that not as much thought had been given to the process of diffusion (Fullan, 1991) as to the desired effects of the innovation, and that what thought had been given concentrated on characteristics of the innovation (in this case, the inclusion of certain task types in the test design) rather than other factors (principally having to do with communication) that are needed to facilitate the appearance of the desired effects (Stoller, 1994).

Summary

The purpose of reading the framework documents and asking the experts for their memories of discussions of intended test impact was to inform the design of the instruments used for conducting interviews and observations in the target institutions. What we understood from this survey was that 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. What was interesting were the assumptions that these features were not already present in preparation classrooms.

17

Little had been published up to that point (and still) about what kinds of teaching and learning occurred in TOEFL test preparation classes. Apart from Alderson and Hamp-Lyons (1996) what had been presented in articles was also quite general rather than based on data collection and analysis (Raimes, 1990; Peirce, 1992). The Alderson and Hamp-Lyons account gave a detailed description of teaching in North America, and we made much use of their work when deciding how to approach our own investigation in the CEE region.

The details of this investigation are presented in the remaining sections of this report. The next section presents our methodology and the rest of the report presents our findings and conclusions.

Methodology

This section presents an account of the decisions we took and the procedures we followed when designing the baseline study and analyzing the data collected during interviews and observations. The section is divided into five parts: Sample, Instruments, Data Collection, Analysis, and Summary

Sample

Sites. The brief was to focus on countries in Central and Eastern Europe. The following personal communication from ETS presents the rationale for this:

> Given the changes these countries are undergoing in many areas, the subcommittee was interested in learning more about current teaching practices there and how they might change after a new TOEFL is introduced. (M. Enright, personal communication, June 27, 2003)

We defined this area as countries that were members of the former Soviet bloc but that had either gained independence from the former Soviet Union (e.g., the Baltic States) or had opened up since the fall of the Berlin Wall.

We were asked to investigate the impact of changes in TOEFL on coaching and formal instruction. We took this to refer to teaching in postcompulsory educational settings since it was unlikely that instruction in primary or secondary schools would be affected by a test designed for individuals planning to pursue tertiary-level studies abroad. We knew from our previous work in the region that the sites that were most likely to prepare students for education in an English-

language setting were private-language schools. It might also be possible to find suitable classes in university departments of English, Public Service Language Centers (not-for-profit bodies, often catering to civil servants), and nongovernmental organizations. The initial criteria for site selection are presented in Table 5.

Table 5

Initial Criteria for Site Selection

The site should	Rationale
offer both TOEFL preparation classes and English for Academic Purposes (EAP) courses	We wanted to observe classes of both sorts run by the same teacher to see the similarities or differences between them (Alderson & Hamp-Lyons, 1996; Watanabe, 1996).
offer two teachers teaching TOEFL test preparation classes	We wanted to see whether any differences between TOEFL and non-TOEFL classes were common across teachers (Alderson & Hamp-Lyons, 1996; Watanabe, 1996).
offer courses preparing students for CBT	We originally believed that CBT candidates outnumbered candidates for the PBT (see Appendix A), and so thought CBT preparation courses would be more representative of what was happening in the region.
	We were also not sure of the future status of the PBT (our understanding was that it would disappear with the introduction of the new TOEFL), and since we planned to conduct a follow-up study (Phase 3) in several years' time, it seemed logical not to visit PBT preparation classes as they would no longer be given in the future.
have good e-mail facilities and personnel who responded to our messages promptly	It was important that there should be good communication with the participating institutions to enhance project efficiency.

Note. CBT = computer-based TOEFL test. PBT = paper-based TOEFL test.

Our initial approach was to follow up personal and professional (first- and second-hand) contacts to identify suitable institutions. We also conducted a trawl of the Internet and asked ETS staff to help us with their contacts. These sources furnished a certain number of possible sites, but many institutions did not respond to messages and others proved to be unsuitable for various reasons.

This caused us to question our original assumptions about where candidates were sitting for the TOEFL. A series of enquiries and requests for clarification revealed that the computer-based TOEFL (CBT) was in fact only offered in 9 of the 21 countries in the region.[1] This had an important effect on the design—if there were no CBT test centers in a country, there would probably be no CBT preparation courses. We widened the search and included institutions that operated in countries where the paper-based TOEFL test (PBT) version was available (see Appendix B).

We identified several more institutions, but few matched the criteria in full. There was less formal preparation being offered in the region than we had originally estimated. We learned this through messages from language teaching professionals responding to requests for cooperation (see Appendix C) and through further information from ETS (Appendixes D and E), which indicated that although substantial numbers of students from the region were taking the TOEFL, the numbers taking it in their native country were relatively low. We suspect that many candidates were taking the test in the United States or other English-speaking countries, perhaps after a period of language development.

We identified 10 sites in six different countries: Bulgaria, Croatia, Lithuania, Poland, Romania, and Slovakia. There were two sites in Bulgaria and Croatia, and three sites in Lithuania. These sites covered four of the eight countries offering CBT by then, including two of the four offering only CBT, and two of the 13 countries offering only PBT (see Appendix B).

The institutions fell into two main categories: private language schools, which offered a range of general language courses including TOEFL preparation courses, and education information centers (one of them affiliated to the U.S. Department of State), which, amongst other duties, offered test preparation courses. Two of the private schools were part of chains but the others were independent operations. All but two sites were located in capital cities. The institutions were generally relatively small organizations, housed in old buildings converted for their current purpose, with teaching rooms that were average or rather small. They differed greatly in the resources they had available, most apparently in their provision of computers (see the section on the role of the computer). Demand for TOEFL preparation was not so high as to require more than one course at a time, apart from in one institution. This was the only institution where we interviewed two teachers.

We found that we could not observe EAP classes to contrast with the TOEFL classes since EAP courses were not offered. In some cases, we were able to observe whatever class the teacher had on his or her timetable in addition to their TOEFL lessons, and in other cases we observed the teacher giving a class elsewhere, outside the school. Table 6 presents a summary of the data we were able to collect at each site.

Table 6

Summary of Data Collected (10 Sites)

Country	Bulgaria [a]		Croatia [b]		Lithuania [b]			Poland [c]	Romania [a]	Slovakia [c]
Site no.	B1	B2	C1 [d]	C2 [e]	L1	L2	L3	P1	R1	S1
TOEFL observation	✓	✓	✓	x	✓	✓	x	✓ (x2)	✓	✓
Non-TOEFL observation	✓	✓	x	x	✓	✓	x	x	x	✓
Teacher interview	✓	✓	✓	✓	✓	✓	x	✓ (× 2)	✓	✓
Student interview	✓	✓	✓	x	✓	✓	x	✓ (× 2)	✓	✓ (× 2)
Director of studies interview	✓	✓	✓	✓	✓	✓	✓	✓	✓	x

Note. × 2 = two sets of data for this category were collected.

[a] CBT and PBT. [b] Only CBT. [c] Only PBT. [d] Site C1 was also the piloting center. [e] Teacher and director of studies were the same person.

Participants. There were 39 informants in all: 10 teachers, 21 students (generally in groups of two or three), and 9 directors of studies (the teacher and director of studies in Site C2 were the same person). Teachers and students were the people who were most likely to be affected by changes in the new TOEFL. The directors of studies provided information about the teaching institutions that the teachers and students might not have. The Henrichsen (1989) model makes it clear that characteristics of the user system—in this case, the teaching sites—can facilitate or hinder the adoption of an innovation.

Of the 10 teachers, 6 were local and 4 were American expatriates. The local teachers all had a high level of English proficiency and were confident and competent users of the language. They had all received formal training and qualifications to teach in the local state school system, while the expatriates had either only basic Teaching English as a Foreign Language (TEFL)/Teacher of English to Speakers of Other Languages (TESOL) qualifications or no teaching qualifications at all. There was a wide range of teaching experience overall—from 2 months to 40 years. One teacher had personal knowledge of the current TOEFL, having taken the test herself, but the others had gained their knowledge via the coursebooks they used and from their students (this is discussed further in the Communication section in this report). For more details on the teachers, see Table 7.

Table 7

Details of Participants: Teachers

Teacher ID	TB1	TB2	TC1	TC2[a]	TL1	TL2	TL3[b]	TP1A[c]	TP1B[c]	TR1	TS1
Nonnative speaker of English	✓	✓	✓	✓		✓				✓	
Native speaker of English					✓			✓	✓		✓
Female	✓	✓	✓			✓			✓	✓	
Male				✓	✓			✓			✓

Note. $n = 10$. The column headings follow this pattern: TB1 means teacher at Site B1—Bulgaria. See Table 6 for an explanation of the site numbers.
[a] Site TC2—This teacher was also the director of studies. [b] Site TL3—No teacher was interviewed. [c] Site P1—Two teachers were interviewed separately.

All of the students were in their late teens or 20s and most were taking TOEFL in order to study abroad, though not necessarily in English-speaking countries. Their English-language competence varied greatly but they were all able to communicate well enough to participate in the interviews effectively. Some had little experience conversing with a native speaker of English, while others had already traveled extensively. For more details on the students, see Table 8.

Table 8

Details of Participants: Students

Student group ID	SB1			SB2		SC1		SC2[a]	SL1		SL2		SL3[a]	SP1A		SP1B			SR1		SS1		
Student number	1	2	3	1	2	1	2		1	2	1	2		1	2	1	2	3	1	2	1	2	3
Female	✓	✓	✓	✓					✓		✓	✓				✓	✓		✓	✓	✓	✓	
Male					✓	✓	✓			✓				✓	✓			✓					✓

Note. $n = 21$. The column headings follow this pattern: SB1 means student at Site B1—Bulgaria. See Table 6 for an explanation of the site numbers. No students participated at Sites C2 or S1.
[a] Sites SC2 and SL3—No students were interviewed.

Of the nine directors of studies, three had come from a TEFL/TESOL background. This proved most useful, as they were aware of the concepts we were discussing and could offer us extra insights. Some were also involved in the teaching program and could therefore talk about issues affecting the teachers. One was not only the director of studies, but also the owner of the school, so he could give us a perspective on how the school as a whole operated, which was motivated by a need for profits as much as by pedagogical concerns. This director of studies had taken the TOEFL and so had interesting views on what was needed for the preparation classes. All the directors of studies were local nationals but spoke English well, so initial plans to engage interpreters in some countries proved unnecessary. For more details on the directors of studies, see Table 9.

Table 9

Details of Participants: Directors of Studies

DOS ID	DOS B1	DOS B2	DOS C1	DOS C2[a]	DOS L1	DOS L2	DOS L3	DOS P1	DOS R1	DOS S1[b]
Female			✓		✓	✓	✓	✓	✓	
Male	✓	✓		✓						

Note. $n = 8$. The column headings follow this pattern: DOS B1 means director of studies at Site B1—Bulgaria).

Site DOS C2—Director of studies was also a teacher. Site DOS S1—Director of studies was interviewed, but tape was inaudible.

Instruments

Sources. In addition to the background provided by the literature review (in particular, Henrichsen, 1989), three sources informed the instrument design:

1. Information on the intended impact of the new TOEFL

2. Details of the nature of TOEFL—both the current versions and the new TOEFL

3. Our own knowledge and experience of previous impact studies in the field of language testing.

Information on the intended impact of the new TOEFL. The first sources of information were the framework documents and the views of TOEFL advisers who had either participated in the writing of these documents or had worked with ETS at a later stage to turn the notions expressed in them into test descriptions (see the second section for efforts to identify intended impact).

Details of the nature of TOEFL. We created a comprehensive picture of the two current versions of TOEFL with reference to a wide variety of ETS documentation, including the Information Bulletin (ETS, 2002b), the content of the ETS TOEFL Web pages (http://www.toefl.org), materials aimed at students and teachers such as *TOEFL Tips—Preparing Students for the Computer-Based TOEFL* (ETS, 1999) and the POWERPREP® test preparation package on CD (ETS, 2000). We then constructed a table to facilitate comparison between the versions, which we completed with information about the new TOEFL (further details follow).

At this stage of the project, ETS had not yet released descriptions of prototypes for the new TOEFL test as it was still under construction. Using the following sources, we had to build a picture of the test.

- The framework documents. These provided basic details such as the planned modes to be tested, and the thinking behind the proposed revision.

- The ETS TOEFL Web site (http://www.toefl.org). The site had some updates regarding the new TOEFL.

- The LanguEdge learning tool (ETS, 2002a). This publication did not contain a description of the new test, but it was intended to serve as a means of introducing the concept of change to TOEFL users.

- ETS personnel. Dr. Mary Enright of the Research team and Tina Wright of the Test Development team provided a great deal of information, especially when checking the draft comparative table.

- Conference presentations given by ETS personnel (Roberts & Tumposky, 2003; Malloy & Tumposky, 2003).

This information enabled the completion of the table comparing the PBT, the CTB and, what was known about the new TOEFL test (see Appendix F). This allowed us to highlight points of similarity and differences between the three versions of the test and, above all, to focus on the aspects of TOEFL *that* might be reflected in the teaching of preparation classes. We would later incorporate these features into the interview and observation instruments.[2]

Knowledge and experience of previous impact studies in the field of language testing. We drew quite heavily on our own experience of working on other impact studies, as well as on the experience of other colleagues. The studies included the Sri Lankan baseline study (Wall & Alderson, 1993), the St. Petersburg baseline study (St. Petersburg Examination Project, 1996), the Hungarian baseline study (Fekete, Major, & Nikolov, 1999), and the instrument design phase of the International English Language Testing System (IELTS) impact study (Alderson & Banerjee, 2001). We also conducted workshops with the Language Testing Research Group at Lancaster University. All of these sources added to an understanding of the type of impact that might be seen after the introduction of the new TOEFL test.

We had planned to conduct a minimum of three interviews at each of the sites: one with at least one TOEFL teacher, one with the director of studies and one with at least one group of students (ideally two or three students in a group). We also had planned to conduct two lesson observations per teacher, one of a TOEFL preparation class and another of a non-TOEFL class.

Interviews. We devised structured interviews to make sure that certain topics were covered with all participants and to provide a basis for cross-referencing between them. However, we also built in space to add further comments from the informants should these seem relevant. Questionnaires would have allowed us to be more confident in the representativeness of

our findings (by allowing us, potentially, to reach more teachers), but this was not the goal. We wished to gather data that would offer deeper insights than questionnaires could provide, even if it meant working with fewer participants.

The areas covered in the interviews were determined mainly by reference to the Henrichsen (1989) hybrid model of the diffusion/implementation process. The Henrichsen model incorporates findings from many researchers in the area of innovation (particularly, but not exclusively, in innovation in education). It divides the process of innovation into three stages: antecedents, process, and consequences. We wished to use the interviews to collect data relevant to the antecedents stage. According to Henrichsen, it is necessary to carefully analyze the characteristics of the intended user system, the characteristics of the intended users, traditional pedagogical practices, and the experiences of previous reformers in order to understand their contribution to the success or failure of an innovation to be introduced in the future. Henrichsen listed the points to focus on in each of in these areas, but these were modified by Wall (1999) in a detailed investigation of test impact in Sri Lanka. The modified version includes the following factors:

- Characteristics of the intended user system: classroom conditions, school factors, educational administration, political factors, geographic factors, economic factors, cultural factors

- Characteristics of the intended users (to be collected for both teachers and students): attitude to education, attitude to classroom teaching, attitude to English teaching, attitude to exams, attitudes to English, attitude to new ideas, level of education, abilities, personal factors, economic situation, interest, goals

- Traditional pedagogical factors: coverage of syllabus, content of teaching, methods of teaching, methods of classroom assessment

- Experience of previous reformers: the outcomes of earlier attempts to innovate, including ideas about why these attempts succeeded or failed

While the focus of the investigation was mostly on the topics listed under traditional pedagogical factors, we also wanted to cover as many of the other areas as possible, without subjecting informants to a very long or intense interrogation. We set a number of questions for

both teachers and students, to get both points of view, and set other questions for the directors of studies. The following are the themes we covered:

- Teachers and students:

    o Course aim, content, materials, and methodology

    o Class-based assessment

    o Test-taking strategies

    o Language development strategies

    o Computer familiarity

    o TOEFL awareness

    o Attitude to TOEFL

    o Attitude to tests in general

    o Predicted impact of the new TOEFL test

- Additional teacher themes:

    o Teacher background and experience

    o Teacher support

    o Predicted impact of new TOEFL test

- Directors of studies:

    o Nature of the courses offered

    o Teacher recruitment and support

    o Resources

    o TOEFL awareness

    o Predicted impact of the new TOEFL test

Copies of the interview schedules for teachers, students, and directors of studies can be found in Appendixes G, H, and I respectively.

Observations. We aimed to observe two classes taught by each of the teachers. There were two reasons for this decision. The first was to see whether there were any similarities or differences in the two classes. The second was to try to tease out the teacher variable from the test impact variable. If we saw, for example, that certain features did not appear in a teacher's TOEFL class, we would not want to claim that this was due to test impact without an initial check as to whether the same feature appeared in the teacher's non-TOEFL class (Alderson & Hamp-Lyons, 1996; Watanabe, 1996).

The observation schedule we constructed can be found in Appendix J. It consists of a front page, which deals with the objectives of the lesson and background information about the group and the course. This section was meant to be filled in before the lesson began.

The next five pages deal with the activities observed in the lesson. There is room for a free description of the activities, as well as a checklist of activities, text types, and question types we expected to see in current TOEFL classes, and possibly in classes preparing for the new TOEFL test in the future. We designed this checklist using the comparative table previously discussed (see Appendix F). We included sections on Listening, Reading, Writing, and Structure. All of these are tested separately on the CBT, with Structure being tested as part of the Structure and Written Expression section on the PBT. Writing is not an integral part of the PBT. Candidates have to take the separate *Test of Written English*™ (TWE®) if they need a result in Writing.

We also included a section on Speaking, for comparison of the number and types of speaking activities taking place at present with what might occur after the introduction of the new TOEFL test. The checklists included activities related to the new TOEFL (labeled NT). We did not expect to see many of these activities in Phase 1, but we included them to provide a point of comparison for Phase 3 (which was expected to be carried out after the introduction of the new examination).

Pages 7 and 8 of the schedule provided a systematic way to capture impressions and observations of features of the lesson, which were felt to be of use in later analysis of the classes. They provided indicators of how communicative the class was and an overall picture of the skills balance in the lesson. Most importantly they reminded the observer to consider ways in which the lessons seemed to be influenced by the current TOEFL. The last page of the instrument was set aside for a map of the classroom, to provide a visualization of each class during analysis. This

also provided a record of such things as classroom layout and resourcing, which might be determined either by the teacher or by the teaching institution.

Piloting. We piloted our instruments in two locations: a private language school in London and a United States information center in Croatia. The modifications we made were not so much in terms of content but rather in the ordering of questions in the interviews (questions probing areas of a more abstract nature were moved to later in the schedule, to give the interviewee time to warm up and feel comfortable with the interview process) and the layout of the observation schedule. We also revised the estimates of the amount of time it would take to cover all of the questions in the interviews. An inspection of the pilot data revealed its usefulness, so we decided that it would be included in the main body of data for analysis.

Data Collection

Timing. We originally planned to gather data between May and August 2003, but it became apparent early in the project that TOEFL preparation classes were seasonal. This was confirmed in later interviews with directors of studies and teachers. The main periods for enrolment were what in Europe were autumn and spring, as the majority of the students were planning to apply for higher education abroad and needed to have their results available in time to submit applications. The data collection period, as a consequence, was between September and December 2003.

Recording. The interviews and observations were recorded on a Sony MiniDisc recorder (Net MD Walkman MZ-N710), which was chosen as it offered the highest sound quality and robustness out of the options considered. It was also small and therefore easy to carry and not very intrusive. Interviews and classes could be recorded without a break, with a recording capacity of 5 hours and 20 minutes on one disc. It was also easy to transfer data between the research team and the staff who undertook the transcribing, and it was convenient as regards storage and archiving.

Analysis

All the interviews were transcribed off site and the transcripts were thoroughly checked when they came back from the transcribers. They were coded using ATLAS.ti (version 4.2). The list of codes was derived from the Wall (1999) adaptation of the Henrichsen (1989) model. The

focus was on traditional pedagogical practices, characteristics of the intended user system, and characteristics of the intended users (see Appendix K for a complete listing of the codes).

We had not managed to gather data on experience of previous reformers so we did not set aside a category for this code. We created codes for characteristics of the innovation and for characteristics of the resource system, which are part of the process section of the Henrichsen model. These were not used often during this phase of the study, but we predict that they will be important in Phase 2, after more information has been released about the new TOEFL test, and Phase 3, after the new test has been administered for the first time.

We also added a number of new codes to those derived from the Henrichsen model. There were three main reasons for this. The first was that some of the original codes proved too wide and important distinctions were being lost in the analysis. An example of this would be Teacher Ability, to which we added the code Teacher Technical Ability to cover the teachers' familiarity with computers. The second reason was that codes for characteristics of the users had to be replicated for each set of interviewees (e.g., the code TAw [teachers' awareness of current TOEFL] became SAw [students' awareness of current TOEFL] when we dealt with student responses, and DAw when we dealt with directors of studies). Lastly, in a very few cases key details were not covered by the extant codes and so new codes had to be created (e.g., the code identifying which version of the TOEFL students expected to sit for).

Agreement between coders. All of the interview transcripts were coded by both researchers. We held regular meetings to discuss the codings and to resolve all discrepancies through discussion. Whenever a new code was added, transcripts coded prior to that were checked and recoded accordingly. Once completed the coded instances were viewed in various ways:

- By institution (e.g., all data from Site 1 in Bulgaria)

- By interviewee type—all data from teachers

- By response to a certain question—all data relevant to a specific question on either the teacher or student schedules or both.

Observations. We dealt with the data from the observations in two ways. First, a summary description of each observation was created to check the impressions gained from analysis of the interview data (see Appendix L for summaries of a TOEFL and non-TOEFL class given by the same teacher). Second, we drew up a table showing which activities in the

observation checklist had occurred in the TOEFL classes we visited (see Appendix M). This, like the summary description, provided a way of checking whether what the individual teachers said in their interviews matched what we had seen in their classrooms. It also allowed us to see whether there were any patterns that all the TOEFL classes followed.

Presentation of findings. We decided that the best way to present our findings was by looking first at the teaching of the language aspects tested by the TOEFL, either currently (Listening, Reading, Writing [CBT] and Structure [CBT], and part of Structure and Written Expression [PBT]) or in the new TOEFL test (Speaking). Vocabulary is no longer tested separately but it featured prominently in the data and thereby merits detailed treatment. We also decided to discuss some of the themes that emerged from the data and which corresponded closely to important aspects of the Henrichsen framework. Time constraints did not allow us to deal with all of these themes in this report, but we have devoted a section each to assessment in the classroom, the role of the coursebook, the role of the computer, teacher training, and communication.

One of the goals of this study is to present the voices of the participants who are most likely to be affected by changes in the TOEFL. We chose to do this in two ways: First, by introducing many of the sections in this report with extracts from interviews that represented key points regarding the theme of the section and, second, by including a number of quotations in the text itself. We believe this method of reporting will provide a richer picture than tables listing the numbers of teachers offering each opinion.

Summary

This section has outlined all of the stages included in the setting up and implementation of Phase 1—the baseline study. Figure 3 provides a visual representation of the stages.

The next five sections cover what was discovered about the teaching and learning for Listening, Structure, Reading, Vocabulary, Writing, and Speaking. (The order of these sections basically follows the order of the CBT. Vocabulary is not tested separately, but teachers perceive it to be closely related to Reading, so it is included after Reading. Speaking is dealt with last as it is not tested now but will be in the new TOEFL test.) The sections after that (Assessment in the Classroom, the Role of the Coursebook, the Role of the Computer, Teacher Training, and Communication) cover the themes emerging from the data that corresponded to key areas in the Henrichsen model. The Summary and Conclusions section presents a summary of findings in all

31

these areas, plus further discussion of the Henrichsen model and other issues of relevance to this investigation.

Preparation	Data Collection	Analysis
• Literature review • Consultation with TOEFL advisers • Analysis of TOEFL • Brainstorming Instrument development Pilots	10 sites in 6 countries: 39 informants in total	Double-coding using ATLAS.ti ↓ Interpretation ↓ Selection of key themes ↓ Report
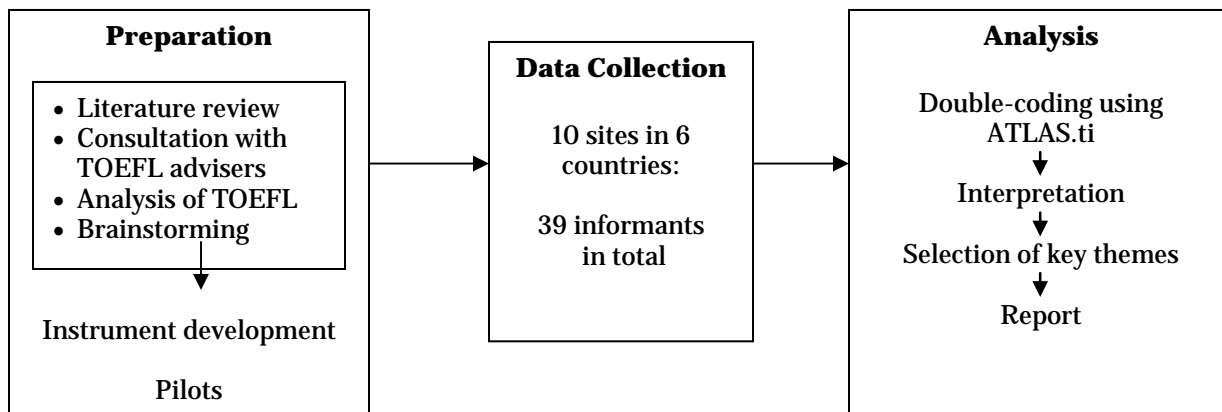
Figure 3. **Stages of the baseline study.**

Teaching Listening

TL2: 293[3]: I find it's very difficult to teach them to hear.

TC2: 613–617: So just listen . . . when you iron your clothes or when you do something else . . . just put a channel on and just listen to it so you can get it into your ear. That's the only way you're going to pass the test properly.

These quotes represent two important aspects of the teachers' approaches to preparing their students for the Listening test. There was evidence, first, of a paucity of techniques to actively improve listening skills, and, second, of the belief that maximizing exposure to spoken English was optimal. Copious practice of test-like items was the most common way of teaching listening in the classroom. This section discusses the teachers' views of what is tested, what they try to teach, how they go about their teaching and some of the issues that arose as we analyzed the data.

What Is Tested?

Most of the teachers seemed to have some understanding of what was being assessed in the TOEFL Listening test, at least on a general level: the ability to understand lectures (all but

two teachers believed this), the ability to infer opinions (all but two), and the ability to guess the meaning of unknown words by using context (all but two). (See Appendix N for the responses teachers gave to Question 95 on the interview schedule, regarding what was tested on TOEFL.) Most teachers also believed that TOEFL tested the ability to understand everyday language (though interestingly the ETS representative who answered the same question said that the test did not intend to test this aspect of listening—see Appendix N). However, this general awareness did not always help teachers to understand how to break listening down into more teachable subskills or to choose techniques that would help students to develop these subskills efficiently. This is exemplified by TP1B, who was not able to judge what her students could already do in listening and therefore was not sure what to emphasize and what to omit in her classes:

> TP1B: 693–697: . . . when we were doing reading and writing there were things that I knew that they understood because I had seen their writing skills to know that they understood how to do it, so I would skip that. But with listening, I go through the exercises as they come along in the book. . . .

The TOEFL preparation books, however, are not helpful when it comes to analyzing what listening consists of or how to develop better skills. The Barron's book, for example, offers little advice other than the recommendation to practice (Sharpe, 1996, p. 50), while the Cambridge book recommends paying attention to context; details and purpose; and listening for meaning, vocabulary, and structure (Gear & Gear, 2002, p. 74–75), without giving guidelines on how to do this.

Listening in the Syllabus

The syllabus of several of the TOEFL preparation courses was designed to reflect the organization of the test itself, with listening emphasized in the first portion of the course, then structure, and then the other skills (SB1: 318 and SB2: 200). In other courses, it was the individual lessons that were divided up this way:

> TC2: 344–345: Basically they have 3 hours. First up they have listening, grammar, and then reading and reading at home a lot.

There were exceptions to the rule, however. One of the teachers had consciously decided not to follow the order of skills proposed in his preparation book, believing that it would be better to leave listening to the end of the course when his students would have had a chance to acclimatize by listening to his voice for several weeks (TL1: 311–315, 400). Another teacher, who demonstrated a high level of reflection and teaching expertise, chose not to do any listening practice at all during class time, apart from giving a few practice tests at the end of the course (TL2: 309 and SL2: 193). She claimed that she could not teach the students to improve and so did not see it as a good use of their time together (TL2: 358). She asked her students to complete listening tasks from a TOEFL preparation book at home and then later briefly discussed in the classroom the difficulties they had encountered.

All but TL2 used English as the medium of communication in the TOEFL preparation classes and, although they may not have been as aware of it as TL1 was, they were probably helping their students to get used to the tone, pitch, intonation, and so on, of the language. The students who had native-speaker teachers had an especially enriched experience in this regard, but all of the teachers had a high level of spoken English and were capable of helping their students to develop an ear for the language. Ironically, however, the listening passages in the TOEFL test itself do not display all the features of authentic speech. According to Buck (2001, p. 223), ". . . the oral features of spoken texts—phonology, hesitations, and the discourse structures typical of unplanned spoken language—are almost entirely absent (in TOEFL).

This may put the students at an advantage in the test situation, however. TL1's former students reported that they had found the TOEFL practice passages to be much slower than their teacher's relatively ungraded, normal speed, native-speaker English (SL1: 737) and therefore easily comprehensible.

What Is Taught?

The interviews revealed that all but one of the teachers mainly followed the exercises in their TOEFL preparation books, using the accompanying cassettes or CD-ROMs. They often followed them quite closely (TB1: 353, TC1: 608, TC2: 1062, TL1: 501, TP1A: 514, TP1B: 204, TP1B: 617, TP1B: 696, TP1B: 1131, TR1: 504, TS1: 803, SR1: 293). TC1 said it was easier for teachers this way (TC1: 617), which is resonant of the teachers interviewed in the Alderson and Hamp-Lyons study (1996) who followed their books so closely that they felt the need for only minimal, if any, class preparation. Two teachers (TB2 and TC1) began with the introductory

exercises covering basics such as vowel distinction (TB2: 704, TL1: 480) which were found in their chosen text, the *Cambridge Preparation for TOEFL* (Gear & Gear, 2002) Other teachers followed the books less religiously: TL1, for example, reported picking and choosing exercises according to his understanding of what was likely to come up on the test (TL1: 480), although he admitted that his primary source of information was accounts of the test given by his former students. Other teachers and students mentioned using exercises that were not from their own TOEFL preparation books but from other TOEFL preparation books available at their institutions. The materials in the most widely used books offer some strategies to facilitate students' ability to comprehend spoken language (such as focusing attention on the conversation, concentrating on the text and the details, concentrating on the purpose, and so on. (Gear & Gear, 2002, p. 74), but they mainly offer exercises which mirror the types of task on the TOEFL.

Teachers also made use of practice listening tests, which were available in the preparation books, in ETS packages (e.g., POWERPREP), and from the Internet. TC2 reported that he only used mock TOEFL tests for listening practice (TC2: 1053). Teachers rarely used activities from more general EFL materials.

When teachers prepared supplementary material they tended to focus on vocabulary, especially the differences between British and American English, the use of synonyms, and idiomatic expressions. Several teachers reported that their students had problems with American usage (e.g., TB1: 357, TP1A: 485) since the majority had been exposed to British English during their schooling. The most common way of dealing with the differences in variety was for the teacher to do a quick translation. It was necessary though, in some cases, to explain new concepts or a new culture:

> TC2: 436–439: So I find that it's important to explain the American college system to them first. So they have a little picture in their head, and sometimes drawing on the board, to tell them what a dorm is, what a quad is and such like, because they don't understand that.

Synonyms were also felt to be worth studying, both for the Listening and the Reading tests (TP1B: 988), as were idiomatic expressions. TP1B gave her students lists of idioms to memorize, which she believed occurred frequently in the Listening test. Her source for this belief was another TOEFL preparation book (Duffy & Mahnke, 1996; TP1B: 232; see also Sharpe,

2001, p. 46). TL1 also felt it was necessary to have a firm grasp of American idioms and alluded to the difficulties this posed for students who had not experienced living in the United States:

> TL1: 1493–1498: But it's almost to the point where they're testing have you spent time in America . . . I've noticed a lot of my students have done Work Travel USA or something like that and say "Oh, I know that," and I say, "How do you know that?" "Oh, I lived in Georgia for 2 years," and the other ones have been sitting in grammar classes or in school and they just don't have the opportunity.

How Is Listening Taught?

The interviews and observations suggested that the majority of the teachers dealt with listening in practice test mode: The students listened to a passage as a group—generally not viewing questions in advance (this will be discussed), read the questions and selected the answers individually, checked their answers as a group under the teacher's control, and finally discussed any difficulties or new vocabulary arising from the task. What variations there were tended to involve more intensive listening. TB2, for example, said that she might play the tape several times (TB2: 651), stopping where students encountered problems (TB2: 709) and providing grammatical explanations where she felt it necessary to help clarify an area of difficulty. TL1 also played the tape more than once, asking the students to concentrate on the general idea the first time round and specific details when the passage was repeated, thus attempting to build up the students' comprehension in steps (TL1: 726). He also asked questions that were different from those in his preparation book, to get students to think about the context of the listening task (TL1: 507, see also TP1B: 66). This method of dealing with passages was not widely mentioned, although is a basic technique of teaching listening in more general English classes.

TS1 stood out from the other teachers, torn between a need to provide his students with the necessary experience to face the Listening test and a desire to create a more communication-oriented classroom. When asked whether he followed the TOEFL preparation book for listening he replied that "the TOEFL exercises were good but they were very concentrated and pretty boring for them" (TS1: 489). He introduced a range of other activities, including work on speeches by famous Americans such as Martin Luther King, Jr. Students received a copy of the text, which they studied at home to ensure they were familiar with all the vocabulary. They

listened to the speech during the next lesson and proceeded to hold discussions or debates around the issues arising from the text. TS1 felt that getting the students to listen to one another was as valuable as getting them to listen to the two usual sources of input (the teacher's voice and recorded native-speaker EFL materials):

> TS1: 469–472: They will go into opinion building and debating and then feeling confident expressing an opinion and listening to one and arguing with someone in a productive setting.

Whether this type of listening practice is appropriate for helping students the current TOEFL is open to debate but this teacher had a clearly stated agenda of not teaching purely to the test:

> TS1: 113: The TOEFL is not the primary goal of this class.

Problems in Teaching Listening

The major methodological challenge for the teachers was to find a way of helping students who could not distinguish between the sounds they were hearing, despite having adequate vocabulary and command of the structures. One teacher expressed the problem in this way:

> TB2: 750–751: Sometimes they can hear it, but they don't understand. They have to read it, to see, because of the intonation, the pronunciation.

Her perception was that pronunciation got in the way of comprehending, but several students indicated that their problem was one of "getting used to rapid listening" (SS1: 756). One student added that

> SS1: 759–761 (see also SB2: 253): . . . after I can see it written down then it's pretty clear to me. I hardly can believe that I couldn't understand it before but then first time is always problem.

It was not clear, however, that any of the teachers had techniques to help students cope with unpredictable input at speed. In fact, the only strategy mentioned that was of direct relevance to the Listening test alone was the advice not to linger on a question once the tape

moved on to the next question (TP1B: 950, SP1B: 78). This advice as given by TP1A and TP1B appears to be directly lifted from the preparation materials since they both quote the exact length of time available to answer each question (12 seconds).

Attitudes Towards the Listening Test

So far this report has discussed some of the findings concerning the what and how of classes devoted to listening but just as interesting were the insights into the attitudes towards this skill. Of particular interest is the belief several teachers put forward that all the students can do to prepare for the Listening test is practice. This made listening the easiest skill to teach for some (TB2: 1990, TC2: 730, and TR1: 1405) since their students got plenty of practice outside the classroom:

> TR1: 1407–1409 (see also TL1: 739): . . . because you know they are very much accustomed to listening to films and music and so on in English so they can cope quite well with the listening part.

TC1 questioned whether the English students were exposed to through popular culture was really of the right nature and caliber, considering the tasks they were expected to do on the TOEFL (TC1: 425). Most of the other teachers seemed to agree with TR1, however, offering plenty of in-class practice (preparation book exercises and mock tests) but also encouraging the students to get as much extra listening practice as they could outside the classroom. They frequently suggested listening to CNN, Voice of America, or any other English-language radio and television program. (TB1: 942, TB2: 1176, TB2: 1397, TL1: 1078, TL2: 392, TL2: 1291, SB2: 300). Similar advice is also offered by the Barron's preparation book, though it specifies documentaries, especially those from educational broadcasting networks (Sharpe, 2001, p. 52). TP1B even had tapes copied for the students so they could do extra practice at home (TP1B:060), and other students demonstrated their commitment to dedicating substantial amounts of their time to test preparation by saying that they would have liked their teacher to do the same (SR1: 865). A number of students confirmed that they did get significant amounts of general listening practice outside class (TC2: 1064, SB2: 977, SL2: 356, SL2: 808, SP1B: 113, SP1B: 1146, SS1: 243) but there was a wish to have more listening in class time as well, especially nearer the time of the actual test (SS1: 730). One teacher (TC1) expressed her view

that since the listening skills were those neglected most at school level the students simply could not do enough practice in class in order to achieve a good standard (TC1: 405).

Students were as divided as the teachers in their opinions regarding whether listening was an easy or challenging part of TOEFL, but they agreed about the importance of familiarity with English and opportunities for extended exposure:

> SP1B: 1654–1658 (see also SP1B: 164, SS1: 1341): I think it's because I haven't had a lot of occasion to listen to American English. For example when you talk to me I understand you because you speak slowly and some of these exercises which we have the people talk too fast and I can't understand everything and that's it.

This contrasts with:

> SP1B: 1689–1692: . . . I think just because I had contact with the American English it's easier for me . . . I have to focus a little . . . but I don't have problems listening.

Three of the teachers who felt listening was the most difficult test to prepare for said they spent more time on this skill than on any other. However, TC2 also spent the most time on listening even though he felt it was amongst the easiest tests to prepare for. TB2 felt listening demanded so much time that she relegated most of the reading and writing exercises to homework (TB2: 675). TL2, as already seen, simply asked her students to practice listening at home. There is clearly little consensus on how this skill is perceived or on how best to approach it.

Issues Arising

There was some concern amongst teachers about the relevance of some of the testing tasks (and preparation material) to students' target language situation, and about the demands that the current TOEFL test made on memory. There were three types of comments relating to the relevance of the tasks to the students' target language situation. The first came from a teacher who had just begun teaching TOEFL preparation who did not understand the relationship between the dialogues presented in her preparation book (which she accepted as representing tasks on the actual test) and the types of listening students would be doing once they were in higher education in their chosen country:

TP1B: 587–594: Since we've been doing listening comprehension we've been doing short dialogues and I don't really know how to relate that directly to something they're going to be doing in an academic environment in the United States. Because if they go to the United States or to Canada they're going to be listening and hearing English all the time but I don't think it's going to make a difference as far as what we've been doing for listening comprehension.

She implied that the material on the test was not authentic and/or relevant enough to be of use to the students wishing to study in North America. Her doubt may have arisen because she had not worked her way though all the listening material, but it showed that she was thinking about the link between the tasks on the test paper and what the students would have to face in the real world. Other teachers seemed to believe that preparing for the Listening test would help the students to get by in the target situation, at least in terms of getting used to pronunciation.

TB1: 1467–1470: . . . the listening section is a good feature because they can learn. They can hear different accents first of all. They can hear people talking and probably something will help them in their studies because everybody has that idea.

There were also comments from students on the relevance of the test content.

SB1: 216–219: I only thought that they would prepare us for the exam but then when I got in to it, started listening . . . I found out that the conversations are mostly about things that happen on the campus and stuff like this.

What none of this takes into account, however, is the fact that not all the students preparing for the test were aiming to study in North America (only 43% mentioned North America as their goal, while others mentioned Europe, Australia, and New Zealand). This brings into question the appropriateness of using mainly North American voices and culture-specific vocabulary in the Listening test.

The second issue had to do with the fact that the demands the Listening test (CBT version) made on the students' memory. Whereas listening in real life would generally be motivated by a purpose, listening on the CBT test students involved hearing a text before being given the questions that were to be answered. This was not necessarily a problem in the early parts of the

test, during the dialogues mentioned by TP1B above, but it did cause some concern in the later parts:

> TL2: 1051–1054: I don't know how they are doing the listening, especially the long lecture parts, because you are not allowed to see the questions before you answer them.

The students were also not allowed to take notes, which clearly contradicted the target language situation:

> TP1B: 596–601: They could understand but they couldn't remember and I told them that, because on the TOEFL in the listening comprehension parts you're not allowed to take notes but if they're listening to a lecture at university, they're going to be allowed to take notes so it's not going to be a problem for them.

There were a number of comments from students about their frustration with this aspect of the test method (SB1: 1560, SL2: 977, SP1A: 1388, SS1: 1833), including some that emphasized the negative consequences of momentary lapses of attention:

> Student 1:   Sometimes it can make some misunderstandings . . . because I really feel like I understand but if there is some, for example, noise or . . . the listening is about a hundred percent concentration and sometimes just an idea comes to my mind and I am disturbed.
>
> Interviewer: And it's gone.
>
> Student 1:   And it can really affect the results.
>
> Student 2:   And actually one dialogue and five questions so if you misheard once you can lose five points. (SS1: 1016–1027)

One of the skills needed for the TOEFL, mentioned by a couple of teachers and students, (TB1: 1231, TC1: 352, TL2: 1238, TP1A: 1109, SP1B: 1481) was stamina. Many students had not experienced such long testing sessions before, and developing the ability to concentrate for a long period of time seemed as important as familiarization with the test and brushing up language skills.

Summary

The major point arising from the interviews with teachers and from observations carried out in their classes was that although they had a general idea of what the listening test was meant to be assessing, they lack techniques for breaking the listening skill down into subskills to facilitate mastery. There was a gap between knowing what would be tested and knowing exactly how to prepare the students for the challenges facing them. This may be partly due to lack of advice given to teachers in the standard TOEFL preparation materials, which seem mainly to advocate giving the students practice. The teachers seem to rely on the preparation books to a great extent, so this lack of guidance transfers to a lack of clear direction in the classroom. Certain techniques that are used in general language classes (such as activating schemata, using pre-listening questions and note taking) and that some teachers seem to be familiar with, are not utilized in the TOEFL classes, probably because they do not seem relevant to the format of the test, which presents questions only after the students have heard the text. In light of this the teachers seem to rely primarily on a combination of exercises that mimic the test format and practice tests, which may build up the students' abilities and strategies eventually but in what seems to be a hit-and-miss rather than a systematic way. T7 says what the other teachers did not state explicitly and sums up the attitude towards teaching preparation for this section:

T4: 1011–1012: I think they need something else, but don't know what.

Teaching Structure

TB2: 154–158: I don't have much time with the TOEFL class to explain a lot of grammar. They are supposed to know it already, at least at upper intermediate level. But what I find is that sometimes they have problems. Like certain grammar. Then I have to take time to go through some grammar.

DOSP1: 805–808: Our teachers pick out the weakest points of the students and then go over it. They don't go over the entire grammar or they don't go page by page of the book. They just do testing, test, test, test, test.

This section examines the ways in which the teaching of structure takes place in the TOEFL classroom. It looks at issues such as the proportion of time dedicated to dealing with the Structure section of the test, what the teachers teach and how they teach it.

The Role of Structure Teaching in Class

Students are assumed to have a certain level of language knowledge if they are preparing for the TOEFL. Most of the students in our study had successfully completed previous levels at their schools or had taken a placement test that indicated they had enough language knowledge to be able to function in the TOEFL class. Nevertheless, all of the teachers said they had to pay at least some attention to grammar. Several said grammar was what they spent most time on (TB1: 381–383, TR1: 459, TL1: 404–406—at the beginning of the course; TL2: 365–367—at the end of the course) and this was corroborated by their students (SB1, SL1, SL2, SR1). (See also Appendix O.) Observations showed that four teachers covered quite a few grammar points in their lessons (see the Structure section of Appendix M). In contrast, two teachers claimed to spend the least time on this area (TP1B: 741–742, TS1: 438–439) as they felt it was not a problem for their students. It is therefore difficult to make generalizations about how teachers perceive the importance of grammar for success on the TOEFL and how much attention they pay to it in their classrooms.

What Is Taught?

When teachers spoke about what or how much grammar they taught it was mainly in terms of reviewing structures that their students were assumed to have acquired elsewhere (TC1: 832–833, TP1B: 362). The students had spent many years learning the language and their experience had probably been of grammar instruction rather than language skills development (TP1A: 469, TS1: 362). The students generally agreed that they had a sound basis in grammar:

> SS1: 1167–1169 (see also SP1A: 1281, SP1B: 1560): We have learned (it) at university, in grammar school, at elementary school. It was almost all the time about grammar rules.

They thus arrive in TOEFL classes with a good level of knowledge (TC2: 477), though they are not necessarily confident in their ability to communicate:

43

SP1A: 1767–1770 (see also SB1: 1618, TB1: 927): This is the way [schools in our country] teach and we don't like it because when you graduate from secondary school you know all the grammar and some people know it perfectly but they cannot pronounce a single sentence and this is a problem.

TL1 felt that it was possible and necessary to take them to a higher level during the TOEFL preparation course. Amongst the language features he got his students to work on were prepositional phrases and articles (TL1: 436). He and another teacher from the same country (TL2: 1340) felt that articles were problematic for students from that particular language background. This was interesting, as the same problem would have applied to most of the countries in the sample, as they shared common language origins. TL1 believed that some of the grammar on the TOEFL was too hard for his students:

TL1: 683–697: For some reason in all the normal grammar classes that I've taught we've done the defining relative clauses and the nondefining relative clauses but TOEFL goes really deeply into these reduced adverb clauses and they're terrifying for the students because they're difficult.

He also felt some of it was deliberately "tricky" (e.g., inverted subjects and verbs, TL1: 1425–1428). In contrast, another teacher reported that her students did not always spot errors in the error identification tasks because they were too easy (TB1: 388). Her students corroborated this, reporting that her advice on these exercises was to "translate, translate, translate" so they could focus in on where the error might be (SB1: 709).

How Is Grammar Taught?
The teaching of grammar mainly involved teacher explanations of individual structures (TL2: 366, TP1B: 343). There was no evidence in the observations of any task-based or discovery activities. TL1 and TC2 mentioned that they used drills with their classes (TC2: 339, TL1: 691) but the majority of the work on structure was through exercises mimicking the TOEFL or on practice tests. Structure was also dealt with through feedback on essays. Few of the teachers used the TOEFL rating scale for essays, preferring to give their students individual feedback that included the correction of grammatical mistakes or errors (TP1A: 643). (See also these sections in this report: Teaching Writing and Assessment in the Classroom.)

Two of the expatriate teachers admitted to having problems explaining grammar to their students (TP1A: 1130–1133; TP1B: 272, 474, and 511), neither having come from a language or linguistics background and both having very limited English teaching experience to date. TP1B went so far as to say that she believed that it would be better for her students if they had a teacher who was not a native-speaker of English:

> TP1B: 511–516: [Teachers from this country] can probably help explain the grammar a little bit better because their first language is not English and I think that's the way TOEFL should be taught. I mean my students obviously know that I've never taken the TOEFL and that even if I did it wouldn't be the same as they taking the TOEFL because English isn't a foreign language for me.

If other teachers faced problems with grammar explanations they did not generally admit it, apart from a few cases (e.g., TL2 found it hard to explain the use of articles [TL2:1340]). In light of this though she managed to pass on very pragmatic advice to her students: If you really can't find an error in the error detection exercises it probably has to do with an article (SL2: 469).

How far did teachers include grammar because they truly felt it was needed for TOEFL and how far was it a part of their standard language teaching repertoire? We were not in a position to observe non-TOEFL classes for all the teachers, as has already been explained in second section. However, our observations suggest that the teachers' approach to dealing with grammar in the classroom did not differ greatly between TOEFL and non-TOEFL classes.

Attitudes Towards the Structure Test

The majority of the teachers were in agreement that the ability to use grammar correctly was tested on TOEFL (see Appendix N). Not all of the students were of the opinion that grammar was vital to TOEFL success (SS1: 1504, SS1: 1653), but the majority believed they could not afford to neglect it (SB1: 2041, SB2: 1826, SC1: 464, SL1: 1505, SP1A: 1542, SR1: 1681). There was a mixture of opinion about the difficulty of the Structure test. TL1 felt it would be hard for the students (TL1: 693), while TL2, the only teacher in the sample who had actually taken the TOEFL, felt it would be easy (TL2: 1087).

Why should there have been differing opinions about how important grammar was to TOEFL success, and whether it was difficult or easy? One possible reason is that the construct

underlying the Structure test is not very clear, or that the way in which it is communicated does not make sense to the teachers or, therefore, to the students.

Summary

Most teachers expected students to have achieved a certain level of English proficiency, given the heavy grammar-translation orientation in schools in the CEE region. In general the students had studied a lot of grammar but, as the evidence from a variety of classes indicated, there were differences in their ability to use this knowledge effectively. The teachers' approach to dealing with grammar mainly involved explanations of language points, which some of the expatriate teachers seemed to find quite challenging, and the use of test-like grammar exercises. There were no task-based or discovery activities in the classes observed.

Teaching Reading

TC1: 629–634: There are four types of questions that you should consider in the Reading section. . . . You have to recognize if this is, for example, a main idea question or inference question and . . . [you] should be told during the course that when you have this type of question this is the way you look for an answer . . .

TP1B: 792–795: That was the biggest problem with the reading comprehension. Just individual vocabulary words and it was just because they were words they hadn't learned in English.

In general, the teachers in this study seemed confident when talking about preparing their students for the reading test, at least in terms of what they needed to teach if not how to go about doing it. This section examines a range of approaches to preparing students for the Reading section of the TOEFL, from those teachers who mostly consigned it to homework to those who felt a large proportion of class time should be dedicated to it. This section discusses what the teachers believed was being assessed, what they taught, and how they went about their teaching.

What Is Tested?

All of the teachers believed that TOEFL assessed the ability to understand a wide range of texts (though this was in fact questioned by the ETS representative who presented the test designers' point of view—see the table in Appendix N, the row labeled D, column labeled ETS

answer), and most believed that it assessed the abilities to understand the organization of texts, make inferences, and understand unfamiliar vocabulary in context. They were also convinced that developing a wide vocabulary was a key to success on this section (TB2: 786, TL1: 452, TP1A: 499, and TS1: 1648, among others). This point is discussed in more detail in the section on teaching vocabulary.

What Is Taught?

All but one of the teachers (and many of the students) reported that they mainly followed their TOEFL preparation book for developing reading skills (see, for example, TC1: 622, TL1: 516). This entailed working their way through sets of exercises that simulated the types of questions found in the test (TP1A: 488, SR1: 306, SS1: 383). Since only one of the teachers had firsthand experience of taking the TOEFL (the paper-based version), most had to rely on feedback from their former students to gauge how closely the preparation books matched what actually appeared on the TOEFL. Most teachers were under the impression that the passages offered were similar to those on the test (TB1: 462, TB2: 804, TL1: 522, TP1A: 491, TR1: 532), both in terms of length and topic. Not all the teachers were of this opinion, however: TL2, the teacher who had taken the TOEFL, thought the passages in the practice materials were shorter than the texts used on the real test (TL2: 429). She also felt the topics did not match those in the test itself in that the topics in the preparation books were more likely to be related to American history or culture (TL2: 441–446).

The teachers and students mentioned a variety of topics when asked about those most frequently found in their practice materials. The full range can be seen in Table 10. It is interesting that scientific topics prevail. This reflects the advice offered by at least one preparation book, *Cambridge Preparation for the TOEFL Test* (Gear & Gear, 2002), which suggests that science and technology predominate, followed by social sciences, with the humanities coming in third (p. 285). Not many comments were made about the topics themselves, other than that they were more academic than the First Certificate in English (FCE) (TP1A: 355), that there seemed to be a disproportionate number of texts about agriculture (TL1: 560) and that the students had more trouble with scientific texts because of the specialized vocabulary (TB2: 800). Opinion was divided about whether the texts in the preparation books were "dry" (TP1B: 778) or interesting (TC1: 447–453).

Table 10

Topics Appearing Frequently in TOEFL Preparation Coursebooks According to Teachers and Students

Topic	Transcript reference
General academic	(TL1: 533)
Agriculture	(TL1: 533)
Archaeology	(TL1: 533)
Astrology	(TR1: 563)
Astronomy	(SL2: 273)
Biology	(TB2: 776) (TR1: 563) (SP1A: 300) (SP1B: 424)
Business	(SL2: 273) (SP1A: 300)
Chemistry	(TR1: 563)
Environment/nature	(TB2: 776) (SL2: 273)
Geology	(TR1: 563)
History	(SL2: 273) (SP1B: 424) (SR1: 337)
Law	(SP1B: 424)
Literature	(TP1A: 506) (SR1: 337)
Medicine	(SP1B: 424)
Physics	(SP1B: 424) (SR1: 337)
Science	(TB2: 776) (TP1A: 506)
Social and health problems	(SP1B: 424)
Sociology	(TL1: 533)

There was more agreement about what the Reading test tested than there was for other tests. Most of the teachers felt that understanding text organization was important, with several noting that this was a skill that could be exploited when preparing for the Writing test. Other frequently mentioned subskills were guessing words in context (TB2: 578, TP1A: 358, TP1B: 810, TS1: 892, SP1B: 402), making inferences (TP1A: 358, SS1: 436), and referencing (SP1A: 268, SP1A: 537). Identifying the main topic of a text was also mentioned (TP1A: 599, SS1: 233), and it was believed to that this required skimming skills (TS1: 895).

Although test-simulation type exercises prevailed in the classroom, some teachers (e.g., TS1) introduced other activities that they thought would help their students manage the Reading

test better, including exercises focusing on vocabulary development (TS1: 517). Particularly popular was working on synonyms (TP1B: 988, SP1A: 798), the reasons for which are discussed in the Teaching Vocabulary section in this report.

TS1 was one of the few teachers who used passages that did not mimic those on the TOEFL. In the lesson observed, for example, he used a humorous text:

> TS1: 542–548: We are also starting going through things that to them are funny, where they look for mistakes in the meaning. For example . . . a list of classified ads where people didn't say it quite correctly. They're starting to look through those and find where the humor is and where the mistake is and where would you change it.

It could be argued that error correction is tested in the Structure section, but in these texts the errors were more semantic in nature than syntactic.

TC2 also stood out from the group by asking his students to complete ordinary TOEFL-like reading exercises at home and using class time to discuss other types of texts (e.g., literature):

> TC2: 1069–1074: At first I spent 40 minutes or so reading with them but I decided, no, it's better if they just do it at home and then I'll just check the answers and bring them some new materials. For example they would get Dostoevsky's stories to read because they are more challenging, and we're talking about them, talking about it, discussing it and sometimes write about it too.

The teachers and students mentioned little in the way of test-taking techniques for reading. One teacher tried to get her students to look at "how questions are arranged" (TL2: 380), meaning analyzing what type of question each was and deciding how best to tackle it. Two other teachers alerted their students in a similar way (TC1: 628, TP1B: 775), the latter going so far as to show them what she referred to as "tricks" for finding the answers. TL2 seemed to believe that there were common features across some of the ETS standardized tests. She claimed there were skills she had developed for tackling the Reading test that she had learned from also teaching GMAT and SAT preparation classes (TL2: 431).

How Was Reading Taught?

The manner in which the reading classes were conducted was relatively uniform: Students worked individually through the exercises in their preparation books or practice tests and then joined together in plenary to check their responses and discuss any issues arising (TC2: 1109, TP1B: 684). The reading work was primarily teacher-led and apart from the discussions of humor and literature mentioned earlier there were no activities that might be seen in a communication-oriented classroom (e.g., prereading tasks or jigsaw reading, which can be applied even in high-level EAP-type classes). One teacher had only recently stopped asking his students to practice reading by reading aloud in the lesson (TC2: 1106). Most teachers did not see lockstep teaching as problematic, but one seemed to be quite frustrated by what he felt was a need to use a teacher-centered method (our term, not his):

> TL1: 461–464: I really hate doing this. I hate teaching in a way where I have to say "Right, okay, learn this!" I like . . . you know . . . "Let's do something interactive. Let's do something interesting. Let's do something fun."

Teachers might have agreed on what the TOEFL tested but they did not agree on how much class time needed to be dedicated to the Reading section for students to be successful. The teacher and students at one site agreed that reading definitely needed more time than the other sections of TOEFL (TB1: 381, SB1: 248). However, another teacher stated that she was confident the students could manage the Reading section reasonably easily (TR1: 462- 480: "They can do one piece of reading in 12 minutes, which is the right time") and so did not require more class time. TP1A felt that students could probably do the work on their own at home, but it may not have occurred to him to ask them to, as he was new in his teaching institution and may not have been sure of what he was allowed to do.

As mentioned earlier, some teachers were quite willing to relegate reading to homework (TB2: 672, TC2: 1067, SL2: 234, SS1: 223, SS1: 371), while others used shorter texts in class and assigned longer texts for out-of-class reading (TS1: 534). This must have seemed a sensible decision to them, as reading is an individual and normally solitary activity, but it also meant that the teacher could not control the timing of the exercises and thereby see whether the students needed to be pushed to read more quickly in preparation for the test. However, no mention was

made of the students' need to read quickly (except for the reference to "12 minutes" given earlier), beyond when students worked through mock tests, generally near the end of the course.

All the teachers claimed to encourage their students to do as much extra reading as possible, and suggested the types of reading that would be most useful and where material was available, such as the local American Center or British Council Library (TL1: 535–548, TS1: 1071, SS1: 519–522).

Attitudes Towards the Reading Test

Three teachers mentioned that they felt the Reading test was the most difficult to prepare students for. TB2 felt that the problems students had were with technical vocabulary:

> TB2: 786–790: This is what makes it difficult for them . . . when it's something with specific terminology, like say, if it's a text about the sea level, or something which they are not interested in, or they don't know the words which go together with such topics.

TB1 had problems trying to demonstrate to students that what they felt was the main idea of a text was not necessarily right, or the only possible answer. She said they "used their own logic" (TB1: 1486–1493) and she found it hard to convince them there was another way of viewing the text. TP1B, a native-speaker of English, claimed that she herself had problems answering some of the questions in the commercial preparation materials she was using. She believed that this was the fault of the test and not because she had reading problems:

> TP1B: 1367–1375: . . . I've had trouble answering the questions just because of the way the texts are written and then sometimes the way the questions are worded. You can't teach what you don't know and I can't explain why this is the answer because I don't necessarily understand why it should be the answer. It's not the case with every question but it's the case with a lot of them. You don't understand—how did they get that answer? I don't see that on the text and they don't see it in a text either, or they don't infer it from a text either.

Note that it was not uncommon for teachers to equate what they found in commercial test preparation materials with the content of the TOEFL test itself, rightly or wrongly. This is discussed further in the Role of the Coursebook section.

The views of TB1 and TP1B are reminiscent of the arguments put forth by Peirce (1992) regarding whose interpretation of a text should be considered legitimate and whose not. Peirce's position is that in testing with high-stakes decisions, the test takers must learn how to adapt in order to survive.

The Relationship Between Reading and Other Skills

It was in the area of reading that there were the most references to the ways that work in language and skill areas might complement one another. However, these were not references to the notion of *integrated skills*, which will be incorporated into the new TOEFL, whereby knowledge acquired by successfully processing material in one section of the test is drawn upon for another section. What was being referred to here was more a matter of how skills developed specifically in one area might prove useful in some way for another, at least from the teacher's point of view.

The most obvious areas where this combining took place were reading and vocabulary. There is no longer a separate vocabulary test but a good command of vocabulary was generally deemed to underpin the ability to do well on all the sections of TOEFL, including reading (TL1: 459–461). It was mentioned that vocabulary caused the most problems for students in this section (TP1B: 793, S6: 454, SP1B: 393, SS1: 230), but it was also believed that reading widely was one of the main sources of vocabulary improvement (TB1: 445–448, TP1A: 355–363).

Reading and writing were covered together by TP1B, and there seemed to be some kind of overlap for other teachers as well. Reading and writing were felt to be mutually beneficial, with reading providing examples and models that the students could then put into practice (TP1B: 730). TL1 recognized the benefits of strengthening the students' ability to recognize the structure of a text so that they could feed that into organization of their own texts (TL1: 1577).

Finally, reading was used to help students to improve their listening. This was done in two ways: first by giving students copies of the passage before they listened to it so that they could concentrate on features other than general meaning (TS1: 495) and second, by giving students a transcript after the first listening, to help them to understand the parts of the passage where they had problems (TB2: 749–752, SB2: 253). Although it could be argued that relying on a transcript is not the best way to build up students' confidence in their ability to identify the gist or the main ideas of a passage, there may be some point in the teachers providing the text for reading if they are convinced that individual words and details are important.

Summary

With one exception, most of the teachers used texts that reflected, more or less, the text types and topics found on the TOEFL. Exercises to practice subskills such as skimming, scanning, referencing, and inferencing were the most commonly observed (and discussed in the interviews). Subskills were practiced but there was an underlying belief that an extensive vocabulary was as necessary as such skills to be successful. This was the focus of several of the lessons observed (TS1, TR1, and TL1). For this reason the next section will deal with the issue of vocabulary even though it is no longer tested separately on the TOEFL test.

Teaching Vocabulary

TC2: 535: I believe a rich vocabulary will help you get a good score on the TOEFL.

SP1B: 1120–1121: I know I should work on my vocabulary because it's pretty essential to pass the TOEFL test really good.

What these quotes (the first from a teacher and the second from a student) highlight is the fundamental belief on the part of most of the teachers and students in the study that a firm knowledge of vocabulary was one of the keys to success on the TOEFL test. Although vocabulary has not been tested separately since 1995 (Read, 2000, p. 144), the importance given to it by the informants justifies discussing it separately rather than subsuming it in any of the sections devoted to separate tests on the TOEFL test.

First, this section discusses the approaches the teachers used to teach vocabulary and then examines their attitudes towards the way vocabulary is tested. This section also presents several issues that the teachers raised in the discussions about this area of language.

How Is Vocabulary Taught?

There seemed to be two main approaches used to develop the students' vocabulary. The first was to distribute lists of words and phrases to be learned by heart. In the case of TL1, these lists contained what he considered to be the key vocabulary from each unit of his test-preparation book (TL1: 380). He reported mixed reactions to this from his students:

TL1: 394–397: When I give them the sheets with the new words on, their faces just fall; but, afterwards they do thank me. They're very happy because, you know, those words could be on one of the tests.

Other teachers felt lists of synonyms were important (TL1: 1770, TR1: 676), and TR1 felt connectors were crucial, especially for use in writing essays:

TR1: 920–924 (see also SB1: 372): I give them lists of words which they can use in writing: in my opinion, to my mind, I for one, as far as I am concerned, as for me, well, to express this or to express that, [and] at the conclusion. If you noticed one girl, she said "needless to say." This came from one of the lists and I was happy she had used it.

Another teacher introduced words that were "quite rare" (TC2: 1085). He did not explain how he chose these words, saying only that he and his students enjoy "playing" with vocabulary.

The second approach to teaching vocabulary was via reading, as explained by TS1:

TS1: 522–525 (see also TB1: 445 and TB2: 685): We don't just do vocabulary on its own any more. The vocabulary is tied into a greater context of things so they're also learning how to discover words on their own through contextual reading.

TC1: 616: In essence the approach involved students finding new vocabulary in context and either discussing it in the class or noting it down and returning to it at home, using a dictionary. There was nothing complex in this method and it seemed to be a natural vocabulary development procedure. One teacher mentioned that she exploited listening passages in the same way.

TS1, however, often asked his students to prepare passages in advance of the lesson, so that they would already be familiar with new vocabulary (and other language points) before he used the passages for listening comprehension or discussion purposes. He gave an example of this way of working:

TS1: 492–499: We would do something like Martin Luther King, Jr.'s *I Have a Dream* speech—very colorful, a lot of vocabulary, things they haven't seen before. So they have a week to pour through it and understand the vocabulary and have some time to look at

the context and be able to understand the words. They've read the words. They've probably said them to themselves. We listen to the tape so they're getting all of it at the same time and then we hold discussions afterwards.

Notes taken during an observation of one of the discussions he referred to revealed that he constantly elicited synonyms as the students expressed their reactions to what they had understood. The teacher said that he did not believe in memorization as it was not "an effective tool for education in general," so the importance of vocabulary on TOEFL was forcing him to "be a lot more creative" in "having them deal and work with these words" (TS1: 1831–1834). What was not clear, however, was what exactly the students did to learn to deal with and to work with words.

Other teachers mentioned working on exercises from their chosen preparation book, specifically, focusing on inferencing and working out meaning from context (TP1A: 357, TS1: 523). They did not mention any other sort of activity, which suggests that there is either a lack of expertise in techniques for teaching vocabulary or that time pressure precludes the use of such techniques.

The students interviewed claimed to rely on memorization to a great extent, including learning lists of idioms (SP1A: 1284, SP1B: 1542, SP1B: 1572). It is not clear whether they used memorization on the advice of their teachers (clearly not TS1 though), or because of habits they learned in their schooling, or in response to advice from other sources. Students said they referred to a variety of Web sites, looking for extra practice materials, and most of these sites offered tips and advice about TOEFL.

Another method that was cited was the use of computer packages designed to help increase vocabulary size (SS1: 1147). One package enabled students to hear the pronunciation of new words, an aspect of vocabulary that is missing from paper-based word lists (SS1: 1150). Although students currently do not have to pronounce words for a TOEFL speaking test, they do need to recognize them in listening passages. This development seems particularly useful then, given the notorious difficulties of written/spoken form relationships in English.

Students said they did lot of extra reading outside class hours (SL2: 242, SP1A: 869). They sometimes worked on tasks assigned by the teacher, but several mentioned reading for pleasure and noting down new lexis that they discovered in the process. One student mentioned expanding his vocabulary by maintaining correspondence with an English-speaking friend.

Overall the picture seems to be that the onus is on students to widen their vocabulary during out-of-class hours. This could be justified since every student has their own internal dictionary, but it could also be seen as neglect of duty on the teachers' part, since they are trying to impress upon students how important it is to enrich their stock of words while not really doing much to help them, other than distributing lists of words to learn and encouraging extra reading.

TS1 clearly felt that vocabulary played such an important role for the students that he was at a loss regarding how best to cover everything he felt they needed (TS1: 1720). He felt they needed a structured approach (e.g., studying suffixes) but he was not confident he could deal with this effectively since he, a native-speaker of English, had not studied this way himself. He may have exposed a gap between what the teachers felt their students needed and their own ability to provide it, that other teachers were unaware of or did not voice.

Attitudes Towards Vocabulary

There was a belief that students needed not only a good basic stock of words but also a high level of vocabulary to do well on TOEFL (TR1: 73, TS1: 441, TS1: 1044). Formal and specialized vocabulary were also felt to be necessary (TR1: 1341, TB2: 786). Where do these views come from? Alderson (2000, p. 105) refers to work by Hale (1988), which suggests that ". . . reading passages used in TOEFL are taken from general readings rather than specialized textbooks, which are intended to be relatively nontechnical and understandable by a general audience."

None of the TOEFL preparation books mentioned any rationale for the vocabulary they presented to students (e.g., no mention of any indicator such as the Academic Word List: http://www.vuw.ac.nz/lals/research/awl/). The new TOEFL has been advertised as being informed by the Spoken and Written Academic Language Corpus (SWAL), so it may be that teachers in the future will have a better understanding of which vocabulary to teach and why — as long as this information is disseminated widely and in terms that teachers can understand.

There are two possible sources for the beliefs both teachers and students held that vocabulary was important and that synonyms in particular were crucial. The first is a possible hangover from when TOEFL included a separate vocabulary test. Teachers do not have easy access to TOEFL teacher training courses (indeed, few seem to exist, see the Teacher Training section) and most devise what they deem to be the most appropriate methodology by drawing on their general teaching experience, their preparation books, and discussions with other teachers

who may have taught or know something about TOEFL (see the Communication section for further details). Thus a belief, even if incorrect, may get passed on from teacher to teacher, and then on to students in the classroom. Another influence on teachers' perceptions and, inevitably, on students' perceptions, might be the verbal descriptors associated with the score points on the scoring rubric for the current Writing test. Four criteria are mentioned at each score point (apart from the lowest one), and one of these covers grammar and vocabulary. The scoring for the Writing test is holistic but teachers might still believe that a quarter of the marks assigned at any level will depend on the quality of the candidate's grammar and vocabulary.

Problems in Teaching Vocabulary

Some issues were raised by teachers, particularly the native speakers of English, about the nature of some of the preparation tasks involving vocabulary. They first made reference to exercises that were supposed to require the guessing of meaning from context. They argued that in some cases not enough context was supplied to allow guessing. This led to a feeling that it was necessary to expand the students' range of synonyms so that they could cope with such questions. As one teacher explained:

> TL1: 455–461: On the TOEFL test they give you a word in the text and the question is which of these four words is closest in meaning to the word. If they don't know that word, these four words are absolutely meaningless so the vocabulary will kill them on the reading section and could kill them on the listening. It could kill them on the structure and that's why vocabulary is so important.

Other teachers felt that the vocabulary items included in TOEFL were not always representative of current language:

> TS1: 1735–1736: There are a lot of words and phrases, which they use in ways which they (students) may never ever hear or ever use.

The teachers wondered why students were expected to recognize what they thought were low-frequency items and antiquated lexis. All of the teachers had a good command of the language and would be able to judge whether vocabulary was useful or obscure. If it is true that the textbooks present this kind of language, it only goes to propagate myths about TOEFL tricks,

questions that testers use to catch students out. As this is not the aim of TOEFL, the dissemination of clear facts about the range and type of language expected might well be of value in the introduction of the new examination.

Summary

In summary, vocabulary was seen as a key to success on TOEFL and the main methods for developing vocabulary seemed to be the distribution of word lists and recommendations to do extensive reading. There may be some misconceptions about the range of vocabulary required for the test, perhaps based on the content of practice materials and arising from the descriptors in the scoring rubric for the Writing test. The requirements need to be clarified and disseminated more effectively to dispel these misconceptions (if they are misconceptions).

Teaching Writing

TC2: 1262–1264: If you tell them "opening statement" or "thesis" they're like "What?" "Paragraphs?" "What is a sentence?" This is how I started originally: What is a sentence?

TL1: 324–326: I say, look, it doesn't need a brain surgeon to answer these questions! They're asking about do you like a big car or a small car!

It has been revealed in earlier testing projects (the Baltic States Year 12 Examination Reform, the St. Petersburg Examination Project, the Peacekeeping English Project) that examination candidates in many parts of the CEE region are not comfortable with the requirement to write in a foreign language. Members of test design teams in several countries have indicated that prior to examination reforms in their countries students in secondary schools did not receive explicit training in writing even in their mother tongues (St. Petersburg Examination Project, 1996, and Fekete et al., 1999, on the Hungarian Examination Reform Project). This, it has been argued, may be related to the fact that secondary school-leaving examinations in most subjects were oral rather than written and students did not need to learn the skill of expressing their thoughts effectively in writing.

Similar views were expressed by the teachers who were interviewed for this baseline study. One teacher indicated that her students had "done all kinds of writing, formal, informal, everything . . . so they are generally prepared" (TB1: 486) but it is likely that the students learned

these skills in English classes at her teaching institution rather than as part of their ordinary schooling. Other teachers made it clear that their students lacked the background they needed to be able to cope with the writing they had to do for TOEFL. One institution anticipated that students registering for their course would find writing the most difficult skill:

> TS1: 263–266: . . . the assumption was that the writing skills would be the lowest and after the diagnostic it definitely was. You could see the way their education in written communication was very low if it existed at all.

Two teachers stated that writing was their students' "greatest weakness" or "weakest point," with one suggesting that this was "cultural":

> TL2: 373–377: People are unable to pattern their thoughts within the paragraphs and within 30 minutes . . . It's difficult to convince them you only have to present a clear structure and not the originality of the ideas.

The students' lack of familiarity with TOEFL writing conventions meant that most teachers devoted a lot of class time to the writing skill. One teacher started working with essays on the first day of his course because writing was "a huge fear for (the students) and they really need to realize that there's nothing to scare them" (TL1: 323). When students were asked which skill their teachers spent most class time on more than half replied that it was writing.

About a third of the teachers stated directly or implied that writing was the most difficult skill to teach, with one teacher saying that this was because "this does not rely on the teacher so much as it rests with them (the students)" (TR1: 1393). This is in contrast to a novice teacher who felt that it was the easiest skill to deal with in the classroom:

> TP1A: 1141–1144: . . . The easiest for me is writing . . . You can do speaking and writing and reading all in one, because you can switch papers and have someone read someone else's and then you can have them speak about it.

His comment did not give any insights into the nature of what was being taught or learnt, however, but only into the types of activities that the students could be asked to do and the fact that they could practice three skills in order to develop one.

What Is Taught?

Only a few of the teachers mentioned that they worked on language when they taught writing: one mentioned using a locally produced grammar book to help with structures (TB2: 867), one mentioned doing "a lot of grammar explanation towards the end of the course" (TL2: 365), and two mentioned encouraging their students to learn linking words or phrases.

Most of the teachers (80%) used their classroom sessions to emphasize the structure of the compositions that students would have to produce for the TOEFL test, and at least half of them follow their chosen coursebook very closely. (See the Role of the Coursebook section.) The coursebook was the main source of information about the examination:

> TP1A: 661–663: I didn't major in writing but I've been following the text and trying to explain to them how it's going to be on the TOEFL.

There were several references to organization at the text level, but these tended to be quite general (e.g., the need to draw up outlines, or the need to divide the essay into an introduction, body and conclusion). One teacher explained the task in this way:

> TP1B: 843–847: You don't necessarily have to have three body paragraphs. You probably need at least two, but you need to make at least two very clear points, support them with examples and with information and introduce them properly in the introductory paragraph and conclude your essay well in the concluding paragraph.

Another teacher saw the introduction not only as a writer's way of beginning an essay but as a way of calming down and breaking up the writing task into manageable steps:

> TL1: 161: I tell them that it's almost like a mathematical formula where we want four things. I want you to memorize these, close your eyes at night and see these four steps. . .

This teacher talked about how his students were in panic the first time they were asked to write (on the diagnostic test at the beginning of the course), but once they had learned his formula they could write introductions in 5 minutes. From the observations, the formula approach to essay writing was very prominent in the lessons of TL1 and TC1 in particular. However, most of the teachers' references were to organization at the level of the paragraph. Frequent use was made of terms such as topic sentence, developing sentences, controlling idea,

and supporting idea—notions that are emphasized in the coursebooks they were using in the classroom.

The teachers did not seem to place as much emphasis on the content of the writing as they did on its structure. One teacher talked about getting students "to focus their thoughts in one smooth direction" and developing a "flow of thoughts to an ultimate conclusion" (TS1: 263). Another mentioned that the way she dealt with content had changed as she became more familiar with the TOEFL (or at least the preparation coursebooks' representation of the TOEFL):

> TL2: 1403–1405: I probably (used to) use a more complicated approach—conceptualizing things, and . . . well, the simplest version of explaining things is the quickest and I came to appreciate that.

Several of the teachers mentioned downloading or asking their students to download topics from the Internet (the TestMagic Web site [http://www.testmagic.com] seemed to be a popular source, while little mention was made of ETS sources), and one drew up her own list of topics such as abortion, alcohol, fast food restaurants, religion, and national versus international governments. She said that her students enjoyed working on the more difficult topics, but she was under the impression that the topics that appeared on the TOEFL test "weren't that difficult":

> TP1B: 848–851: I mean they're pretty broad . . . and they don't require a lot of background information. It's mostly opinions stuff . . . and if someone has been to high school or college they should be able to write about them.

Another teacher mentioned that his students had two problems with content. The first was similar to the cultural problem referred to earlier, which was mentioned by a teacher from the same country:

> TL1: 331–334: [The students from this country] like this argumentative thing. When it says do you agree or disagree they like to get off the track. I was giving the example of riding a horse. Horses have these blinders on. Don't go that way. Go that way. Stick with it. It's very easy. Then they start catching on.

The second problem was that some students thought too deeply about "the truth value" of what they were writing. The teacher's approach was pragmatic:

TL1: 340–342: I say, TOEFL people don't care what's true deep in your heart of hearts. They want you to answer a question. They're not going to come to (town name) to find out whether your uncle really has a big car!

In both cases, the message seemed to be that students should not worry too much about what they were saying but rather concentrate on how they structured their writing at a macro- and micro-level. Although several teachers got their students to brainstorm before they began writing, presumably to generate ideas on the topic, only one mentioned giving students any written or oral material to react to in their writing. The notion of integrated skills did not seem common.

One teacher stood out from the others by emphasizing creative writing in his classes, trying to help students overcome the fear of writing, which they had developed in school (TC2: 539). He gave the example of asking his students to read stories by Dostoevsky and then discussing them before writing about them. He praised his students for using their imagination and for their use of metaphor. It was only during the analysis of his interview, however, that we realized that he might not have understood that he was being asked about the way he taught writing in TOEFL classes rather than in general language improvement classes. This may explain why he seemed so off message when compared to the other teachers.

There was only one other teacher who assigned tasks that did not mimic the tasks the students would be asked to do on the TOEFL. This teacher concentrated on "authentic tasks": getting his students to write the kinds of texts they might have to produce in their target situation, including letters for scholarships, job applications, and so on. (TS1: 620). This seemed surprising until we recalled that the students attending his TOEFL classes were not all preparing for the test; some were just taking the highest-level course his institute had to offer.

How Was Writing Taught?

At least half the teachers based their teaching of writing on the coursebook they were using, using the explanations provided (presenting them orally in the session) and getting their students to practice either while they were in the lesson or by doing homework for the next

session. One teacher got her students to write in class from the beginning of the course so that they would get over their fear of the blank page:

> I say, "Look here, you think you can't do it but take the pencil in your hand and start doing it now!" (TR1: 586–587)

Others asked their students to write in class because they knew they would not do homework between their lessons. Some of the teachers were surprised when they realized that their students would not write outside the classroom, with one feeling particularly discouraged:

> TP1A: 631–632: I'd like to teach them English but if they're not motivated to learn it kind of slows the ball down.

This teacher had managed to find a way of dealing with writing that gave the students some of the practice he felt they needed without taking up too much class time (reviewing the advice given in the coursebook, getting students to write a paragraph in 10 minutes, asking them to swap papers and comment on one another's efforts, and reacting as a sample of them read their writing aloud), but another teacher reported that her students had complained when she asked them to do their writing during the lesson (TP1B: 730). She was now having to work out the best way to tackle the problem, since they were reluctant to spend too much of their class time writing but when she gave them tasks to do between lessons only a few handed in their assignments. Neither of these teachers talked about *why* their students did not do their homework. It is not known whether they knew or whether they understood the constraints their students were working under.

Two teachers insisted that their students did their writing at home and that they used the computer, so that they could get used to timing themselves and to typing their essays (TB2: 897 and TL2: 490). It was curious that other teachers who were preparing students for the computer-based TOEFL did not mention the problem of using the keyboard (see, however, the Role of the Computer section). It is not known whether their students were already proficient in typing or whether they planned to write their essays by hand.

Other teachers were able to count on their students writing at home and said they used the classroom sessions to go over the input from the coursebooks, get their students to brainstorm ideas for their assignments, and perhaps start them off on writing that they would complete as

homework. A couple of teachers mentioned that they sometimes used reading passages to illustrate how paragraphs could be organized (but not to give ideas for content) (TB2: 842). Others mentioned using sets of essays that had been given marks at different levels of the TOEFL writing scale (often from coursebooks) but there did not seem to be one standard way of exploiting this material. The teacher whose students "had done all kinds of writing" encouraged them to figure out for themselves why the essays had received certain marks:

> TB1: 493 –496: I give them an essay and that essay is . . . marked 5. So what do you like about it? What don't you like? Do you think there's something that could be improved? And through that I think they learn. They find and learn for themselves.

This teacher felt that if she told students the way to do it or the formula that they would get bored. (However, she also stated during another part of her interview that she did not always agree with the marks the essays had been given so she may have found it difficult to justify them to her students. See the Assessment in the Classroom section.)

Another teacher said that her students had marked the essays themselves and then compared their own marks with the official marks (TR1: 602). The remaining teachers only indicated that their students had worked with the material without explaining how. There were at least two teachers who did not use the sample essays with their students. The first (TP1A) did not seem to know that they existed (he was also unaware of the official TOEFL writing scales), and the second (TP1B) said that she did not understand the scales. Both of these teachers were native English speakers who had just started in their first teaching jobs.

What Kind of Feedback Did Teachers Give?

Most teachers who gave homework assigned one essay a week. The most common way of giving feedback was to react to each student's particular problems. These could be in any area, from mechanics to organization to questions of content. Two teachers stated that they marked their students' essays the way college professors would mark them (TP1B: 573, 579; TS1: 1833). One of these teachers also taught at a local university, so was presumably familiar with that university's standards for written work (this assumes, of course, that there were common standards), but the second teacher had not done any teaching at university level before. She had

just finished her undergraduate degree at a university in the United States and her claim was based on a belief that all university teachers mark the same:

> TP1B: 570–580: I grade them like my own essays were graded at [name of her university]. Some of the things that I fix they're not necessarily wrong—they're just things that I've come to avoid because I know what a lot of American professors like. . . . If you go to America and you write an essay for a professor they're going to expect such-and-such and it would be better if you learned how to do that now. I don't know what's acceptable essay form (in this country), but I know that my professors would go crazy if they saw something like "In conclusion let me state" or "I think that"—things like that.

It can only be inferred that the teacher was referring to the use of *me* and *I*. She also mentioned stressing the need to stick to the topic, "changing the wording a little bit if I think it's awkward," and fixing grammar, punctuation, and spelling mistakes (TP1B: 870). She insisted that she did not mark essays "like they do on the TOEFL." She stated in another part of the interview that she did not understand the TOEFL writing scale (TP1B: 874), but it is not clear whether she rejected the TOEFL way of marking because she did not understand it or because she was convinced her own method was better. In any case her students may have been receiving messages that were not in accord with TOEFL criteria.

Another teacher who stuck very closely to the advice given in the coursebook admitted that he was not always sure what standards to apply in marking. This teacher did not know where his own writing would stand on the TOEFL scale since he had never taken the test (TP1A: 668). In fact, only one of the ten teachers interviewed had taken the TOEFL (TL2) and so had any familiarity with the standards of the test rather than with what the preparation books said about it. However, this teacher did not use the TOEFL scale when reacting to her students' writing, feeling that grades did not motivate them and that they could well stand in the way of the development process:

> TL2: 529–533: I feel that by letting them keep writing the essays without grading there is a sort of . . . dynamics of evolving of a proper structure. If I keep writing those grades, they may form an expectation of a grade and I do not want to take this responsibility.

Only a few teachers employed other means of providing feedback on essays. A couple of teachers asked students to read their work aloud to the group and they would give on-the-spot oral feedback (TP1A and TR1) and two teachers encouraged peer marking (TL1 and TP1A). Interestingly though, all of these teachers either also gave detailed written feedback (or correction) or expressed willingness to do so if their students turned in their homework. Only one teacher talked about encouraging her students to monitor their own improvement: she marked each of their essays using the TOEFL writing scale and then asked them to look back through their previous writing and compare their marks to see whether there was any progression (TB2: 939).

Summary

The students in the sample, as is often the case in the CEE region, did not generally have a strong background in writing so their TOEFL teachers felt they needed to devote considerable time to developing the students' writing skills. They concentrated on the structure of the essay, as it was presented in the TOEFL preparation books, using a formulaic approach. They did not say much about the content of essays, and seemed to feel that the lists of topics they found in TOEFL preparation books and on the Internet were adequate for their students' needs. As regards homework, most teachers assigned it but some also made their students write in class, feeling that this was one way of ensuring they actually put pen to paper. Most teachers claimed that they and their students were familiar with the TOEFL writing scales but few teachers used the scales when giving feedback on their students' writing. Most teachers' views of the writing tasks and marking criteria stemmed from their understanding of their TOEFL preparation coursebooks or from their own experience as university teachers or students, and not from their own experience of taking TOEFL.

Teaching Speaking

TB1: 416–417: Generally speaking, nobody wants to sit for the spoken test. They don't need it. Nobody requires them to take it.

TC1: 581– 582: We do . . . speaking the least, because it's just not on the TOEFL test. It's not necessary.

These quotes reflect two of the most salient features of the TOEFL classes in this study: speaking was not generally focused on, and the reason for this was that (nearly) everyone

realized, teachers and students alike, that it would not be tested. What is worthy of note, however, was that in all classes but one English was the predominant language of communication. What lies beneath this general picture is a variety of beliefs and practices regarding the use of spoken English in the classroom. This section describes the approaches we heard about and observed. We will first examine what was happening in the classroom and then examine some of the beliefs that shaped some of the classroom practices.

Approaches to Teaching Speaking

The TOEFL classes in this study can be placed on a continuum: at one end there was a teacher conducting lessons in which virtually no English was spoken (TL2: 152) and at the other there were teachers and students who not only used English most of the time but also participated in activities specifically designed to develop speaking skills. These included conversations to start off the lessons (TP1A: 233), and discussions and debates based on reading passages (including the students' own work (TP1A: 1148, SP1A: 308) and/or also listening passages (SS1: 197). At one institution the TOEFL teacher organized extracurricular activities that included movie nights, to encourage further discussion (TS1: 601). This practice focused on content rather than form, grounded in the teacher's belief that the students needed to make mistakes and worry less about the grammar in order to progress (TS1: 354). Students seemed to appreciate this (SS1: 197, SS1: 668) but some expressed the wish to be corrected more frequently:

> SS1: 668–673: There is one thing I would welcome. He never or almost never corrects us, when we're speaking, so I would welcome this thing. I know he doesn't want to embarrass us or something but I want to know exactly and if he stops me or interrupts me I'm sure I will remember it. So I like this method but as for other students I don't know.

Features of a typical communicative language classroom and the teaching philosophy on which it is founded can be recognized here, but this may be at variance with the students' prior classroom experience. Between the two extremes lay the majority of classes, where English was used mainly to convey the content of the lesson (TC1: 595) and where students used the language either due to teacher encouragement in the case of local teachers (TB2: 853; SB2: 307) or out of necessity if their teachers did not speak their language. What students were doing in

class was mainly responding to teacher cues (TP1B: 759) and any spontaneous speech was in the form of asking questions related to the tasks they were working on.

In some classes the focus on test preparation meant that there was little time left for true speaking, as opposed to using English as a means of practicing other aspects of language:

> TC2: 338–339: There is not that much conversation going on and debating. It's drills, drills, drills—very structured.

The two teachers who focused most on speaking for the sake of speaking, organized their lessons to ensure that there was time for oral practice. TP1A stated that he had not yet given a complete practice test during class time, as this would not leave sufficient time for speaking (TP1A: 716). TS1 was able to find class time by assigning most of the writing and reading exercises as homework (TS1: 383).

Rationale for Teachers' Choices

If we examine what might have led teachers to organize their classes in these ways, we can find several clues in the data. One explanation for why teachers chose different approaches to speaking in their TOEFL classes might be related to what they considered to be the main aim of their courses. If we compare responses to questions regarding aims and methods adopted, some patterns emerge. The most obvious connection appears in the case of TL2, who stated very clearly that her aim was to prepare students for TOEFL and who chose to not include speaking at all:

> TL2: 152: I teach taking a test, not language skill.

TB1 also stated that since none of the students would sit for the *Test of Spoken English*™ (TSE®) there was no need to focus on speaking (see the second quote at the start of this section). In contrast TS1 felt that students needed a wider experience of English learning (and even had students in his class who did not intend to take the TOEFL), and he was one of the teachers who encouraged speaking the most. This picture seems to appear clear cut, until the other teachers were looked at. They also stated that their main goal was getting candidates through TOEFL successfully (although three said some additional language development was also incorporated), but they all decided to encourage speaking in English. In the observations, we witnessed frequent correction of pronunciation, which suggested a (perhaps unconscious) need to help students

68

improve their spoken communication, even when this did not fit in with the stated aims of the class.

Other factors that may have influenced the teachers' decisions were detected. TP1B did not focus on speaking because her school ran a separate conversation class for the TOEFL students. It was curious then that TP1A, who worked at the same institution and whose students also had separate conversation classes, was the teacher who did not do complete practice tests during his lessons so that there would be enough time for speaking practice. One explanation for this difference in teaching may be that TP1A was very inexperienced, having only been teaching a few months and never having taught a TOEFL class before. His main source of ideas and guidance was his initial teaching certificate, which focused on techniques for communicative language teaching. He stated that it was very challenging to apply these techniques to the TOEFL preparation book he had been directed to use. TP1B had not yet taken an initial teacher training course and therefore was not indoctrinated with communicative ideas.

It is also the case though that TP1A did not seem very certain about whether a speaking test was included in TOEFL or not:

TP1A: 582–584: . . . I assume on the official exam there's a speaking section as well but I'm not sure. Is there a speaking section on the TOEFL?

There were clearly differences between these teachers in training and test awareness, which may have influenced the decisions they took when planning their classes.

Another way in which school factors might have played a role in the presence or absence of speaking in the TOEFL courses was seen in the case of TC2, who claimed that his school had a reputation for being "English only" and that this was one of its attractions to potential students. Unfortunately, we were unable to gather data from students at this institution (and TC2 was the director of studies as well as the TOEFL teacher), so we cannot confirm that this was the situation—but it seemed plausible.

A similar school ethos was described by one of the directors of studies, who had won a scholarship to study in Japan while she was at university. She learned there that written English was not sufficient and she formed the opinion that language teaching needed to include speaking. English as the medium of instruction was therefore encouraged in that institution under her influence (DOSR1: 760).

The material that was used also played a role in how much speaking took place in the classroom. A teacher who in her non-TOEFL classes used English throughout and who was herself a highly competent speaker claimed that there was little speaking in her TOEFL course as a direct result of the nature of the materials: her classes consisted almost entirely of students working individually through practice exams on computer (TB1: 570). There was little need for the students to talk beyond asking for assistance if there was a technical hitch. The same phenomenon was observed in one of the other centers during the second half of each lesson, which was when students worked individually at computers (TB2: 1059).

Another reason teachers gave for their own use of English was that it provided additional input for developing the students' listening comprehension. But what was the advantage to students of spending time practicing speaking if it was not going to be tested? The answer may be in the responses they gave when asked about their attitude to the new TOEFL. Many of the students were in favor of the idea that speaking should be included in the test. Some said they felt their overall ability could not be assessed without an oral test (SL2: 1104, SR1: 1314). Indeed, one student (and a teacher) expressed surprise that TOEFL did not already test speaking (SR1: 1139, TC1: 1553–1554).

However, the directors of studies expressed some concerns about the idea of a speaking test. There were comments about the extra stress that it would create (DOSC1: 669) and that some candidates would probably be afraid of such a test:

> DOSC1: 514–516: I know that it's supposed to include a speaking component because that's what everybody's talking about and that's what everybody fears.

Preparation for Study Abroad

One reason that students gave for wanting speaking in the TOEFL courses was that they needed to prepare for study in an English-speaking environment abroad. They needed practice because oral skills had not been prioritized in their state education systems:

> SB2: 682–684 (see also SP1A: 1763): Actually I think this is the problem of English lessons in [our country]. We don't speak enough English. And when we have to speak to native speakers, it's oops!

However, this was not a view universally held: One of the students believed that if students had a good foundation in other aspects of the language, as attested by a TOEFL score, then they would soon manage to communicate effectively once immersed in the English-speaking environment (SS1: 1863). TS1 went further, suggesting that speaking was not necessary in the target use situation:

> TS1: 1845–1880: Their listening skills may be great. Their academic skills as far as grammar and writing can be excellent. But maybe their pronunciation is absolutely horrible. Is that necessary for someone to study abroad? I don't think so. . . . I mean, I did very little speaking. Most people do very little speaking in their university. You understand, you take notes, you take a test. What do you need to speak for?

What was interesting about TS1's view, however, was that he was one of the teachers whose classes featured extensive speaking.

Comparisons With Other Tests

One of the most interesting areas of comparison made between TOEFL and other international tests had to do with the existence or absence of a speaking test. It was reported that many candidates took TOEFL rather than another test serving the same purpose because they did not have to do speaking (DOSR1: 826). However, DOSB2 expressed the opposite view, saying that students (maybe not all) liked the speaking tests offered by other agencies (Cambridge was mentioned in this case) as they allowed them to see their true ability. He claimed that the students were often pleasantly surprised by their results, and that this carried another advantage:

> DOSB2: 881:
>
> Director of studies: Students who take an exam where there is a speaking section, they are happy to understand that they have performed well in the speaking section.
>
> Interviewer: Yes, of course.
>
> Director of studies: And this compensates for any . . . how should I say . . . any failures in the grammar part, or the reading part.

71

The belief that a high performance on one section of the test could compensate for a low performance on another section was not necessarily accurate, but what matters is what the users believe to be the case.

Speaking in New TOEFL

The introduction of a speaking test seemed to be the most notable feature of the new TOEFL for the teachers in the study: all seven who reported that they had heard about the new test mentioned the speaking test when asked what they knew about it. This was sometimes the only new feature they were aware of. The same pattern could be seen amongst the directors of studies, where only one of the six who knew there would be a new TOEFL was not aware of the speaking test. A smaller proportion of the students (half) had heard of the changes. They may not have relished the extra challenge, but felt that it was a necessary development (SB2: 1346).

Summary

All of the teachers but one used English as their means of classroom communication, but they rarely focused on how to develop the speaking abilities of their students. They used English for a variety of reasons, including helping their students to develop their listening ability, but it was generally not felt important to do specific work on speaking since it was not part of the standard TOEFL and none of the students needed to take the Test of Spoken English.

Assessment in the Classroom

TL1: 663–674: . . . After we do the structure we do a practice test. We finish that. We're still working on the reading then if we finish that we do a practice test there. And at the very end of the class we do the complete TOEFL test. It actually takes two sessions, but we do the essay, we do everything, and I take that all from the book just because it's there and that's what it's there for.

Essential to any study of classroom practices are questions concerning assessment: what are the functions of assessment in a particular setting, what is assessed, and how does the assessment take place? Both teachers and students were asked about assessment practices in their institutions (see Questions 43 to 51 in the Teacher Interview in Appendix G and Questions 42 to 46 in the Student Interview in Appendix H). No uniform pattern emerged which applied to all

settings, but several themes appeared frequently enough to be of interest in this investigation. This section discusses what we learned about the types of testing that were carried out in the schools, the materials that the tests were based on, and how teachers and students reacted to the results.

The Screening Function

There seemed to be several functions for assessment within the schools, the first of which was screening. We specifically asked whether students were screened before they were allowed to join TOEFL preparation courses, in order to get some impression of the range of student ability the TOEFL teachers had to cater for. We found that less screening took place than we would have imagined, mainly because the students in the courses had often been enrolled on other courses in the institution and had worked their way upwards through the levels. The screening that did take place was for outsiders, students who were enrolling in the institution for the first time. The results were rarely restrictive: If the students did not do well on the screen test, they would be advised to build up their knowledge and skills before taking the TOEFL course later; but, they would not be barred from joining the course if they wished to go against the advice given.

> DOSB1: 188–193: We would prefer students who are upper-intermediate or advanced level. Sometimes there are exceptions. Somebody with intermediate level just insists in getting into the course and it's not the appropriate level for the TOEFL, but still if you want to try you can try. It's up to you.

Where screening was carried out, it was generally via commercially available tests, either in the original format or adapted, or other in-house. However, one school simply asked potential students to fill in a self- evaluation form containing level descriptions, and then allowed them to change classes if they had not evaluated themselves accurately (DOSR1: 221).

Diagnostic Tests

Diagnostic tests were mentioned by many of the teachers and directors of studies. These were tests students took shortly after commencing their courses. They were taken from the TOEFL preparation coursebooks being used at the institution. It was not clear what these tests revealed about the students beyond basic information, such as which skill area they performed

well on, and, therefore, some indication of where their strengths and weaknesses lay. There was some value in this as a consciousness-raising exercise, as this student indicated.

SS1: 291: On the first test we realized that we don't know how to write essays

There was, however, no mention of any profile indicating which subskills the students needed to pay more attention to. This would clearly provide useful information to both the teacher and the students but it did not seem that the instruments could cope with such a level of detail. The term *diagnostic test* was used widely in the coursebooks. It had been adopted by the course providers and the students, but the tests were only practice tests, appearing early on in the commercial preparation materials.

We also asked teachers whether they used tests during their courses to assess students' progress in the specific skills tested on TOEFL, but none of them reported doing so.

TB1: 570–576:

Interviewer: Have you ever given them any other sorts of tests which aren't from the books, to see how they're doing in reading or listening?

TB1: No. I didn't feel like it was necessary because in the previous courses that they've been to they've done that.

What this quote seemed to suggest was that measuring progress was not appropriate on a TOEFL course. We did not follow up the point to find out why. It may be, however, that the notion of progress, which implies movement from one point to another, is not relevant on a course which consists of an accumulation of points to learn rather than a systematic build-up of skills—with no notion of going from something easier to something harder (but see comments on essay writing that follow).

Practice Tests

As mentioned often throughout this study, the courses seemed to be shaped by the format and content of the coursebooks used. Where these included practice tests the assumption was that they were parallel in form to the real TOEFL test. Teachers reported that their former students had told them that the practice tests were similar to the TOEFL, but it was not clear if any of the teachers had queried how parallel they really were. We doubt that the students had the skills

necessary to make this kind of judgment or, if they did, that they would have had the time or inclination to an analysis if they were actually concentrating on performing well as they took the test (see the Communication section for further details of student feedback).

Feedback on Writing

One aspect of assessment that seemed important to ask about was the use of scales for marking the Writing section of the TOEFL test. We expected to see teachers using the 0–6 TOEFL test rating scale in order to offer students an indication of what score they might receive on their essay. There was evidence that the scale was used by some teachers, but it was used in a variety of ways. Some teachers did not seem comfortable using the scale to give students a score for a piece of writing (TB1: 515). They did not feel confident that they could give scores in line with the examples provided in the Bulletin and on the Web site. This was not because they were unfamiliar with scales in general: Most, though not all, had used them before, some extensively. They simply were not sure about how their own assessments matched up with the examples given and they worried about giving their students the idea that their writing was better than a TOEFL essay marker might deem it to be. One teacher expressed her doubts in this way:

> TB1: 516–519: I read a lot of essays which are on the Internet in different and specialized Web sites, and I saw essays which were marked, let's say, 5. If I were the teacher I would never give a 5 for that essay. That's why I don't want to mark them.

At least one teacher, who was teaching her first TOEFL course, admitted that she did not understand the scale:

> TP1B: 872–877: . . . But I don't score like they do on the TOEFL, like 1 to 6, just because I read the criteria for scoring for each number and it didn't make sense to me. It's not very clear to me what a minimal score is, what an adequate score is, what a good score is. I don't know what the TOEFL scorers are looking for when they score it.

These examples suggest that teachers may need more guidance before they can competently and confidently use the scales with their students.

Other teachers implied that their students might have problems understanding the scale. One was going to introduce it later in the course, which suggests that he felt it was too much for

them to deal with earlier. He felt the scale acted as a restraint, which was not beneficial for their writing development (TS1: 791). Another teacher seemed to feel that the students were better off receiving encouragement to do their best without worrying about understanding the scale:

> TB1: 556–557: If they have the wish, they can always read the criteria but it wouldn't mean anything to them.

The data confirmed that this teacher's students were not familiar with the scale.

Other teachers used various methods to familiarize their students with the scale and to get them to mark their own or their peers' writing. One got students involved in marking their own work early on:

> TR1: 611–612: At the beginning I have given them samples and asked them to mark them and compare their marking.

She also elicited scores from student peers as each week some students in the group read aloud the essay they had prepared at home (TR1: 602). Students on another course also read each other's work, with the aim of assigning scores and correcting errors (SL2: 297).

However, teachers mostly offered comments on written work to help the students see where to improve at a detailed level rather than giving a score from the TOEFL scale (TL2: 49, TP1A: 634). Some claimed to give quite detailed feedback:

> TB1: 519–522: I review the essay. I point out, in my opinion, which are the weaknesses, which are the strengths, why they shouldn't write that, why it's a mistake and I write everything down.

Students seemed to appreciate this attention to detail:

> SS1: 590–591: At this level we don't need any grades. We can do it ourselves. We like to have some comments on how to improve.

Also:

> SP1B: 562–564: [Teacher's name] always writes me some corrections—what was wrong, what was good, and so this is very helpful to improve your writing skills.

It was clear that some students equated grades with being at school again, where marking seemed a random process, with no feedback, and, therefore not informative or formative. However, a variety of opinions will emerge in any survey, and the opposite view was expressed by a student in another setting:

SL2: 600–606:

Student: I would like to get a score.

Interviewer: Why do you think that would help?

Student: Well, if my score would be like four and a half I would be more assured about my TOEFL exam because now I don't know how I am writing. Maybe it is only two points.

A classmate remarked, however, that she did not want any scores, as it would make her nervous. It is hard to know how teachers can satisfy such a variety of perceptions and expectations amongst their students.

Using Computers for Self-Assessment

There was little use made of computers for test practice in the classes observed, and test method effect was not taken into account in most classes. The teachers seemed to underrate the value of sitting through a complete practice test under test conditions—above all, timed conditions. This was in spite of the fact that, as TB1 observed, the students encountered unfamiliar item types when they moved from general TOEFL practice to working on computer-based TOEFL materials. (TB1: 857).

Summary

There were several functions for assessment in the institutions providing TOEFL preparation courses. These included screening, diagnosis (in the broadest possible sense), and giving the students practice dealing with the TOEFL tasks and format (though often without the benefit of computers). Many of the tests used came from TOEFL preparation coursebooks. They may have served a useful function in terms of raising consciousness (or confidence building), but there is no evidence (other than student report) that the tests were truly parallel to the real TOEFL.

Most other assessment was done in the area of writing. Teachers tended to give detailed feedback to students rather than giving a mark using the TOEFL writing scale. Some teachers did not seem confident using the scale.

The Role of the Coursebook

TL1: 907–908: I joke: This [the TOEFL preparation book] is the TOEFL Bible. I'm God. Read the Bible every day . . .

One of the clearest findings in this study was that most of the teachers running TOEFL preparation classes relied greatly on the commercial preparation coursebooks that they or their schools had selected. This was not the case for every teacher, but the preparation coursebook was at the heart of the majority of the courses investigated. The findings therefore parallel those of Alderson and Hamp-Lyons (1996), who concluded that "most teachers just seemed to do what the book says and what they claim the students want" (p. 286). This section discusses in what ways, and to what extent, this is true for the teachers in our study.

Belief in the Coursebook

The religious reference in the opening quote is apposite, in that teachers tended to show a great faith in the TOEFL preparation coursebooks they were using:

TL2: 1357–368:
Teacher: I think that people feel very safe with the TOEFL preparation materials.
Interviewer: You think they're quite well produced materials?
Teacher: Now there's that belief that you do not have to take anything apart from those materials
Interviewer: They'll do the whole job for you
Teacher: Yes, because there wouldn't be such reliable items elsewhere.

Only 1 of the 10 teachers had taken the TOEFL herself, so all the others relied on secondhand information understand what the real TOEFL test was like. Most used the coursebook as their main point of reference, as this teacher did in the teaching of writing:

TP1A: 600–602: . . .we've been working on paragraphs and I follow the book closely because I'm assuming this is what they're going to be graded on when they take the TOEFL exam.

Again, this situation was similar to that found by Alderson and Hamp-Lyons (1996), who warned that "it may be difficult to untangle test effects from textbook effects" (p. 282).

Who Chose the Coursebook and Why?

From follow-up data gathered from directors of studies, subsequent to the interviews with them, we clarified why the schools use the books they do. Table 11 shows who chose the coursebook in use.

Table 11

Who Chooses the TOEFL Preparation Coursebook?

Person choosing	# of instances
DOS[a]	2
Current teacher	1
Past teacher	1
Combination: DOS and teacher	1

Note. Six replies are recorded in this table. A seventh institution did not rely on any one coursebook but rather an in-house compilation of several coursebooks.

[a] DOS = director of studies.

Three reasons were given for why certain coursebooks were chosen rather than others. The first reason had to do with the book's coverage:

TL1: 441–443: That's why I chose the book that I use. I think that it goes into every topic but it doesn't spend too much time. It gets what you probably will need and that's why I like it.

The second had to do with gaining a market edge. One school imported its coursebook at great expense to ensure that it offered materials that were not only new to its students but definitely not being used by any other TOEFL preparation providers locally. The third reason

that was given (mentioned by two schools) was that certain books were cheaper than others, due to discounts offered by the publishers.

How Did Teachers Use the Books?

The coursebook acted in many cases as the syllabus for teaching, as highlighted in this example:

> TB2: 603–615:
>
> Interviewer: How do you work out what to cover in the 72 hours?
>
> Teacher: Oh, I just have the book, and I stick to the book.
>
> Interviewer: You use that?
>
> Teacher: I think this is what I am supposed to do.
>
> Interviewer: So that acts as the course outline. How much of it do you try and get through in the 72 hours? The whole thing?
>
> Teacher: Yes.

TC1 was another teacher who followed her coursebook very closely. As she was a substitute teacher she may not have felt that she had the right to divert from the plan she was given. However, she also hinted that she depended on the materials because of her lack of experience in TOEFL teaching (TC1: 1088). This point will be revisited later on in this discussion.

As already mentioned, various teachers indicated that they taught the TOEFL skills areas in the order they were presented in the coursebook. This often followed the order the skills were dealt with on the test (Listening, Structure, Reading, Writing; see, for example, *Cambridge Preparation for the TOEFL Test* [Gear & Gear, 2002] and *Barron's How to Prepare for the TOEFL* [Sharpe, 1999]). Only one teacher queried the rationale for this order or attempted to do things in an alternative manner. He believed that his students would benefit from working on the listening section last rather than first so that they had time to get used to his voice before attempting to listen to recordings of other English speakers (TL1: 311–315, 400).

Most teachers worked through their coursebooks systematically, according to their perceptions of their students' needs. However, two of the teachers worked in a slightly different manner. TC1's school had produced its own coursebook, using materials from several

commercially produced courses. What they taught and when they taught it was dictated by this coursebook. TS1 had made a conscious decision not to follow his coursebook(s) exclusively as he felt that his students needed more than just TOEFL test preparation. Indeed, as mentioned elsewhere, he had several students in his class who did not intend to take the TOEFL test and he chose, for their sake, to give all the students materials which he saw as more interesting and stimulating, including texts (such as poems) without a TOEFL-test focus.

Recognizing the Books for What They Are

Fundamental to a judgment about how well the books fulfill their function, is being clear about what the nature of their role should be. TL2 felt that the preparation books were there only to help students to familiarize themselves with the TOEFL test:

> TL2: 198–200: This is not language learning material and you don't have to bother about learning the tenses thoroughly. You only have to be warned of what will be tested, and about the types of the test questions.

Her attitude was reflected in the way she conducted her classes, which were based solely on practice materials and the specific knowledge and skills needed to get good results. TS1, on the other hand, seemed rather surprised that the materials were so focused on the TOEFL:

> TS1: 843–846: Their goal is to pass the test and it's written in all the material. I find a lot of the books are . . . really structured, not for the students to learn generally, but to pass the test.

This last point leads to the nature of the books themselves. Both TP1A and TL2 mentioned that the books were designed for self-study rather than for class use:

> TP1A: 195–197 (see also TL2: 176): Well, in my opinion this is a good book and sometimes I feel as though the students themselves can go through the different reading sections and then if they have questions come back to me.

Why Do Teachers Rely on the Coursebooks?

Several teachers stated that they adhered to the format of the coursebook because this was what students expected:

TP1B: 324–330: I've even occasionally pioneered doing things a little bit differently, not following the examples in the book exactly, and I haven't met with the best reception. . . . The people I know in that class who are really serious about it, I know they just want to do exercises for the TOEFL. They just want to go through the book and get as much practice with what's going to be on the test as possible.

This extract is a reminder that since the majority of schools were commercial operations the relationship between students and teacher was also one of client and service provider. Another reason given was that the books covered everything needed for a test preparation course:

TL1: 1969–1971: One of the things that I love about this book is that if they have problems with the grammar . . . it says you need to study this section and this section and you immediately go to it.

A third reason was that teachers had few other resources to fall back on. They had not received any training to teach the TOEFL test and, on the whole, had worked out their own syllabuses and methodology on their own, mainly using nothing more than a preparation coursebook.

TL1: 373–376: I didn't know what TOEFL was. I'd just come from America. I walked in and the director said, "Teach this class." I was like "What is it?" "It's a test. Teach it," and I was like, "OK" and so I learned the hard way.

TC1 was unusual in having had a TOEFL "mentor" and the opportunity to observe and indeed join some classes and see how TOEFL teaching was carried out at her institute. The other teacher did not have this support. It is hardly surprising, given this background, that the teachers' views of what were important and the techniques they used often mirrored those in the coursebooks.

Within the field of EFL/ESL a movement has arisen (called *Dogme*, after the naturalistic film movement) which criticizes "materials-driven pedagogy," claiming that the overuse of commercial materials and technology undermines teachers' ability to truly address students' needs and promotes a one-size-fits-all approach to what happens in the classroom. Though many of the arguments are quite convincing, the movement does not fully address the role of the

coursebook as a support for novice teachers who are just building up their knowledge and skills base, or for teachers who lack confidence in their own foreign-language abilities (whether with justification or not). There may well be an element of this (the coursebook as crutch) in what was seen among the teachers in the study, who all varied regarding their general teaching experience, TOEFL test teaching experience and, though to a lesser extent, English-language ability.

The role of the coursebook, in courses that exist solely for test-preparation purposes is one that needs to be discussed further. One argument which has been presented forcefully is that teachers and students are merely "part of huge test preparation industry fuelled by students' anxiety to succeed on this high-stakes test" (Alderson & Hamp-Lyons, 1996, p. 293). However, implicit in this argument is the assumption that the main participants are not aware of what they are doing. Students invest time and money to do well on the TOEFL test in order to achieve ambitions of study or work abroad. This will allow them (they hope) to enjoy a better life style in the future. Most of the students in this study were very open about their aims, which were not to improve their English-language ability or even to prepare for higher education, but to get a good enough result to allow them to pursue their ambitions. This problematizes somewhat the notion of publisher exploitation.

The question that is most challenging is not "What other sources can the teachers and students use?" but "How can we make sure that these sources truly reflect the constructs underlying the examination and give sufficient guidance to teachers to enable them to teach effectively?"

Summary

Most of the teachers were very dependent on their coursebooks and worked their students through either the whole book or the sections that the teachers thought the students were in most need of help. Many of the teachers had learned about the TOEFL test from the coursebook, and it would therefore be very hard to say that the TOEFL test had any influence that was not mediated by the coursebook. The teachers seemed very familiar with the messages contained in their coursebooks, but the books did not include much, if anything, in the way of teacher guidance. The criteria for buying coursebooks were not always pedagogic; they also included price and the desire to have materials that no one else could offer.

The Role of the Computer

One of the key questions posed by Kirsch, Jamieson, Taylor, and Eignor (1998) in research investigating potential construct irrelevant variance with the introduction of computer-based testing of the TOEFL test in 1998 was: "Do all international test takers have comparable access to and experience with using computers?" (p. 2). This section looks at the issues of access and experience (encompassing familiarity and competence) in order to build a picture of the current situation of TOEFL test candidates in Central and Eastern Europe. This is important for two reasons. First, the Kirsch study is several years old and with the speed of change in the CEE region and regarding computer technology (which has a knock-on effect on cost and therefore access), the situation may now be quite different. In addition, the informants in the Kirsch study were mainly non-European, and the Europeans studied came from Western Europe. Students from the countries in this study were, therefore, underrepresented.

Research previously undertaken by Pelgrum , Janssen Reinen, and Plomp (1993) had included three countries of Central and Eastern Europe (Bulgaria, Latvia and Slovenia), only one of which, Bulgaria, is included in our baseline study. Pelgrum et al. reported that about 20% of secondary school students in those countries did not use computers at all. The temptation must be resisted of seeing these countries as homogeneous though, since they had very different political profiles during the Soviet era and now have very different economies and opportunities. The fact that there was a large proportion of students who did not use computers at that time says little about what the situation might be at present. A study undertaken for ETS by Taylor, Jamieson, Eignor, and Kirsch (1998) states:

> It is not known how characteristics of the TOEFL examinees who will take the operational computer-based test in the future may change. At this point, forecasting possible changes in the TOEFL test-taking population, either as a result of the introduction of computer-based testing or other international social and political events, would be speculative. (p. 27)

The purpose of this section is to look at the role of the computer in TOEFL test classrooms in 2003, the year we collected data. First discussed are the computer facilities (and other technical facilities) available at the institutions visited. The computer competence of the students is discussed, as is the computer familiarity and computer competence of the teachers.

84

Computer Facilities

On of the first things we wished to investigate was the kind of computer provision available in the ten institutions in the study, and whether the amount or type of provision related to the version of the TOEFL test (PBT or CBT) that the students were preparing for. This information is presented in Table 12.

Table 12

Computer Provision in TOEFL Preparation Institutions

Country	Bulgaria		Croatia		Lithuania			Poland	Romania	Slovakia
Sites	B1	B2	C1	C2	L1	L2	L3	P1	R1	S1
Computers used by students in class?	yes	yes + technical support	no	no	no	no	no	no	no	no
Computers available to students for self-access?	yes	yes	yes	yes?	yes?	yes	yes	yes?	no	yes
Computers available to teachers?	yes	yes	yes	yes	yes	yes	yes	yes	no?	yes
Which TOEFL version available in country?	CBT in Sofia + 1 PBT center		CBT only—in Zagreb		CBT only—in Vilnius			PBT only—in 6 centers (Nearest CBT in Berlin—very far, expensive to travel)	CBT in Bucharest + 3 PBT centers	PBT only—in 2 centers (Nearest CBT in Vienna—close, good access)

Note. ? = Conflicting information from different informants.

We expected that we would not see computers being used in class in the institutions where students were preparing for the PBT version of TOEFL test, and this was indeed the case (see Sites P1 and S1). We expected that we would see more use of computers in the classes where students were preparing for the CBT, but this was only the case in two of the eight institutions in this category (Sites B1 and B2).

We saw a wide range of technical provision across the institutions. In one school, there were no computers at all and the teacher did not even have access to a CD player, but only a cassette recorder. At the other extreme was a school with a dedicated computer lab for teachers to hold classes in, as well as a self-access lab for students where a technician's assistance was available (B2). One school did not have a dedicated computer lab but did have eight computers which were set up in one of the classrooms for every single TOEFL test class and then stored away again so the classroom could be used for regular classes (DOSB1: 564). These two schools were the only ones to have computer-based classes featuring regularly on their TOEFL preparation program. The other schools had either a few computers available for students to use during out-of-class hours on a self-access basis, or had computers in the staff room for teachers' access only. One school had negotiated an arrangement with the nearest Fulbright office so that its students were allowed to experience one computer practice test free of charge during their teaching program. This was meant to compensate the students for the school's lack of computer facilities.

There was quite a difference in use even where schools offered computers to their students for out-of-class use. In one school, a booking system was in operation and the director of studies reported near constant use, whereas in another school students were not very sure what facilities the school offered them. We later discovered from the director of studies that four computers were available for self-study (DOSP1: 655), but the students did not know this and therefore did not use them.

When we asked students if they felt anything was missing from their TOEFL preparation courses, those at two sites raised the matter of computer access. In both cases the students would be taking the computer-based TOEFL but no class time was dedicated to computer-based practice.

SL2: 757: As I said I would like more practice on the computers.

It seems that most students had access to computers outside of class hours so it is not clear why access at school was felt necessary. The most likely explanation is that the students wished to use the computers during class time (with teacher assistance available) rather than purely for independent study. One teacher, from the school with the fewest facilities, hoped that the school would soon deal with its technical constraints:

TR1: 437–444:

Teacher: I need a book with cassettes and most of the modern books are with CDs so we can't use them.

Interviewer: Of course. So that's another consideration of what to choose.

Teacher: I trust this school will probably buy computers in the very near future and so on but for the time being that's not the case.

At another school, the director of studies talked enthusiastically about a new computer laboratory that was being built and which promised improved access for students preparing to take tests (DOSL3: 128). Subsequent correspondence with this director of studies has confirmed the students are now able to work in this new facility. The moveable computer lab mentioned earlier was the best solution found by the owner/director of studies, who was faced with few classrooms and needed maximum flexibility. He also discussed hopes to expand and offer better computer facilities (DOSB1: 581), as did another director of studies who spoke of plans for Internet access for students and teachers (DOSP1: 662). It was felt that technological development was going to be necessary for schools to catch up or keep ahead of their competitors.

We felt it was important to ask about students' access to computers away from the institution. Most students seemed to have ample opportunity for off-site computer-based practice, for general language development exercises as well as test preparation, outside of class hours. Quite a few claimed to use the Internet for gathering information about the TOEFL test and finding language exercises; they also claimed to use various CD-ROMs to undertake extra practice (SB1: 420, SB2: 1029, SP1A: 429, SP1B: 1201 and 1227, SS1: 834). Access did not appear to be a significant problem for the sample of students interviewed who lived in capital cities and large provincial towns.

TB1: 870–880:

Interviewer: How many of them will do extra computer-based practice outside of class hours?

Teacher: All of them. They all have computers at home or even if they don't they know a friend with one. They can go and practice. All of them have personal discs so they can share.

DOSB1 supported this perception, but with a qualification:

DOSB1: 633–643:

Interviewer:	Do you get the impression that most students have easy access to a computer?
Director of studies:	I think so. There are plenty of Internet cafes. . . . It's not that expensive . . .
Interviewer:	Right, so even if they haven't got them at home then they don't have access problems. Would the university students have access to them at university as well?
Director of studies:	Only at the American University.

Students' Computer Competence

This section discusses the issue of the informants' experience with computers, incorporating not only the concept of familiarity but competence as well. This is of importance because, according to Taylor et al. (1998),

[M]any in the field of language testing are concerned that the introduction of a computer-based TOEFL test in 1998 will confound language proficiency with computer proficiency and thus bring in construct-irrelevant variance to the measurement of examinees' English-language abilities. (p. i)

The focus of the Taylor et al. (1998) study was whether the students' level of computer literacy would affect their ability to do well on CBT. It was possible at that time to find TOEFL candidates who were not familiar with computers and who needed to take advantage of the CBT computer-familiarization tutorials to be able to demonstrate their language ability. We assumed that computer competence among students would have increased since that time due to lowering of costs for technology, the increased use of computers in schools (this is from our knowledge of the education system in the region), and a strong will for modernization in this part of the world. It is not possible to make a direct comparison between this study and Taylor et al. (1998), due to the fact that different countries are represented in each sample; however, the current situation can be described.

None of the students interviewed reported unease with using computers and, indeed, most claimed to feel at least comfortable using them, if not very competent.

SL1: 368–382:

Interviewer: And how do you find doing practice tests on the computer? Is it useful?

Student 2: Yes, I think it's more useful than a book—than a testing book.

Interviewer: How do you see the differences? Why aren't you so keen on the ones in the book?

Student 2: Much easier on computer.

Interviewer: Right, that's interesting. So you find it a little bit easier. Is it the actual content of the test, the things they asked you . . . or is it just easier because it's on the computer?

Student 2: No, I feel better on the computer.

Some of the students were clearly proficient and regular computer users, needing to use them for their field of study (e.g., mathematics) at university level (SL2: 908).

Of course, this sample was self-selecting (they volunteered to cooperate in the study), so they may have been more confident than other students in many respects—not only in their ability to communicate in English, but also their computer skills. We have no way of knowing this for certain, however, so we assumed that they are representative of their cohort. They gave the impression that students of their generation were generally familiar with the basics of computing at least and had no major problems in this respect. One director of studies doubted that this was the case (DOSC1: 680) but she could not have had much evidence to base her opinion on given that computers were rarely used in class. Some other directors of studies were responsible for helping students to register for the TOEFL test and the students' lack of familiarity with certain functions, such as making payments over the Internet, might have given them the wrong impression of their overall abilities. The truth is though that there is little evidence of how well students coped with computer-based practice since it was so rarely incorporated into the classes observed. We thus had to rely on students' self-reporting of their abilities and comments from teachers who might or might not have had direct evidence.

Familiarity with the format of the computer-based test was a separate matter. The two institutions that offered computer practice in class were helping students to get used to the CBT

interface, but the teachers running classes without computers seemed to assume that students would become test-aware by working on their own, out of class. The aim of their courses was to cover content, language development, and question types, but not to simulate how it feels to sit in front of the computer on the actual day of the test. Some directors of studies were aware that this could cause anxiety, as demonstrated by these two extracts (the second projecting ahead to introduction of an oral test with the new TOEFL):

> DOSP1: 490–493: I can imagine it must be very hard for a person who is practicing paper TOEFL to suddenly sit on the exam in front of the computer like where the person doesn't know the commands on the computer, where to click and stuff like that.

> DOSC1: 669–673: I'm sure it's going to be quite frightening. I'm not sure how exactly it's going to be administered but I imagine there will be a microphone. There should be one and I'm sure it's going to be quite stressful. There is no exam as such—yet.

> One issue that emerged as problematic was that of typing essays for the CBT Writing test:

> SR1: 1166–1170: They say it will not be a problem for computer-based because we will show you, but to make an essay on computer you should have been practicing before that. When I was learning to write on computer it was very hard, because "Oh, where is A? Over here. Where is M? Over here."

This view was echoed by students in Interview SB1, who felt that some students would still prefer to write their essay by hand. Indeed, one of the students was intending to use this option when taking the CBT, as she did not feel confident she could use her time well due to poor typing skills. Familiarization with the mouse and scrolling (the focus of the CBT pre-test tutorial) is something which most computer are probably able to do, but the ability to type effectively and efficiently under test conditions and not waste precious time finding keys and correcting (especially when the student is used to an alternative alphabet) may introduce a certain amount of construct-irrelevant variance into the testing situation.

Another problem related to the extended use of computers was eyestrain. As the CBT exam was several hours long (and the new TOEFL test will be at least that long) this issue could

not be overlooked, and could surely affect some candidates more than others (i.e., those not used to looking at a computer screen for extended periods of time). As one teacher reported:

> TL2: 047: People keep complaining that their eyes are very sore after very intense watching the screen.

Another student raised the problem of students with weaker eyesight being disadvantaged (SC1: 1706). Further studies are needed to identify the actual effects of the computer and to ensure that such factors are not affecting final scores.

Teachers' Computer Competence

We asked teachers about their own computer competence and whether it affected the way they ran their classes. The teachers varied greatly in their computer ability, with fully comfortable and competent users at one end of the spectrum and a self-designated "technophobe" at the other. One teacher, nearer the latter end, felt she had "missed the boat" somehow. It was not necessarily the case that she wanted to incorporate computers into her classes, or that she felt her lessons were poorer in quality without them (she was a competent and highly experienced teacher), but she wished to be more computer-literate for affective reasons:

> Interviewer: Do you feel your confidence in your skills has any effect on your teaching the class? Are you worried about it?
>
> TR1: 1043–1053:
>
> Teacher: I am.
>
> Interviewer: In what way?
>
> Teacher: Because you know . . . there is a gap between me and this young generation. This is a computer-born generation, which I am not.

As for the issue of whether the teachers' computing skills affected their teaching of TOEFL classes, it again has to be remembered that some of the teachers were preparing students for the PBT version of the TOEFL test and it would not be expected of them to use computers or even refer to them. However, the issue would seem to be relevant for the teachers preparing students for the CBT. These teachers gave mixed responses when we asked them whether they thought their confidence and ability in using computers had any effect on their ability to teach a

TOEFL test class or their methodology. Some felt such skills were clearly needed but others felt they were not. The former were in fact the ones using computers in the classroom already, and they may have been more aware of the challenges of doing so.

The teachers mentioned two ways in which their computer skills might be drawn on when teaching. The first was in dealing with any technical problems that might arise, as this teacher explains:

> TB1: 10670–1072: . . . if you don't know about computers, I think it would be difficult to teach computer-based TOEFL because sometimes something happens [with the] equipment or machines and you have to know how to react.

> TL1: 119: The second was in understanding the students' difficulties when they used the programs and being able to empathize with any difficulties they were facing.

An interesting feature of one school was that while the director of studies talked at length about using computers for a range of other courses (including preparation for the London Chamber of Commerce and Industry [LCCI] examinations and the First Certificate in English [FCE]), the TOEFL teacher chose not to use computers in the classroom. This teacher seemed to have more say in how he would run his classes than other teachers felt they had (TP1A). He nominated himself a "computer idiot" (TL1: 939) and demonstrated a clear preference for traditional methods. This was not a general pattern, however.

Another area of interest was whether a teacher's computer skills might have consequences was for job selection. We asked the directors of studies what selection criteria they employed when recruiting TOEFL teachers. Only one, whose school offered computer-based courses, specifically mentioned competence in use of computers (DOSB1: 323–324). Others gave priority to other factors such as experience of teaching adults, time spent abroad studying, experience of the U.S. education system, being a native speaker of English, or having a minimum of 5 years of teaching experience (DOSB1: 317, DOSB2: 342, DOSL3: 346). It will be interesting to see whether this changes once the new TOEFL becomes operational.

Predicted Impact

> DOSC1: 517–521: How is it going to change for us as a center? The most important thing is the actual administration: the fact that we might be Internet-based, and that will change very much the way we work here. That's something that we're more concerned with.

It would be logical to believe that the changes in TOEFL will have a strong impact on resourcing, yet this might be too simplistic an assumption. We presumed that schools would have felt computer classes were a necessity when the CBT was introduced in 1998. This was not the case, as this study shows. Some schools offered classes without ever putting the students in front of computer screens, materials which demonstrate screenshots, and practice tests which are accepted by teachers and students alike as being near enough to the real TOEFL test. Will schools manage to do the same and adequately offer preparation for the new TOEFL? The answer is "Yes, perhaps," once a sufficient variety of materials is available—if judged by the present practice.

For the schools that do offer more highly computer-based courses another issue of resourcing will probably arise: that of teacher training. As seen, the computer abilities of the teachers in the sample varied greatly from institution to institution. The common thread running through the majority of the schools, however, was the lack of training that teachers received in using computers in general and using them in class in particular. The only institution to mention the availability of any systematic training was the best-equipped school (B2), yet even here the training was optional:

> DOSB2: 542–548:
>
> Interviewer: You've mentioned that some teachers are very computer literate, and they tend to do the multimedia classes, and some teachers maybe aren't so comfortable with the computers. Do you offer training if they're interested in getting up to speed with their skills? Can you offer that training?
>
> Director of studies: Yes, we do, we do—if they are interested.

We can imagine a future scenario where computer use might be seen as more of a necessity for adequately preparing students for TOEFL, and where, if teachers show the same level of confidence and competence generally observed in the study, the need for further training

will arise. A second scenario is that younger teachers who have this computer competence already will be hired in preference to those of an older generation who may well prove to be less competent in this area.

> DOSC1: 67–68: The notion of testing on the computer is still quite a big thing and then to do something where you are supposed to give an oral response—I think it's going to be like a lot to take on.

What will classrooms that consist primarily of computer practice look like? In the case of the two courses where computers are already used, what is seen is reduced use of spoken English as students become autonomous units, interacting with programs and working at their own pace; the teacher fills the role of adviser and helper if they request. The difference in methodology from the teacher-led lockstep classes observed at other sites is notable. Since preparation for the new TOEFL speaking test will also be needed, the teachers will have to organize their classes carefully. This assumes, of course, that the nature of the speaking test requires the development of interactional skills, which the current TOEFL Academic Speaking Test (TAST), understood to be a prototype for the new TOEFL speaking test, does not seem to do. If exercises, as at present, mimic test formats then we can expect to see a growth in commercial materials which are computer-bound, using more tasks which are compatible with a computer-mediated environment than tasks encouraging oral/aural interaction.

Summary

We have identified three concerns corresponding to the three groups interviewed. For students, there is not much worry about computer literacy in a general sense, as might have been predicted a decade ago, but there is some concern about the need to have efficient typing skills. With the ever-increasing use of computers for leisure, work and study, students generally seem prepared psychologically for taking computer-based tests.

For the teachers, it is a slightly different matter. The introduction of a very different format of testing may have an impact on their employment in the future, though this may be counterbalanced with demands for other skills, such as a high level of language ability. It is equally possible that, as now, the format of the commercially prepared materials will dictate

what is taught in class and the test simulation aspects of preparation will be left to the students to cover independently, outside of class hours.

For the directors of studies, issues of resourcing and the consequent reorganization of programs are the main concerns. Their dilemma is how best to gain a local market edge while bearing in mind the investment that will be necessary in any re-equipping. The effect on the directors of studies is not discussed in great detail here; however, there are ways in which they too are affected, in terms of resourcing provision and in terms of teacher training. As one director of studies pointed out:

> DOSB1: 401–405: It used to happen, a teacher coming and "Oh, it's not working" and "What's wrong again?" It's stupid. There's only a couple of buttons. Can you imagine them with a PC?

This director of studies clearly despaired of the technological abilities of some of his teachers. Yet, whose responsibility will it be to make them more computer literate—the employees' or the employers'? Again, for the type of institutions in our study, where commercial viability is all important, market forces are more likely to dictate what happens than pedagogic ideals.

Teacher Training

The teachers in the sample had a range of ability and experience, but one factor they had in common was that they had received no training in how to run a TOEFL test preparation course. Several teachers had had to work out a course outline and decide on an appropriate methodology themselves, with little or no guidance from others (TB2: 1494, TL2: 953, TR1: 984, TS1: 1334). This was not always an easy task, as this teacher made clear:

> TB1: 1040: . . . it was extremely difficult because it was my first approach to that test. I had not sat it, and it was very different to everything I've done before. I had to read all the books we had here. I had to go through the sites on the Internet and generally try to find and learn absolutely everything. It was very difficult.

We asked whether teachers used materials from ETS for guidance in their lesson planning. Most teachers claimed not to, but a few reported using them for content for their lessons: downloading information about the exam or lists of essay titles, and so on, from the Web

site. No one referred to any materials with more detailed information for teachers, such as suggestions for how best to structure a course or how to check that the skills and subskills tested on the TOEFL test were adequately covered. We also searched for material to establish the exact nature of the TOEFL test and to provide support for teachers but we did not find any such documents either.

None of the schools offered anything in the way of TOEFL-specific training. The closest was a workshop for teachers comparing the requirements of the IELTS essay with the TOEFL essay (DOSB2: 497). This was not surprising, considering that in most cases TOEFL courses constituted a relatively small proportion of the schools' operations. The schools were not likely to provide training for the small numbers of teachers involved with TOEFL, as this director of studies indicated:

> DOS21: 446–543: . . . most of the time, you're just concentrating on the cash flow. We need to survive. That's the most important thing. We have the young teachers who are very willing. I'm glad to see their enthusiasm. They want to develop and to get better and see different things but sometimes I say it's difficult to spend five hundred [local currency] on something when we need to something for the school like dictionaries or whatever. So you have to balance things sometimes.

On the whole teachers were expected to manage their own professional development, both in term of both time and finances:

> DOSR1: 522:
>
> Interviewer: So you can offer some training?
> Director of studies: No, no. They have to try and get their qualifications and do what
>   they can.

Observation was the only form of TOEFL training mentioned, but this only occurred in one institution (Site C1). New teachers were expected to observe the TOEFL teacher for several lessons before they were allowed to serve as substitute teachers. This was an informal arrangement rather than a planned training program, but it was more structured than any other training available in our sample.

The same school stood out in its approach to TOEFL preparation in another way, in that some self-initiated training was evident. The TOEFL teacher had identified some schools abroad that were offering TOEFL courses and realized she might learn something from visiting them. Her institution helped her to raise funding for a visit. She visited classes at another and talked to teachers and then returned with ideas for improvements (DOSC1: 384). This institution was unique in its efforts to develop its courses but its status as a nongovernmental organization and its links to the U.S. State Department (as an information center) may have accounted for this to some extent.

Not only had the teachers not received TOEFL-specific training, but most had not received training EAP (English for Academic Purposes) training either. Only one had experienced formal EAP training and was also teaching EAP at the local university. Most of the others were teaching classes at various levels at local language schools. They made references to other exam courses, most often relating to the Cambridge suite of exams, but they did not mention receiving training for these courses either. (However, DOSL2 offered a form of training to teachers teaching FCE classes, which attracted larger numbers than TOEFL classes. Teachers at this institution observed more experienced teachers before they started giving lessons.)

It is not the case that there was no professional development available to these teachers, however. The problem was that the content of general courses on teaching English were not considered relevant to the TOEFL courses by the teachers in the sample. For example, TC1 had undergone teacher training for state school and also attended various workshops and seminars, mostly provided locally by the British Council:

TC1: 1107:

Teacher: It seems these things are usually more useful for regular teaching . . . how to improve motivation, how to improve these skills or how to make your students talk, how to use their multiple intelligences. These are specific examples.

Interviewer: So not necessarily things that you can bring to the TOEFL class?

Teacher: No, absolutely not.

This exchange shows that the teacher saw TOEFL teaching as different from other types of teaching, so different that it did not require motivating activities or any of the other features she

mention. TP1A was in a similar situation, finding it challenging to adapt what he had learned on his recently completed TESOL certificate course. He had been taught to teach communicatively, but this did not match what he found in the TOEFL preparation coursebook he was expected to teach from. The mismatch was causing him some concern:

> TP1A: 167: I'd feel much more comfortable as a teacher if I knew how to teach this text more effectively. To convey the material. I don't know if there's preparation courses for teachers for TOEFL or not, but this is what I was given as material to use.

He did not yet have the skills to adapt his new techniques to the TOEFL situation, although he showed signs of using communicative activities when helping students to develop their writing skills.

What is interesting though is that other teachers made no pretence of teaching their groups communicatively. Observations of TL2 in a non-TOEFL class showed that she was quite capable of this kind of teaching (see Appendix L), but she had made a conscious decision to run her TOEFL classes differently. If the intention of the new TOEFL is to encourage positive washback in the preparation classroom, leading to more than more than simple simulations of the test and test-wiseness training then teachers (and their managers) will need to be convinced that teaching differently will make their classes more effective (and profitable).

Summary

The picture that emerged from the data was of teachers working mostly in isolation to develop a program for their students. This did not seem to be an issue of great concern for the majority as they judged the success of their course by their students' feedback and results. There was some indication that changes were being made to courses but this tended to be driven by availability of materials rather than by a reconceptualization of the methodologies employed. It is hard to predict whether teachers will feel the need for further training once the new TOEFL is introduced. They coped with the changes when CBT was introduced and continued to prepare courses in much the same way they had done previously. They will need to understand that their current methods are inadequate if they are to be convinced to change, and their institutions will need to understand that investing in training is worth their while. If training were available externally, via the Internet for example, and if it led to the school gaining a market edge in some

way, then this might influence the management's decision to encourage teachers to take up such an opportunity.

Communication

> DOSC1: 574: The average student has no idea about the ETS and why this is so big. They only see me. So I'm very much concerned with having the right information and on time, and that it's current, comprehensive and all that.

This section discusses issues relating to the communication of information about the TOEFL test. We were interested in finding out what the teachers, students and directors of studies thought they knew about the TOEFL and the channels of communication they used to gain this information. This is important because good communication is central to the process of the diffusion of innovation (Henrichsen, 1989). It is only by examining the channels of communication available before the introduction of innovation (in this case a new test) that we can understand why or why not it embeds itself satisfactorily into the target context.

Sources of Information

We might expect one of the main sources of information to be ETS itself as producer of the TOEFL test. ETS's means of disseminating information about TOEFL include the electronic medium, in the form of its Web site (http://www.toefl.org), and paper medium, in the form of the Bulletin, which is sent to everyone who registers to take the test and is available on request to others. The practice materials available from ETS (e.g., POWERPREP) also serve as a means of familiarization with certain, but limited, aspects of the test, such as possible item types and the length and the order of the sections. These formats offer broad coverage, available to everyone worldwide (wherever there is Internet access at least). Less wide-ranging coverage exists in the form of conference presentations. The teachers referred to the ETS Web site to some extent, and to the Bulletin, but none of them mentioned conferences. Their economic position (based on extrapolation from what we knew from our work on other projects in the CEE region) would make it difficult for them to attend conferences, especially international ones, unless financial support was available. As mentioned in the Teacher Training section in this report, private institutions are often not willing to fund such training activities, especially if the teachers' courses make up only a small percentage of their overall business. There were three further

sources of information mentioned: personal communication, the Internet, coursebooks, and other practice materials.

Personal communication. Personal communication was as important, if not more important than, direct communication from ETS. Two teachers referred to their director of studies as their main source of information (TB1, TS1), and TC1 owed much of her understanding to her mentor, whom she had observed teaching and been able to talk to before taking on classes herself (this is discussed in more detail in the Teacher Training section in this report). The most commonly mentioned source of information was former students, who gave their teachers feedback after sitting the examination (TB1, TB2, TL1, TR1). The teachers seemed to rely on this feedback to inform their decisions about what to focus on in their next TOEFL courses, as this teacher explains:

> TL1: 643–649: I've got a pretty good idea of what they're [ETS] looking for, just from the students of the past that have taken it. I know what marks they got. I know what the question was and I can kind of gauge what they need to be doing. You know, I've read twenty essays that they've written so I basically know how good their writing is and probably what they did on the test, so after 5 1/2 years I've got a fairly good idea what TOEFL's looking for.

This dependence on student feedback was a significant feature of the whole group of teacher informants.

The Internet. Another commonly mentioned source of information was the Internet, but not necessarily the official ETS Web site. In fact, the informants often had trouble remembering whether the sites they had visited were ETS sites or not:

SL2: 1071–1074:

Interviewer: You've been to the official TOEFL Educational Testing Services Web site as well as some different ones, I understand?

Student: Well, I don't remember exactly because there are so many Web pages.

DOSB2: 928–940:

Director of studies: I heard about the new version of the TOEFL a year ago.

Interviewer: And can you remember what, where you heard about it?

Director of studies: Yes, I was browsing the Internet, and I just read. Perhaps it was the official TOEFL site or some other site, or, I don't know, some institution that offers TOEFL classes, from around the world, not in [this country]. But I read about this, and I thought, well, that's a good idea, but I just don't have any more information than that.

Both teachers and students used a range of sites offering information, tips, practice tests and language development or checking exercises. The quality of material on such sites will naturally vary and it would be worthy of further investigation to establish what criteria students and teachers use to select which sites they feel are worth using and which they reject.

Coursebooks and other practice materials. Even more cited than the Internet were the various commercially produced preparation materials, which every course was using to greater or lesser extent. As was the case with Web sites, the informants did not necessarily distinguish between official ETS materials and the ones produced by various publishing companies in both the United States and the United Kingdom. (See Appendix P for a list of the coursebooks used by the informants.)

Only one of the teachers had information about TOEFL from having sat TOEFL herself (TL2), but others had worked thoroughly through the practice tests provided by ETS:

TL1: 1127–1135 (see also TB1):

Teacher: I can explain a bit of the question or say what they want you to do in this bit is to put a sentence into a paragraph and that's all you have to do.

Interviewer: And you've done it yourself.

Teacher: And I've done it myself so I can explain it. That was very helpful for me because I didn't want to be sitting there with a student asking me what do we do here: "Oh, I've no idea."

Another important point arising from the data was how students found out about TOEFL before they even got to their preparation classes. One important source of information was friends:

SS1: 631–635: The first question was like "Where can I find that information?" So because I didn't find too much information for example in magazines or media or

something like that I had to ask some of my friends of mine so they gave me some advice like go to the agency.

The other was, as mentioned above, the Internet:

SP1A: 1022–1028:

Interviewer: What's your main source of information about the TOEFL examination?

Student: Actually it was the Internet and the TOEFL organization Web page and my friend who is going to take the same exam. She was just very interested in it and she was searching for some materials about TOEFL so I know I think quite much from her exactly.

Level of Familiarity With TOEFL

In some of the countries we visited, TOEFL was generally more well-known than other examinations for reasons which were not entirely clear:

TB1: 273: The difference is that TOEFL is more popular and students probably from the eighth and ninth class at school know what TOEFL is because all their relatives or brothers or sisters have sat for that exam and they know everything about it.

This extract suggests a self-generating cycle in that as more candidates hear of the TOEFL test, more decide to take it, therefore more hear of it, and so on. Another important factor may be the presence of a branch of an American university in that particular town. A director of studies in another country stated, however, that all the teachers knew IELTS better than they knew TOEFL. She felt this was the result of a good dissemination program by the British Council and questioned why there was not a similar drive on the part of ETS. What was significant about this opinion was that the director of studies was someone who was in reasonable contact with ETS, the center being affiliated with ETS and acting as an ETS test center, as well as an information dissemination center. She was part of the dissemination mechanism but still felt dissatisfied with dissemination efforts (DOSL2: 875).

Shelf Life

An area where there was some confusion, and which had implications for how the TOEFL's cost and value were perceived, was the test's shelf life—how long the test result was

valid for. One misconception appearing several times was that if a test result had a fixed life span (2 years in the case of the TOEFL test) the individual had to keep taking the test regularly, even after they were accepted at and actually began attending a university (SP1A: 1491). The short shelf life of the TOEFL test (in the students' eyes) was reported to be deterring potential candidates (DOSC1: 988). Ironically, this shows a lack of awareness of other tests (such as IELTS), which also, rightly in testing terms, have a limited shelf life.

Dissemination of News About the New TOEFL

Most of the teachers reported having heard that changes were to be made in the TOEFL test, but three (i.e., nearly one third) had not heard anything. Those who had heard something quoted a range of sources, including the ETS Web site, another Web site dealing with various tests, the local British Council office, a director of studies, and students. All of the directors of studies referred to the changes in one way or another, and again they cited a range of sources, including ETS via the Web site (two had heard this way) or conferences. Whereas no teachers reported attending any conferences, three of the directors of studies had been to the NAFSA Conference (Association for International Educators, based in Washington, DC), and one had particularly sought out the ETS presentations. Others had heard from their local Fulbright office. Two directors of studies first heard of the new tests when we contacted and invited them to participate in the study. We found it somewhat surprising that they had not heard anything earlier, considering that the test was supposed to be introduced in less than 2 years' time and its introduction would have implications for planning and allocating resources. Some students had not heard anything about the new TOEFL, but this was not surprising, as the changes would (probably) not affect them directly.

Information Centers

It appears that access to up-to-date information about developments in TOEFL is not uniform across this part of Europe, and the process of dissemination relies primarily on Web sites being accessed by interested parties. In some areas, education information centers are in operation, but their coverage is not even across the region. Two of the sites were information centers, but information about the location and function of such centers is not easily found on the Internet (their function was only discovered when visited). They reported good contact with ETS, but they only had limited resourcing to distribute the information they received and seemed

to respond to requests for details from the public rather than being proactive. One of the managers explained it thus:

> DOSC1: 623–635: In some places you only have like a one-man show—like one person who does advising within the Embassy, and then they would know a lot less than we do, particularly by virtue of being within the embassy and very few people coming in. And then you have Fulbright centers in countries where Fulbright programs work, where they have a commission of their own, like in the UK they have in London. . . . And also our model, which is quite uncommon—like a NGO (nongovernmental organization), which is independent and has taken on this as one of our activities. It's not all we do but this is one of our activities. So you can expect that after all that I'm telling you, you expect different levels of knowledge about the testing and so forth. It depends on a lot of factors.

The second information center manager was clearly concerned that she was limited in her ability to get information out to relevant groups of teachers and students and she made some suggestions about how this situation could be improved:

> DOSL2: 790–797: We will have to prepare more handouts which will explain how these tests will differ from each other and . . . I will have to add to my international education week program a special session for English teachers to inform them about changes. . . . Already I did this but as I do not have a special session for teachers just a course occasionally. When I met English teachers in different schools I just told them the TOEFL is going to be changed.

Summary

Figure 4 shows the sources that contribute to teachers' and students' understanding of the TOEFL test. To summarize, the ETS Web site is the most widely used of the sources of information directly emanating from ETS. Teachers also count on information from other Web sites and from their past students. Students-to-be count on various Web sites and their friends. The most influential source of information, however, is the TOEFL preparation coursebook (of which more has been said in the Role of the Coursebook section). As Alderson and Hamp-Lyons (1996) found in their study, it is difficult untangling the effect of the coursebook from the test. There are thus implications about the accuracy and focus of the message received if it is not

conveyed directly from the test producers, but through an intermediary who may have less than full understanding of the purpose, construct, and aims of the test.

```
                              ETS
        ┌─────────────┐              ┌─────────────┐
        │   Bulletin  │              │  Web site   │
        └─────────────┘              └─────────────┘
                  ┌──────────────────────┐
                  │  Practice materials  │
                  └──────────────────────┘

┌─────────────┐                              ┌─────────────┐
│ Colleagues  │                              │   Friends   │
└─────────────┘                              └─────────────┘

      ┌─────────────┐          ┌─────────────┐
      │  Teachers   │ ───────► │  Students   │
      └─────────────┘ ◄─────── └─────────────┘

  ┌─────────────┐      ┌──────────────────┐    ┌─────────────┐
  │ Other Web   │      │ Commercial course-│   │ Agencies (e.g.,│
  │   sites     │      │     books        │   │  Fulbright)   │
  └─────────────┘      └──────────────────┘    └─────────────┘
```
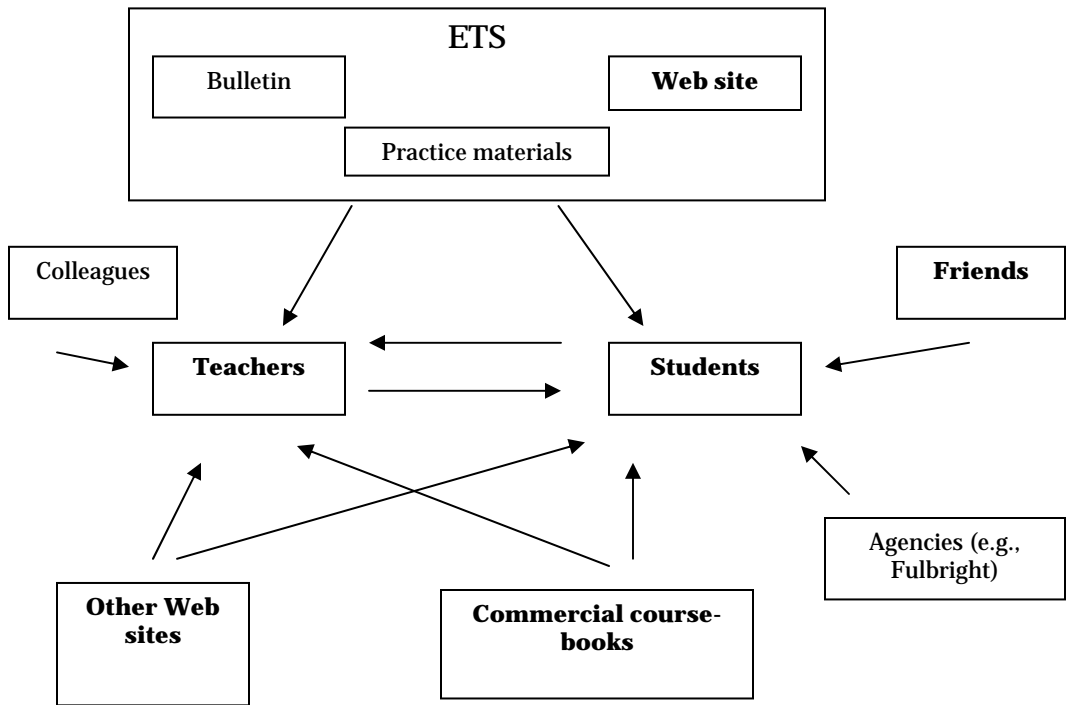
Figure 4. **Sources of information about TOEFL.**

Note. The most widely used sources are depicted in bold.

Summary and Conclusions

As mentioned at the beginning of this report, the purpose of this project was to conduct a baseline study which would give an accurate picture of what teaching and learning were like in TOEFL preparation courses in Central and Eastern Europe. This would provide a point of comparison for a later study that would investigate whether the introduction of the new TOEFL test had had any impact on the teaching and learning taking place in such courses. We wished to answer three questions in the baseline phase of this investigation:

1. What was the intended impact of the new TOEFL?

2. What were the characteristics of TOEFL preparation classes before the introduction of the new examination?

3. What further steps were needed in order to track the impact (or lack of impact) that the new TOEFL test would have on teaching and learning?

The purpose of asking the first question was to help us to understand what features we needed to look for as we interviewed teachers, students, and directors of studies, and as we observed TOEFL preparation lessons. The Intended Impact of the New TOEFL section in this report describes the steps taken to get this information and summarizes the findings. First, we looked for statements about desired impact in the TOEFL 2000 framework documents. There were few of these and they were mostly quite general (e.g., there would be more communicative teaching). There was mention though of a specific change in the teaching of writing (that the writing would be more academic) and some hope that students would learn to communicate orally. We then wrote to experts who had been involved in the conceptualization phase of the new TOEFL, to ask what they recalled of early discussions regarding the impact of the new examination. Much of what they reported was also at a general level (e.g., changes in test-preparation practices, improved academic language and skills), but several mentions were made of integrated skills teaching and a greater emphasis on speaking. This indicated that we should look for both integrated skills work and the teaching of speaking when we visited the target institutions in the data-collection phase of this study.

We also constructed a table comparing the two current versions of the TOEFL test (the paper-based and computer-based versions) and what we knew then about the new TOEFL. We described this undertaking in the Methodology section of this report and presented it in tables in Appendix F. Once this content was approved by ETS, we took what we understood to be the most important features of all the tests and incorporated them into the data-collection instruments—in particular, into a checklist at the core of the observation schedule. In this way, we made sure to note whether certain features that were already present in current teaching would be present in future classrooms after the settling-in period of the new test.

In order to answer the second question, we needed to design data-collection instruments and procedures for data analysis. These are described in the Methodology section of this report. The theoretical framework was Henrichsen's hybrid model of the diffusion/implementation process (Henrichsen, 1989) as modified by Wall (1999). The aspect of the model of most interest during the baseline study was the antecedents section. This relates to the conditions in place prior to the introduction of an innovation—in this case, new TOEFL.

The sample consisted of 10 institutions in six countries. We conducted interviews with 10 teachers, 21 students, and 8 directors of studies, and we observed nine TOEFL lessons and five non-TOEFL lessons. The bulk of the present section is devoted to summarizing the findings regarding classroom practice: recounting the most important features in the teaching of the skills that have been and will be tested on the TOEFL test. Also summarized are the findings relating to five further themes that emerged from the data. These summaries are related back to the Henrichsen model near the end of the section.

Regarding the third research question, the baseline study offers a useful picture of current practice but we will need more understanding of teachers' abilities and attitudes, the teaching context and other factors in the Henrichsen model before determining whether any changes occurring after the introduction of the new TOEFL can be linked back to the new test constructs and format. This will also help to explain why any changes take the form they do. Our recommendations for Phase 2 of this research—the transition study—are presented at the end of this section.

The following subsections are summaries of the baseline findings.

Teaching Listening

Teachers spent varying amounts of time on listening in the classroom, with some preferring to tell their students to do the coursebook exercises at home. Those who taught listening mainly went through the same exercises, sometimes playing the passages again when students did not understand and sometimes asking the students to study the transcripts. Some teachers distributed lists of synonyms and/or Americanisms to help the students to expand their vocabulary. One of the few strategies they suggested was "listening for key words" but they offered little concrete advice about how to do this successfully. All but one of the teachers conducted their classes in English, with several advocating this as a way of giving additional listening practice. All advised their students to practice listening to authentic material in their out-of-class hours, but they gave little to no advice about how to take advantage of this input systematically.

The main point emerging was that teachers did not use techniques for teaching listening in a nontest-like way. There was also a belief that the current TOEFL test (CBT version) tested memory since questions are presented after students hear the passages. Students are thus unable to practice strategies such as listening for gist or detail, as they do not know what they need to

listen for in advance of hearing the passages. It is not clear whether this is why few such techniques are practiced in class or whether the teachers were simply not aware of the techniques students could be using.

We concluded that teachers understood what the Listening test tested in general terms, but they could not break Listening down into a set of teachable subskills, in the way that they seemed able to for reading. It is as if they assumed or hoped that their students would improve through some form of osmosis.

Teaching Structure

Most teachers counted on the fact that their students had studied a lot of English grammar in the past, and many saw the attention they paid to language structure in the classroom as being a matter of brushing up only. There were exceptions to this, however: teachers who felt that the grammar that students had to deal with was too difficult or even tricky. It is hard to generalize about how much time teachers devoted to the study of grammar, but what work they did do usually consisted of teacher explanations of particular language points (not always easy for native speakers of English, who had not studied grammar systematically), going over the exercises in the coursebook, occasional drilling, and focusing on grammar when giving feedback on student essays. The coursebooks did not give much guidance on how to teach this area more effectively (Alderson & Hamp-Lyons, 1996), offering only on a succession of practice exercises and explanations.

It will be interesting to see whether the teaching of language structure survives after the introduction of new TOEFL, since it will not be tested separately in the new examination. If it does continue, it will be important to record whether the teachers' techniques change or whether they will still use materials that amount to little more than mock test questions.

Teaching Reading

The teachers generally seemed confident in their teaching of reading and were more able to talk about the subskills involved than they were able to do for listening. The notion of four types of questions (main idea, inference, etc.) was prevalent and seemed to be a legacy of the way reading was handled in the exam preparation coursebooks. The teaching of reading followed the same pattern as for listening and structure, with students working through coursebook exercises and answering questions on their own, and then offering their answers to the rest of the

group and listening to the correct answers given by the teacher. Reading aloud was also observed in some of the lessons. There was no evidence of prereading or any other reading development techniques encouraged in communicative coursebooks.

Some teachers relegated reading exercises to homework since the answers to the questions were often found in the coursebooks. This did not ensure that students learned to read quickly, but this did not, in any case, appear to be a priority in the classrooms we visited. All of the teachers encouraged their students to read widely in their out-of-class time.

The teachers' main worries were that technical vocabulary presented special problems (this may have been a coursebook-generated problem rather than a test problem since, according to Hale [1988], TOEFL test passages are nontechnical), that the wording of questions was often hard to understand, and that students sometimes had problems understanding why their interpretation of passages was not correct (Pierce, 1992).

Vocabulary development figured highly in the teaching of reading for TOEFL purposes. There was a concern that some questions that were meant to test the understanding of meaning in context were, in fact, testing straight vocabulary knowledge. We do not know whether the coursebooks included faulty questions that led teachers to believe that the test itself was problematic.

Teaching Vocabulary

There was a unanimous belief that having a rich vocabulary was crucial to success on the TOEFL test. It was not clear how the words that were presented in the coursebooks were chosen (i.e., frequency lists, corpus analyses), but teachers readily compiled lists of words and phrases (often synonyms) for their students to learn by heart. Some teachers emphasized vocabulary during the reading lessons, sometimes asking students to read passages at home and learn the words to allow the lesson to proceed more smoothly (which may have defeated the goal of learning to read for a more definite purpose). Besides memorizing lists, students spent time doing vocabulary exercises on the Internet and using CD-ROMs created for the purpose.

It was not clear why so much attention was given to vocabulary because it has not been tested separately since 1995 (Read, 2000). The fact that it figures prominently in the TOEFL rating scale for writing may have something to do with this, as may the belief that some teachers held that it was not always possible for students to guess the meaning of words in context because there was not sufficient context available to them.

Teaching Writing

The teachers' choice of content was heavily influenced by the preparation coursebooks, with a focus on paragraph structure, the importance of introductions and the 4-or-5-paragraph essay formula. There was some feeling that once students learned the formula for introducing a topic and arranging paragraphs, their writing improved quickly. Teachers paid less attention to the content of writing than to the organization, although some of them talked about their students' inability to argue succinctly. Quite a bit of class time was devoted to writing because many students had not received much instruction in their earlier schooling, but the students were generally expected to produce their essays at home in their own time. (There is little evidence of students writing under timed conditions. Some concerns were raised by students about having to write using computers due to poor typing ability, especially if their own language was written in another script. This feeling was expressed even though students can currently choose whether to hand-write or word-process their essays.

Most teachers were aware of the TOEFL rating scale for writing, but they did not use it when marking their students' writing. They might introduce it in class and do some scale-familiarization exercises with them, but any further work had to be done by the students working independently, either marking their own writing or that of their peers. Teachers tended to give ad hoc feedback to their students, focusing on problems that occurred during the production of a particular piece of writing, with some native speakers of English commenting that they marked as college professors would.

One of the key questions regarding writing was why the teachers did not make more use of the TOEFL rating scales. For some it was a matter of confidence in being able to use them as intended, which was probably due to a lack of training. Teachers were familiar with the concept of scales (some used them when preparing students for other exams), but some process of standardization needed to take place before they could be sure they were marking correctly. Merely reviewing sample essays at different levels, without an opportunity for debate or questioning, is not enough to provide teachers with the appreciation they need of TOEFL standards. Further training might also be useful to help teachers to use the scales in the classroom. Scales have their role in the writing development process, but there was not much evidence of them being used this way.

Another important finding had to do with the use of the formula approach to teaching writing. The idea of a formula may appear simplistic but it seemed to satisfy the students. This approach did not emphasize the accuracy of content but whether each part of the essay fulfilled a particular function. It will be interesting to monitor whether this changes when integrated tasks are introduced in the new TOEFL, when content (quality of ideas) will have to take on greater importance.

Teaching Speaking

Most teachers did not focus on the speaking skill and there seemed to be a strong connection between this and the fact that speaking is not tested on the current TOEFL test. However, all but one of the teachers conducted their classes in English. They did this for various reasons: the reputation the school had for teaching in English, the fact that some teachers were native speakers of English and did not know their students' language, and the view that some teachers had that this would provide extra input to the students and help them to develop their listening abilities. The fact that speaking would be tested in the future was the only thing that many of the teachers and students knew about the new exam, and the teacher who taught in the students' first language declared that she would have to switch to English in the future.

It will be important to watch how teachers manage to fit focused speaking work into an already crowded timetable. It is reasonable to assume that they will need to reduce their attention to other skills in order to make time for speaking. With the introduction of integrated skills testing it is hard to imagine they will want do less reading, listening, or writing. It will also be important to see what kinds of speaking they choose to work on, since a test which is similar to the current TAST test will not require much in the way of interaction or sociolinguistic skills.

Assessment in the Classroom

In one sense, the teachers were assessing all of their students all of the time, since most of the classroom work they did involved going through coursebook exercises or practice tests and making judgments about whether students answered questions correctly or incorrectly. However, there were also occasions when more formal assessment took place, for the purposes of screening, diagnosis, and practice. Screening tests were sometimes given before the start of a course to determine whether students had enough proficiency to enter the TOEFL courses. These

seemed to be of little consequence in some cases, however, as students who were determined to attend a course would not be barred on the basis of poor scores.

Diagnostic tests were administered at the beginning of the course to give students and teachers an idea of what their general learning priorities should be (e.g., Which skill were the students best at? Which skills were the weakest?). They did not diagnose particular problems though or give teachers an indication of what exactly they should focus on when working on certain skills.

Practice tests were taken at certain points during the course to give students a general impression of how they were doing. They did not necessarily measure true progress, if this implies a movement from easy to difficult. The coursebooks that made up the syllabus of the courses seemed to be based more on the notion of an accumulation of language and practice rather than a progression from less to more difficult (Hamp-Lyons, 1998).

Most of these tests came from the preparation coursebooks. We do not know of any empirical work done to establish how valid they were or how similar they were to the real TOEFL. We also have reservations about their reliability, given that they were administered under a range of conditions unlike the TOEFL test. The tests seemed to play a useful role as a means of test familiarization, especially if they were done on computers. We stress, though, that they can only give the students some indication of their level of preparedness for the TOEFL test rather than a true picture of their language proficiency.

The Role of the Coursebook

Most of the teachers were highly dependent on their preparation coursebooks, and had learned what they knew about the TOEFL test by working their way through these publications. It is hard to determine what role the actual test plays in shaping TOEFL teaching and learning since nearly everything is mediated through the coursebook (see Alderson and Hamp-Lyons' 1996 discussion of the difficulty of trying to separate test effect from textbook effect). It is important to recall that the coursebooks are not always selected because of their pedagogic value, but because of price or other pragmatic considerations.

The coursebooks contain the basic information that students need about the TOEFL test (information about the overall structure of the test and item and task types, texts and tasks, general advice about what to focus on) as well as providing answer keys, but they generally do include advice to the teachers about how to teach the target skills effectively. They do not

encourage deep understanding of the TOEFL test construct or of ways of teaching outside the teacher-centered lockstep completion of exercises and checking answers. TOEFL courses are above all revision courses (apart from when they deal with writing skills, as discussed), so it would be useful if the coursebooks offered suggestions about how to deal with language students have been introduced to elsewhere. We did not come across any coursebooks that provided ideas for revising language in a thought-provoking manner.

The Role of the Computer

Two of the courses were not preparing for the computer-based version of the TOEFL so it was not expected that computers would figure highly in their teaching operations. Only two of the other courses made any use of computers in their classrooms. Some institutions had computers available for self-access work, but one not only lacked computer facilities but also other technological teaching equipment such as CD players.

It was assumed by teachers and confirmed by students that the latter had sufficient experience of computers to feel comfortable working on them. Most students seemed to have access to computers outside the school, and were capable of using the Internet and commercial CD-ROMs to practice language and certain skills. One area of concern for the students was being able to get around the keyboard quickly when typing essays.

Not all of the teachers felt confident in their own computer skills. Some doubted that being more competent would help them to teach more effectively, but this may have been because they could not envisage the opportunities that more expertise might open up to them. Some teachers seemed to be able to choose not to use computers in their classes even though their institution had adequate facilities. It is not known whether this will remain a possibility in the future or whether computer competence will be a criterion for job selection.

Most CBT courses seemed not to be affected by the fact that the test was administered on computer, using paper-based materials in the classroom. We wonder whether the changes in the new TOEFL will be dramatic enough to have resourcing and training consequences for the institutions, or whether teachers and students will continue spending class time working out of the coursebook, with students practicing doing the test on their home computers.

Teacher Training

The institutions we studied provided little, if any, professional development work in EFL or EAP. None of the teachers had received any training in how to teach for the TOEFL test. They had had to develop their teaching approach on their own, using materials from ETS or, more frequently, studying the preparation coursebooks. One school did offer a new teacher the opportunity to observe an experienced teacher and discuss problems related to TOEFL teaching but this was the exception rather than the rule.

We concluded that one of the key reasons why coursebooks played such a prominent role in the classrooms was this lack of TOEFL-specific training. Such training could offer ideas about different methods and techniques, and perhaps provide an alternative to the teacher-led approach observed in almost all of the classes.

Communication

Teachers received their information about TOEFL from various sources, including ETS (the Web site, the bulletin, various books and learning packages such as POWERPREP), other Internet sites, former students, colleagues and, most importantly, the preparation coursebooks. Only one teacher had firsthand experience of taking the TOEFL test. Others had to trust that the messages they were receiving were accurate and up to date. Students got their information mainly from the Internet (though not necessarily the ETS Web site), education information centers or organizations such as the Fulbright offices located in many capital cities, and friends. The centers received information direct from ETS, but they did not always have enough staff or other resources to spread news about changes in the examination to teachers or students in the area.

Three of the ten teachers did not know that the TOEFL test was about to undergo a change, even though it was, at that time, less than 2 years before the examination was due to go live. The remaining teachers knew only that there would be a speaking test. All of the directors of studies were aware that there would be a change, but some of them were alerted to this fact by us, when we contacted them and asked them to participate in the study.

We were surprised by this general lack of awareness, which persisted throughout the data collection period and also to the end of the baseline study (June 2004). This had implications in terms of the institutions' ability to cope with the impending changes and to plan ahead successfully.

We were also surprised by the number and types of sources the participants used to get information about the current TOEFL. There are always possibilities of dilution, if not distortion, of information, if messages arrive through unofficial channels. This also has implications for planning since effective decisions can only be based on accurate information.

Application of the Henrichsen Model

It is evident from these summaries and the sections in this report they are based on that TOEFL teaching did not look the same in all the institutions; however, there were some themes that appeared frequently. This subsection focuses on what the Henrichsen model can explain about the new TOEFL based on the information in this report. See Figure 1 on page 5.

Antecedents

As mentioned in at the beginning of the report, the function of a baseline study is to fill in the details which correspond to the part of the Henrichsen model called *antecedents*: the situation existing before the introduction of an educational innovation. What has been presented in the summaries of the Teaching Listening, Structure, Reading, and Vocabulary sections of this report corresponds to the traditional pedagogic practices part of the antecedents section: what teaching looked like before the key participants knew (very much) about the new TOEFL examination and what it would require of them.

We also tried to collect data to fill in the characteristics of the user system and the characteristics of the users sections of the model.

For *characteristics of the user system,* we gathered information about each of the sites, including what kind of institution each was (private-language school or education information centers), what the administrative/management was like (was it part of a chain or an independent operation, large or small), what the resourcing was like (principally computers and other technology, but also resourcing for investment in training), and what the classroom conditions were like. We could see that some of these factors may have an influence on the way the new TOEFL test is dealt with in the future (e.g., that the education information centers have the potential for being better informed about new developments because they have more direct links with ETS than the private schools). The institutions that already use computers may find it easier to switch over to preparing for the new TOEFL test if the commercial materials that will undoubtedly appear soon are CD-ROM- or DVD-based rather than book-based. Larger

institutions, which offer a variety of courses, could conceivably use some of the profit gained from other courses to contribute to funding for the training of their teachers of the TOEFL test. An important factor in the user system is the director of studies and his/her awareness of and interest in the new examination. It is difficult to know at the baseline stage what factors will matter most since it is not until the process stage that they begin to interact with new factors such as *factors within the innovation* (what the new test looks like) and *characteristics of communication* (how messages concerning the new test are disseminated).

For the characteristics of the users section of the model, we gathered data on factors such as the teachers' experience and training (native-speakers or nonnative speakers of English; their qualifications for teaching in general, EFL, EAP, and TOEFL; novice teachers or very experienced ones), teacher and student awareness of the current and future examinations, their attitudes towards the current examination, and their technical abilities and their goals for the future. We feel that some of these factors made a difference in how the teachers were handling their classes during the baseline phase—principally whether the teacher was local to the country or an expatriate (expatriates tended to be more confident in their views of what was important in the target language situation, and less confident about their ability to explain grammar), whether the teacher felt that the classes should be exam-focused as opposed to more generally humanities-focused, and whether the teacher was computer-competent or less confident in handling technology. We were able to observe half the teachers teaching non-TOEFL classes, which gave us an opportunity to see whether there were aspects of personal teacher style which might override the impact of the new test or which might themselves be overridden. All but one of the teachers taught in the same way regardless of whether they were teaching for the TOEFL test or not.

We also looked at some of the characteristics of students, including their occupations outside the TOEFL class (whether they studied at university or were working), their ambitions for the future (mainly what and where they intended to study), and their language and computer skills. We were not able to ask individuals many detailed questions as the interviews had to be done in groups of two or three, but our impressions from the data are that no factor stood out from any others as likely to make a difference in how students approach their study for the TOEFL. It may not be realistic to expect to learn much about student learning when we had such

116

limited exposure to students, and we might need to rethink our means of getting information if we are to get more than general insights into how students prepare for the new examination.

We do not have much information about the experience of previous reformers' section of the model as we devoted all of our resources to finding out about the user system and the users. We have noted, however, that the fact that the TOEFL test adopted a computer-based orientation in 1998 seems to have made little difference in most of the institutions we visited. What types of change will have to appear in the new test before changes have to take place in the classroom?

Process

Although this study was meant to capture baseline data, it has also dealt with various aspects appearing in the *process* section of the Henrichsen model. The Intended Impact of the New TOEFL section in this report presented information about the source of the new TOEFL test (this chapter discusses the framework documents and the views of the test development experts). Appendix F presents what was known of the message section of the Henrichsen model. The Communications section in this report covered characteristics of communication and awareness of the receivers.

We are not able to comment on ETS's plans and strategies for dissemination of the message as we did not have access to this information.

Since ETS began releasing information about the new version of the TOEFL test, it has become possible to begin work on collecting further data corresponding to the remaining areas of the model. The style and nature of the message are now more concrete and are not just announcements that a change will take place in the future. This will allow the *receivers to be studied* as they learn more about and think more about the new TOEFL test. This will form part of Phase 2 of the project—the transition study.

We will look at factors that facilitate or hinder change to understand the nature of the innovation itself. This will be possible now that sample tests and more information on the new TOEFL have been made publicly available. The initial work on the message section of the model (Appendix F) can now be verified and expanded since there are concrete materials to react to.

It will be particularly important to look again at the characteristics of the users, which Wall's amended version of the Henrichsen model argued needed to be included in investigating the process stage (Wall, 1999). We feel that the students that we studied, for example, are representative of upwardly-mobile young people in this region, but the recent opening up of the

European Union to include three of the studied countries (Lithuania, Poland, and Slovakia) and young people's consequent ability to not only study but work in the UK and other countries where business is conducted in English, may mean an even greater change in the population hoping to do well on the TOEFL test. This is in line with the advice given by Taylor et al. (1998), that important changes in the candidature may take place over little time.

It is also important to keep track of changes that occur in the user system, features such as classroom conditions, school factors, and educational administration, as well as political, geographic, economic, or cultural factors. To discuss causality of any outcomes, we need to monitor where any factors have altered.

Not all aspects of the model can be completed in full, however. One aspect in particular, characteristics of the resource system, would involve looking at the way the introduction of the TOEFL test has been managed within ETS (e.g., Does the system have the resources it needs to introduce and successfully deal with the innovation? Do all parts of the system work in harmony?). We do not have access to this kind of information and therefore feel that this part of the model is beyond the scope of this investigation.

Once the examination has been administered for the first time, it will be important to visit each of the institutions again, to interview all the key participants and observe teaching, in order to assess whether any significant change has taken place since the baseline visit. In addition to questioning teachers and students about their beliefs and practices, we will interview the directors of studies to find out how conditions within the institution affect the curriculum and teaching program. Special attention should again be paid to the roles of the coursebook, computers, teacher training, and communication in the shaping of the impact of the examination. With this kind of data we can examine elements in the final stage of the Henrichsen model—the consequences.

Findings in Context

It will be important in Phase 3 to relate findings back to other studies in the area of test impact. It is particularly important to mention the work of Alderson and Hamp-Lyons (1996), whose study of TOEFL preparation classes in the mid-1990s deals most closely with the issues dealt with here. Although many of our findings match theirs, in particular the teachers' reliance on the preparation coursebook, there are some differences that must be noted. The teachers in this study were, on the whole, not as negative about the work they were doing in the TOEFL

118

classroom as the teachers in the earlier study. Those teachers admitted that they felt guilty because they did not put much time or effort into the preparation of their lessons. The teachers in this study were not asked directly whether they prepared less for the TOEFL classes than for their other classes, but their reflections about what they were doing and why leads us to believe that they took their job very seriously.

Two teachers expressed dissatisfaction with having to restrict their creativity, and a couple organized activities that did not resemble "typical TOEFL teaching;" but, those who kept within the pattern of closely following the coursebook seemed fully prepared and gave every impression that they had thought carefully about how they would carry out their teaching duties.

Perhaps the biggest difference was that the teachers in this study did not find it difficult to accept that their job was to get students through the examination: they stated, and their students stated, that getting through the examination was the aim of the teaching—not (usually) preparation for life in an English academic institution and certainly not language development. It was recognized that the students had had years of English study before coming to the TOEFL classes, and their purpose was to learn enough to do well on the TOEFL test. The students were clients paying for a service, rather than (as was the case in the Sri Lankan impact study, for example, Wall, 1999) students in a school system where the curriculum objectives were much greater than the objectives of the final examination and where examination caused "curriculum distortion" (Vernon, 1956; Smith, 1991). In the case of these institutions, the contents of the TOEFL test *were* the curriculum, and examination preparation was the students' only reason for being there.

Predicting Impact

Most of the views the TOEFL advisers gave to our question about desired TOEFL impact were quite general. They did, however, express more specific hopes regarding the effects the test might have on the teaching of speaking and of grammar (see Table 3). For instance, they felt there would be more emphasis on speaking and less on the study of decontextualized grammar and vocabulary. This section discusses our predictions regarding these two language areas, using what we know of TOEFL classes after the interviews and observations.

Teaching speaking. The advisers' comments seemed to assume that speaking did not figure highly in pre-new TOEFL classrooms. We were surprised to see how much spoken English there actually was, but it was used for classroom management purposes and to give

students the opportunity to practice listening rather than being focused on for its own sake. It is possible that the amount of spoken English will not change after the introduction of the new TOEFL test; but, we will need to monitor *how* it is used by teachers and students. We predict that there will more speaking development work than there is at present, but it is not yet clear what aspects will be stressed and what methods will be used to help students to perfect their skills. It is likely that teaching will mirror the tasks on the new speaking test, judging by how closely tasks in the majority of TOEFL classes in this study reflected what appears on the current TOEFL test. One of the specific expert predictions was that there would be more study of the pragmatic force of utterances. We feel this unlikely, given the nature of the practice materials that were beginning to appear as Phase 1 of this study was closing.

Teaching grammar. The second theme to look at carefully will be the teaching of grammar. One of the expert predictions was that there would be a reduction in the study of decontextualized grammar (and vocabulary), due to the nature of the new writing tasks that would require work at the discourse level. Another factor affecting the fate of grammar is the disappearance of the independent Structure test. These influences notwithstanding, we predict that teaching grammar will continue to feature highly in classes. This is based on observations of how much vocabulary teaching was still taking place, despite the fact that vocabulary is no longer tested separately,

Use of the textbook. We predict that commercially produced TOEFL preparation books will continue to be enormously influential. What will be interesting to observe in the next few years is whether the coursebooks that take the place of the current books will offer any more in the way of teacher guidance, especially concerning the constructs underlying the new TOEFL test and different techniques that teachers can use to develop their students' skills (especially speaking and tasks requiring skills integration). In order for this to happen, it will be necessary for the coursebook producers to have good insights into what the TOEFL test is meant to assess. This leads back to the importance of communication from ETS and the need for TOEFL-specific teacher training: two of the areas of concern that emerged as themes from this study.

Use of computers in the classroom. The use of computers in the classroom is another key area; it seems sensible to predict that practice on the computer will be vital for effective preparation. However, the introduction of the CBT did not seem to cause major changes in the use of technology in the classroom. It may be that directors of studies' worries about resourcing

will prove unnecessary and that classes will continue to run without the benefit of new technology.

All of the factors in the Henrichsen model need to be taken into account when trying to predict what the impact of the TOEFL test will be in the future. They correspond, in a major way, to what Chapman and Snyder (2000, p. 457) call *intermediate links*. Given what is known from a decade of research into impact and washback, it is no longer logical to predict that because there will be speaking and integrated tasks on the new TOEFL test there will automatically be adequate and beneficial practice of these skills in the classroom. Further analysis needs to be done regarding the message that ETS is trying to send to its users (which has not been particularly clear until recently), the means of transmitting the message, whether this will convince managers of schools to invest money in developing their schools' capacity, whether they will invest in the right things (e.g., Is it better to buy a computer or invest in some teacher training?), whether teachers will understand what they are supposed to do, whether they will be able to do this effectively, and whether this will ultimately affect student performance on the test and their more general learning.

Future Research

It is now time to focus on some of the teachers in the institutions that we have studied in order to find out how they and the institutions respond to the emerging news about the new examination. Phase 2 of this research—the transition study—will include an in-depth study of teachers' reactions to TOEFL news as it becomes available. Will teachers understand what the new TOEFL examination requires of them and their students? Will they understand what needs to go into materials and what the most effective ways might be of teaching the required skills? Will they have the knowledge and other forms of know-how to be able to react appropriately to the new demands? This study will track teachers from late 2004, as they begin to learn about the details of the new examination, until mid-2005, when, we predict, they will be quite far along in their arrangements for new TOEFL preparation course. We will elicit the teachers' reactions to the examination and record the decisions that are being taken in their institutions We hope that the information we gather will help to identify whatever "evidential links" (Messick, 1996) there may be between the introduction of the new TOEFL and the way teaching and learning look in the future.

References

Airaisian, P. W. (1988). Measurement driven instruction: A closer look. *Educational Measurement: Issues and Practices, 7*(4), 6–11.

Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.

Alderson, J. C., & Banerjee, J. (2001). Impact and washback research in language testing. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, et al. (Eds.), *Experimenting with uncertainty: Essays in honor of Alan Davies* (pp. 150–161). Cambridge, UK: Cambridge University Press.

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

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

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

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

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.

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

Butler, F., Eignor, D., Jones, S., McNamara, T., & Suomi, B. (2000). *TOEFL 2000 speaking framework: A working paper* (TOEFL Monograph No. MS-20). Princeton, NJ: ETS.

Chapman, D. W., & Snyder, C. W. (2000). Can high-stakes national testing improve instruction: Re-examining 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., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper* (TOEFL Monograph No. MS-18). Princeton, NJ: ETS.

Duffy, C., & Mahnke, K. (1996). *Heinemann TOEFL preparation course*. New York: Macmillan.

Enright, M., Grabe, G., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework: A working paper* (TOEFL Monograph No. MS-17). Princeton, NJ: ETS.

ETS. (1999). *TOEFL tips—Preparing students for the computer-based TOEFL*. Princeton, NJ: Author.

ETS. (2000). *POWERPREP. Preparation for the computer-based TOEFL.* Princeton, NJ: Author.

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

ETS. (2002b). *TOEFL 2002–2003 information bulletin.* Princeton, NJ: Author.

ETS. (2002c). *TOEFL Test score and data summary 2001–2002.* Princeton, NJ: Author.

Fekete, H., Major, E., & Nikolov, M. (1999). *English language education in Hungary: A baseline study*. Budapest: The British Council Hungary.

Frederiksen, J., & Collins, A. (1989). A system approach to educational testing. *Educational Researcher, 18*, 27–32.

Fullan, M. (1991). *The new meaning of educational change*. London: Cassell PLC.

Gear, J., & Gear, R. (2002). *Cambridge preparation for the TOEFL test*. Cambridge, UK: Cambridge University Press.

Hale, G. A. (1988). Student major field and text content: Interactive effects on reading comprehension in the Test of English as a Foreign Language. *Language Testing*, *5*(1), 49–61.

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

Hamp-Lyons, L., & Kroll, B. (1997). *TOEFL 2000—Writing: composition, community and assessment* (TOEFL Monograph No. MS-5). Princeton, NJ: ETS.

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

Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English medium university in Turkey. In A. Hughes (Ed.), *Testing English for university study* (pp. 134–153). London: Modern English Publications.

International computer-based testing center list. (n.d.). Retrieved June 2004, from the ETS Web site: http://etsis4.ets.org/tcenter/cbt_fr.cfm

Hughes, A. (1994). *Backwash and TOEFL 2000.* Unpublished manuscript.

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 framework: A working paper* (TOEFL Monograph No. MS-16). Princeton, NJ: ETS.

Kaplan. (2002). *TOEFL CBT with CD-ROM*. New York: Kaplan.

Kellaghan, T., Madaus, G., & Airaisian, P. (1982). *The effects of standardized testing.* London: Kluwer Nijhoff Publishing.

Khaniya, T. R. (1990). *Examinations as instruments for educational change: Investigating the washback effect of the Nepalese English exams.* Unpublished doctoral dissertation, University of Edinburgh.

Kirkland, M. (1971). The effect of tests on students and schools. *Review of Educational Research, 41*(4), 303–350.

Kirsch, I., Jamieson, J., Taylor, D., & Eignor, D. (1998). *Computer familiarity among TOEFL examinees* (TOEFL Research Rep. No. RR-59). Princeton, NJ: ETS.

Li, X. (1990). How powerful can a language test be? *Journal of Multilingual and Multicultural Development, 11*(5), 393–404.

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: University of Chicago Press.

Malloy, J., & Tumposky. D. (2003). *Preparing students for reading and writing in the next generation of TOEFL.* Paper presented at the TESOL Conference, Baltimore.

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

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.; pp. 12-103). New York: Macmillan.

Peirce, B. N. (1992). Demystifying the TOEFL reading test. *TESOL Quarterly, 26*(4), 665–689.

Pelgrum, W., Janssen Reinen, I., & Plomp, T. (1993). *Schools, teachers, students and computers: A cross-national perspective.* Hague: The Netherlands: International Association for the valuation of Educational Achievement.

Popham, W. J. (1987). Two-plus decades of educational objectives. *International Journal of Educational Research, 11*(1), 31–41.

Raimes, A. (1990). The TOEFL Test of Written English: Cause for concern. *TESOL Quarterly, 24*(3), 427–442.

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

Roberts, G., & Tumposky. D. (2003). *Writing tasks for the next generation TOEFL.* Paper presented at the TESOL conference, Baltimore, Maryland.

Rogers, E. (1995). *The diffusion of innovations*. New York: The Free Press.

Sharpe, P. J. (1996). *Barron's how to prepare for the TOEFL test* (7th ed.). Hauppauge, NY: Barron's.

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

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

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.

St. Petersburg Examination Project. (1996). *A baseline study.* St. Petersburg, Russia: Author.

Stoller, F. (1994). The diffusion of innovation in intensive ESL programs. *Applied Linguistics, 15*(3), 300–327.

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

Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). *The relationship between computer familiarity and performance on computer-based TOEFL test tasks* (TOEFL Research Rep. No. RR-61). Princeton, NJ: ETS.

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

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. (1999). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory.* Unpublished doctoral dissertation, Lancaster University.

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., & Alderson, J. C. (1993). Examining washback: the Sri Lankan impact study. *Language Testing, 10*(1), 41–69.

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

Weir, C., & Roberts, J. (1994). *Evaluation in ELT.* Oxford, UK: Blackwell Publishing.

Wesdorp, H. (1983). Backwash effects of multiple-choice language tests: Myth or reality? *Practice and Problems in Language Testing, 5.* Arnhem, the Netherlands: CITO.

Notes

[1] Bulgaria, Croatia, Czech Republic, Georgia, Hungary, Lithuania, Romania, Russia, and Ukraine. The Czech Republic CBT test center closed in June 2003, however.

[2] The information presented in Appendix F was current *at the time of the construction of the instruments*. ETS has since revised the new TOEFL test specifications, but this table records what was known at the time of this study. Updated information can be found on the ETS Web site at http://www.ets.org/toefl.

[3] In the excerpts for the transcripts used in this report, references to participants will follow this pattern: TL2: 293 means Teacher L2 (at Site L2—Lithuania), Line 293 of transcript.

[4] See ETS (1999, 2000, 2002a, 2002b, 2002c), Malloy & Tumposky (2003), and Roberts & Tumposky (2003). See the ETS Web site for details about TAST, TWE, LanguEdge: http://www.ets.org

List of Appendixes

Appendix A

TOEFL Candidates by Native

Country July 2001–June 2002

Country	CBT	PBT
Albania	697	191
Belarus	529	14
Bosnia & Herzogovina	406	61
Bulgaria	3,251	46
Croatia	471	3
Czech Republic	766	9
Estonia	172	122
Georgia	460	226
Hungary	928	7
Latvia	323	159
Lithuania	656	5
Macedonia	353	114
(Formerly Yugoslav Republic of)		
Moldova		43
Poland	2,748	70
Romania	2,893	76
Russia	5,059	188
Slovakia	490	70
Slovenia	96	11
Ukraine	2,348	48
Yugoslavia	732	144

Note. CBT = computer-based test. PBT = paper-based test. Data are from *TOEFL Test Score and Data Summary 2001–2002,* by ETS, 2003, Princeton, NJ: Author. Copyright 2003 by ETS. Adapted with permission.

TOEFL Test Centers in Central and Eastern Europe

Table B1

Test Types Offered by Test Centers in Central and Eastern Europe

Country	CBT	PBT
Albania	0	1
Belarus	0	1
Bosnia & Herzogovina	0	1
Bulgaria	1	1
Croatia	1	0
Czech Republic	0	1
Estonia	0	2
Georgia	1	0
Hungary	1	0
Kosovo	0	1
Latvia	0	1
Lithuania	1	0
Macedonia	0	1
Moldova	0	1
Poland	0	6
Romania	1	3
Russia	2	12
Slovakia	0	2
Slovenia	0	1
Ukraine	1	3
Yugoslavia	0	1
Total number of centers	9	39

Note. CBT = computer-based test. PBT = paper-based test. Only one version of TOEFL (CBT or PBT) was offered in any test center. Data are from *Test Center List—Paper-Based TOEFL Test Administrations,* by ETS, 2003, Princeton, NJ: Author. Copyright 2003 by ETS. Adapted with permission.

Table B2

Total of Test Types

Test type offered	CBT only	PBT only	Both	Total
Number of countries	4	13	4	21

Note. Data are from *International Computer-Based Testing Center List,* n.d., retrieved June 2004, from the ETS Web site: http://etsis4.ets.org/tcenter/cbt_fr.cfm. Copyright by ETS. Adapted with permission.

Appendix C

Correspondence With Language Schools in the CEE Region

Regarding TOEFL Preparation Courses

Croatia

"Thank you for your message. I'm afraid there are not many schools in Croatia involved in TOEFL exams, the ALTE Cambridge exams being more popular." (Zagreb, March 21, 2003)

Czech Republic

"Thank you for you enquiry, but I am afraid that we do not presently offer enough TOEFL preparation courses (only one a year) to be of assistance." (Prague, March 24, 2003)

"We would be pleased to participate in this project; however, first I have to inform you about the situation with our TOEFL courses. Recently we have been teaching these courses only on an occasional, one-to-one basis. The amount of teachers with experience in teaching TOEFL preparation courses also corresponds with this fact. Currently, our clients are more interested in Cambridge exams and we offer these courses both for groups and individuals." (Prague, May 29, 2003)

Hungary

"In my experience though, schools in Hungary offer TOEFL prep courses on an irregular basis based on demand or interest." (Budapest, May 29, 2003)

Latvia

"Thanks for informing me about possible cooperation with Lancaster University. Unfortunately, we don't prepare for TOEFL and, what's more, I don't know anyone in Latvia who does. Maybe private tutors?" (March 13, 2003)

"As far as I know organized TOEFL preparation classes are only at the Language Learning Center in Riga. Their branch in Liepaja offers materials on TOEFL test but doesn't have computer-based training. At Liepaja university (where the pilot study was carried out) there are no organized classes Generally, students who need to prepare hire private teachers or order preparation CDs from TOEFL." (Riga, June 20, 2003)

ROMANIA

"As TOEFL is valid only for 2 years, the Romanian students are more interested in obtaining a lifetime certificate, such as the Cambridge certificates." (October 30, 2003)

Appendix D

Number of Candidates by Test Center Location:

Computer-Based Tests, 2003

Country	City	Test center	No. of candidates
Bulgaria	Sofia	8007	2,553
Croatia	Zagreb	8009	417
Czech Republic	Prague	8014	522
Hungary	Budapest	8042	1,004
Lithuania	Vilnius	8058	457
Romania	Bucharest	8076	1,437
Russia	All centers		2,144
	Moscow 1	8951	148
	Moscow 2	8961	1,680
	St. Petersburg	8953	316

Note. No test data received from Estonia, Latvia, Macedonia, Poland, Slovakia, Slovenia, or Ukraine. Data are from a personal communication with M. Enright, May 28, 2004.

Appendix E

Number of Candidates by Test Center Location:

Paper-Based Tests, 2003

Country	City	Test center	No. of candidates
Bulgaria	Varna	M316	550
Czech Republic	Prague	M515	51
Estonia	All centers		188
	Tallinn	P476	150
	Tartu University	P479	38
Latvia	Riga	P485	194
Poland	All centers		998
	Szestoehowa [sic] 1	P200	38
	Szestoehowa [sic] 2	P201	3
	Poznan	P204	124
	Warsaw 1	P205	403
	Warsaw 2	P209	418
	Gdansk	P213	12
Romania	All centers		141
	Bucharest	P403	1
	Cluj	P405	94
	Iasi	P415	46
Russia	All centers		158
	Krasnojarsk	P552	4
	Yekaterinburg	P553	35
	Krasnodar	P554	11
	Rostov-Na-Donu	P555	3
	Samara	P558	4
	Moscow	P560	1
	Khabarovsk	P561	1
	Tomsk	P566	15
	Vladivostok	P567	75
	Volgograd	P568	9
Slovakia	All centers		75
	Bratislava	M502	53
	Kosice	M505	25

Note. No test data received from Hungary, Lithuania, or Croatia. Data are from a personal communication with M. Enright, May 28, 2004.

Appendix F

Comparisons of Different Versions of TOEFL (June 2004)

At the time this report was written, no specific test details on the new TOEFL had been released by ETS. All information in this appendix was based on a variety of sources, including LanguEdge (ETS, 2002a) and the ETS speaking test TAST, which were promoted by ETS as practice materials; various conference presentations by ETS staff; and personal communications with ETS staff.[4]

New Features

New features in the new TOEFL include the following:

- Four modes (skills) are tested—including speaking.

- There is no separate structure paper—instead, structure is evaluated via the four skills sections.

- Integrated tasks are included as well as independent tasks.

- Test takers are required to synthesize information from various texts required in reading test.

- Tasks in both reading and listening sections are more complex than previously.

- Writing tasks are similar to the past TOEFL essay, but the focus is on giving and supporting opinions (agree/disagree).

- Diagnostic information for test takers is provided.

These new features are shown in Tables F1 through F5, which compare various sections in TOEFL across three versions: the paper-based and computer based versions of TOEFL and what was known at the time regarding the new TOEFL. Figure F1 shows the components of the new integrated tasks.

Table F1

Three Versions of TOEFL

	Date of introduction	Availability	How often held?	Timing	Scoring	Notes
PBT	1963–1964	Offered where CBT not offered	5–6 times a year (plus some extra dates in certain centers)	Less than 3 hours according to Web site or 3 ½ according to Information Bulletin	Until July 1998 possible range: 200 to 677 Now possible range: 310 to 677 (to avoid overlap with CBT scores)	
CBT	1998	Where there are ETS CBT centers (see ETS Web site)	On appointment basis	Max—3 hrs 20 Min—2 hrs 35 + untimed tutorial and break	Possible range: 0–300 Concordance tables available to determine comparability	TWE introduced in 1986. Revised 1989. Offered with TOEFL from 1990.
New TOEFL	2004 on practice basis 2005 on stream					

Note. CBT = computer-based test. PBT = paper-based test.

Table F2

Comparison of Listening Section

	Paper-based test	Computer-based test	New TOEFL	
			Independent task	Integrated task: Listening/Reading/Writing
Construct	Ability to understand English as spoken in North America use vocabulary and idiomatic expression as well as grammatical constructions frequently used in spoken English	• Comprehension of main ideas • Order of a process • Supporting ideas • Important details • Ability to make inferences and categorize topics and objects	• Basic comprehension • Gist questions • Detail questions • Pragmatic understanding • Identify function • Identify stance • Connecting ideas • Organization question • Content questions • Link content	See Table F4
Item types	MC (4 choices)	MC (4 choices) • Select a visual (or part of visual) • Select more than one answer (e.g., 2 out of 4) • Order events or steps in process • Match objects or text to categories	MC comprehension	Not applicable— input only
Text (stimuli) types	3 sections • short conversations • longer conversations • talks	A. Short dialogues (maximum: 2 turns) B. Short conversations (More than 2 turns) OR mini-lectures/academic discussions	Conversations/lectures (not necessarily only traditional style)	Lecture

137

(Table continues)

Table F2 (continued)

	Paper-based test	Computer-based test	New TOEFL	
			Independent task	Integrated task: Listening/Reading/Writing
Timing		40–60 minutes (15–25 for answering) Self-paced	25 minutes to answer questions	15 minutes (to write)
Text length (stimulus)		Up to 1 minute Up to 2.5 minutes/2 minutes	(probably longer lectures than for computer-based test)	200– 250 words/2 minutes
Number of questions/tasks	**[a]	30–50	6 passages/35 questions approximately	
Scoring		0–30 scale		
Notes		Computer adaptive[b] No note taking allowed	Computer adaptive[c] Note taking allowed Hear once only	Personal opinion not required Note taking allowed Hear once only

Note. MC = multiple choice question. Blank cell indicates no data was available.

[a] There were a minimum of 140 items in total, made up of listening, reading, and structure items. The exact number for each skill varied from administration to administration, but the total number of items was always ± 140. [b] Based on Spoken and Written Academic Language corpus (produced by ETS). [c] We couldn't get a clear picture about this from consulting all the sources..

Table F3

Comparison of Reading Section

	Paper-based test	Computer-based test	New TOEFL		
			Independent task	Integrated task Listening/Reading/ Speaking	Integrated task Listening/Reading/ Writing
Construct	• Ability to read nontechnical reading matter and understand short passages that are similar in topic and style to those that students are likely to encounter in North American colleges or universities • Understand information stated or implied	• Comprehension of main and supporting ideas • Ability to analyze meaning • Cognitively complex comprehension; ability to understand organization of passage—including cohesion and coherence • Ability to comprehend discrete language segments and use context clues	• Vocabulary comprehension • Reference • Identifying essential details from complex sentences • Understanding lexical, grammatical & logical links • Factual info stated in text • Verify true info • Inference • Rhetorical purpose	See Tables F2 and F4	
Item types	MC	• MC: Select option or visuals • Click on word/phrase/ sentence/paragraph that corresponds to a vocabulary item or structural feature • Insert a sentence	• Identify synonyms (MC) • Identify reference (MC) • Sentence simplification (MC) • Find false option (MC) • Insert text • Prose summary • Schematic table	Not applicable—input only	

(Table continues)

139

Table F3 (continued)

| | Paper-based test | Computer-based test | New TOEFL | | |
			Independent task	Integrated task Listening/Reading/ Speaking	Integrated task Listening/Reading/ Writing
Text types	Academic	Academic	Academic: Exposition, argumentation, historical, biographical/ autobiographical		
Timing		70–90 minutes	25 minutes per passage = 75 minutes		20 minutes to write
Text length (stimulus)		4–5 passages of 250–350 words	3 passages		200–250 words
Number of questions/ tasks	**[a]	44–55 (11 per passage)	36 approximately (13 approximately per passage)	1	1
Scoring		0–30 scale			
Notes				Text on screen, disappears, then prompt given Note taking allowed	Note taking allowed

Note. MC = multiple choice question. Blank cell indicates no data were available.

[a] There were a minimum of 140 items in total, made up of listening, reading, and structure items. The exact number for each skill varied from administration to administration, but the total number of items was always ± 140.

Table F4

Comparison of Writing Section

| | PBT | CBT | New TOEFL | |
			Independent task	Integrated task: Listening/Reading/Writing
Construct	TWE taken with PBT (no separate fee)—Same construct as for CBT writing test-> Also Structure and Written Expression (joint paper)—measures ability to recognize language that is appropriate for standard written English	• Ability to generate, organize, develop ideas • Support those ideas with examples or evidence • Compose a response in standard written English/ including use of various syntactic forms and selection of appropriate formal vocabulary/(review)	• Produce extended text • Give and support opinions • Organize ideas • Use a range of grammatical features and vocabulary	• Comprehend an academic text • Distinguish essential from nonessential info • Organize info • Produce grammatical sentence without directly copying the source • Ability to understand info in lecture • Convey that info accurately • Demonstrate high level of lexical & grammatical accuracy
Item types	Short essay—Assigned randomly from pool by computer. Topics listed in the Information Bulletin/ETS Web site		Short essay	Short essay
Text types		Defending a position	Opinion/preference	Explanation of how lecture and text are related—No opinion/preference
Timing	30 minutes	30 minutes	30 minutes	Unclear [b]
Text length (output)			300 words [a]	
Number of questions/tasks	1	1 topic	1	1

(Table continues)

141

Table F4 (continued)

	PBT	CBT	New TOEFL	
			Independent task	Integrated task: Listening/Reading/Writing
Scoring	0–6 scale. Rated by 2 independent raters—average awarded in 0.5 raw score increments		1–5 scale	1–5 scale (different descriptors to independent task scale)
	TWE score printed separately	Composite score given (see Structure)		
Notes	Online registration possible	Can hand-write or type (Notepad based—no tab, spell or grammar check). Get scores quicker if typed.	Typed only (no choice)	Passage not visible while listening—visible while writing response. Student may refer to notes.

Note. Blank cell indicates no data was available.

[a] Word counts are given as guidelines for an adequate answer not as strict limit. [b] We couldn't get a clear picture from the information available, but there were actually different pictures being projected—a problem, according to innovation theory.

Table F5

Comparison of Structure and Written Expression

	PBT	CBT	New TOEFL
Construct	Knowledge of important structural and grammatical elements of standard written English	Ability to recognize appropriate standard written English and assess a variety of structural features including plural forms, tenses, and word order.	Not applicable
Item types	• MC (4 choices) • Selecting incorrect phrase in sentence (4 choices)	As per PBT	
Timing		15–20 minutes	
Number of questions/tasks	** [a]	20– 25 tasks	
Scoring		Composite score of Structure section (0–30) + Writing section (0–6) on 50%/50% basis	
Notes		Computer adaptive	

Note. MC = multiple choice question. Blank cell indicates no data was available.

[a] There were a minimum of 140 items in total, made up of listening, reading, and structure items. The exact number for each skill varied from administration to administration, but the total number of items was always ± 140.

Table F6

Comparison of Speaking

	PBT	CBT	New TOEFL		
			Independent task	Integrated task: Listening/Reading/Speaking	Integrated task: Listening/Speaking
Construct	N/A	N/A	Ability to describe a choice and explain and support personal preference	Ability to accurately combine and convey key information	
Item types			Task 1 　Talk about familiar, 　low-risk topics Task 2 　State & support 　preferences	Task 1 　Fit & explain Task 2 　General-> specific	Task 1 　Problem-> solution Task 2 　Summary
Text types			Domain: personal Independent: General topics	Domain: academic Based on a reading (600–700 words) + a connected listening passage	Domain: academic Based on academic lecture
Timing			15 seconds of preparation + 45 seconds to record	15 minutes to read + 90 seconds of preparation + 90 seconds to record	Listening time + 60 seconds preparation + 90 to record
Length (output)			60 seconds of speaking	90 seconds of speaking	90 seconds of speaking
number of questions/tasks			2	2	2
Scoring			5 band holistic scale	5 band holistic scale	5 band holistic scale
Notes			TAST due to become speaking component of New TOEFL in 2005		

Summary of Tasks

Figure F1 shows the independent tasks for each skill area in the three versions of TOEFL (paper-based and computer-based tests and the new TOEFL). See Tables F1-F6 for the number and types of tasks.
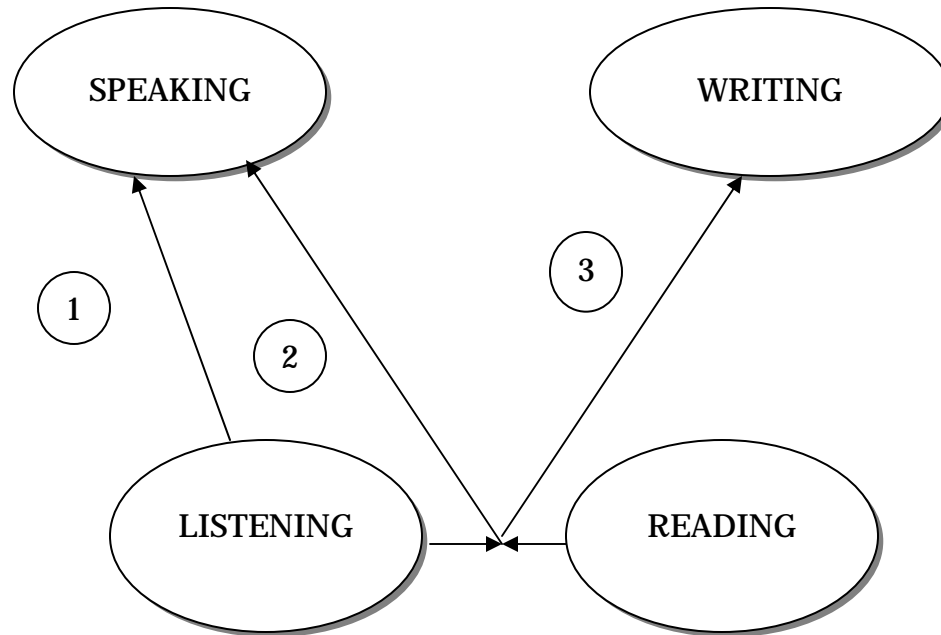
Figure F1. **Summary of tasks.**

Note. This figure has three integrated tasks combinations (and sequences): 1 = Listening (input) and Speaking (output); 2 = Listening + Reading (input) and Speaking (output); and 3 =Listening + Reading (input) and Writing (output).

Appendix G

Structured Interview With TOEFL Teachers

For admin—not part of the interview:

Name of interviewer: _____

Name of teacher: _____ male ☐ female ☐

Name of institution: _____

City: _____

Date of interview: _____

Time of interview: _____

Location of interview: _____

(Age):

_____

Notes to interviewer:

- (Anything in parentheses are notes to the interviewer—not to be read out loud.)

- If so & if not questions: Only ask if relevant.

- Please mark on this sheet any questions NOT asked.

- Remember first of all to refer to the observed lesson and thank the teacher.

_____

Post-observation link:

1. Was that a typical lesson?

2. Do you feel you reached your objectives for that lesson?

3. *If not*, why not?

4. Do you think the TOEFL influenced your methodology in that lesson?

5. To what extent did TOEFL influence the content?

Teaching experience

6. How many years of experience as a teacher do you have? (– subject(s)?)

7. How many years of experience as a teacher <u>of English</u> do you have?

8. How long have you been teaching TOEFL preparation classes?

9. How much of this time has been teaching TOEFL <u>CBT</u> preparation classes?

10. What teacher-training courses have you attended?

11. Were they before you became a teacher or after? *(PRESET or INSET)*

12. Do you work with any exam board or exam bodies?

13. In what role?

14. Have you taken the TOEFL exam yourself?

15. *If so,* do you think this experience influences how you teach test preparation classes? (*draw on questions 7-11*)

The TOEFL preparation classes—aims

16. (SHOW prompt #5) Do you see the aim of the class primarily is to
    i. work on the right things for passing TOEFL
    ii. to improve the students' general English
    iii. to prepare students for working in an academic environment?

17. How do you feel about that?

The TOEFL preparation classes—selection

18. Are students **screened** (preselected) in any way before they can join this TOEFL class?

19. *If so*, how?

20. Why do you think students choose to come to this particular school?

The TOEFL preparation classes—methodology

21. Have you ever taught or do you teach other EAP (English for Academic Purposes) classes?

22. *If so,* how is this TOEFL class different or similar to them?

23. Which language (L1 or L2) do you use most in class?

24. On what do you base your decision about which language to use?

25. Which language is used most in class by the students?

26. How do you feel about that?

27. Which of these different working arrangements do you use most in class? (SEE PROMPT #2: pair work, group work, individual work, whole class, etc.)

28. What do you tend to use most in a typical class?

The TOEFL preparation classes—course content

29. Is there a course outline for the TOEFL courses from this institution? *(ask for a copy)*

30. Who produced it? What's it based on?

31. Over the length of the course which **language elements** do you spend most time on and which least? (SHOW PROMPT #3: listening/reading/writing/speaking/grammar/ vocabulary/other)

32. What is your rationale?

33. So, are all **four skills** practiced in class? (*To double check—especially re. speaking*)

34. What types of activities do you typically use to practice each skill?

35. What activities *(gloss to "exercise," if necessary)* do you do to develop **listening**?

36. What kinds of texts do you usually give students to **read** in the lessons?

37. How long are they on average?

38. What topics do they generally cover?

39. Do you do any activities working on **two 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 what you heard?

40. Is **speaking** improved during class -time at all?

41. *If so:* What kinds of activities do you do?

42. What do you do in the course to improve **writing**?

The TOEFL preparation classes—assessment

43. Do you set **essays** to write for TOEFL preparation?

44. *If so*: Do you give the students marks for the essays?

45. What system of marking (grading) do you use?

46. *If not mentioned*: Are you familiar with using **scoring bands** to mark written work?

47. *If not mentioned*: Are you aware of existing marking criteria for the written section of TOEFL? *(see example provided—ask for copies if an alternative scheme is used)*

48. Do you use them for marking classwork or homework—often/at all?

49. Do you feel comfortable using them? Why/Why not?

50. Do you refer to them in class? (i.e., Are the students familiar with it?)

51. Do you give tests to check other skills (reading, listening)?

The TOEFL preparation classes–test taking techniques

52. Do you cover test-taking techniques in your lessons? (e.g., analyzing questions—*refer to observation)*

53. *If so*: Are these activities very different from other activities you do in the lessons? *(paraphrase if necessary)* (*This aims at identifying specific test-focused vs. more general lang. development activities)*

54. What proportion of the whole course is spent specifically on exam practice?

55. At which stage of the course do you do most exam practice? (beginning/middle/end/throughout)

The TOEFL preparation classes—teaching materials and resources

56. Which of these **materials** are available at your institution for teaching TOEFL prep? (SHOW PROMPT #4—uncover <u>one by one</u>)

57. Which of these materials do you use <u>most</u> in class? Why?

58. Are you familiar with these materials? *(show example of ETS materials e.g., POWERPREP, LanguEdge)*

59. Do you use them?

60. *(If relevant* (see (f) on prompt): What resources do you draw on to help you?

61. Do you use or make reference to the ETS Web sites in class?

62. Do you spend time in class on **computer-based** tasks? (*To double check. Refer to PROMPT again)*

63. *If not*, why not?

64. *If so,* How much of the course as a whole is typically spent on computer-based tasks/practice?

65. *If relevant:* How many computers are available?

66. Do students do computer practice outside of classes?

The TOEFL preparation classes—atmosphere

67. How would you describe the general atmosphere in your lessons?

68. Was today's lesson typical concerning that?

Students' independent language development strategies

69. Do students do any study outside class to help prepare for TOEFL, as far as you know?

70. What kind?

71. Is this prompted by you (e.g., by giving tips or ideas for what to do)?

Test-taking strategies NB: Ts don't have to answer this if they don't want to:

72. Have you heard of any methods for cheating which candidates try?

73. Is it common?

74. Why do you think some candidates try to cheat?

Teacher Support

75. Do you refer to official TOEFL materials (e.g., booklets such as TOEFL Tips or the Web site) to give you guidance on how to teach these classes?

76. Have you ever had any training to teach EAP preparation?

77. *If so*, from where?

78. Have you ever had any training to teach TOEFL preparation?

79. *If so*, from where?

80. Is any training on how to teach computer-based courses available?

81. *If so*, from where?

82. Do teachers tend to take up training opportunities offered (*re former 2 Qs- not in general*)?

83. *If not*, why not?

84. *(If relevant)*: How far has this training influenced the content of your lessons?

85. How far has this training influenced the methodology of your lessons?

Computer skills

86. How confident do you feel in your computer skills?

87. Do you feel your confidence in using computers has an effect on your ability to teach TOEFL preparation?

TOEFL Awareness

88. What are your **sources of information** about TOEFL? (if Web sites or books are mentioned, get details of exactly which one(s) if possible)

89. Which do you find most helpful? Why?

90. *(If not mentioned already):* Are you familiar with the TOEFL **Web site**?

91. How would you describe it (*If necessary add examples:*—e.g., useful/interesting/easy to navigate/hard…?

92. As far as you know, what preparation materials are available from the producers of the TOEFL exam (ETS)?

93. As far as you know, what are the main **differences** between the paper & pencil version of the exam and computer-based version?

94. What can you tell me about the format of CBT? (e.g., questions types, topics, exam procedure)

95. In your opinion does it test the following things? (*Yes/No answers required, plus brief expansion if they wish*)

- ability to use grammar correctly
- ability to use a wide range of vocabulary appropriately
- ability to use idioms correctly
- ability to understand a wide range of texts
- ability to express original ideas in writing
- ability to translate from your native language to English and vice-versa
- ability to take an active part in an academic discussion or seminar
- ability to understand lectures
- ability to infer someone's opinion, when it is not stated clearly
- ability to understand the organization of a text
- ability to write formal letters
- ability to analyze information from several texts
- ability to make inferences from information in a text
- ability to give a presentation
- ability to understand unfamiliar vocabulary from context clues
- ability to state your opinion on a given topic and support it
- ability to understand language used in everyday situations and conversations
- ability to speak for an extended period on a familiar topic
- ability to write an academic style article

Attitudes to TOEFL (NB CBT not NT)

96.   From your experience, do you believe TOEFL scores reflect students' real language ability?

97.   What language skills and subskills does a candidate need to do well on TOEFL, in your opinion?

98.   What knowledge or skills <u>apart from language</u> does a candidate need to get good TOEFL scores, in your opinion?

99.   What do you think are the good features of the TOEFL exam (if any)?

100.   What do you think are the bad features of the TOEFL exam (if any)?

101.   Which section/aspect is hardest to teach?

102.   Which section/aspect is easiest to teach?

103.   Do you feel there are any topics which are used very frequently in TOEFL?

104.   Do you feel TOEFL is expensive?

105.   Would students be prepared to pay more to take it?

Attitudes to Teaching TOEFL

106.   Are TOEFL classes popular with teachers? Why/Why not?

107.   Do you <u>personally</u> like teaching TOEFL? Why/Why not?

108.   Are you yourself learning anything by teaching these classes?

NT awareness

109.   Have you heard about the **new version of TOEFL**, which will be introduced next year? *(if not, go to q. 30)*

110.   *If so:* What do you know about it?

111.   *If so: H*ow did you find out about it?

112.   *If so:* As far as you are aware, how does it differ from the current CBT TOEFL?

113.   What's your reaction to the <u>new</u> TOEFL?

Attitudes to Tests (D)

114.   Do you agree or disagree with the following statements? And why?

- 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

(If enough time)

115.   What has most influenced your teaching over the years?

116.   What has most influenced your teaching of TOEFL?

If there was no possibility to observe this teacher taking a second EAP/Advanced class ask the following:

- Do you also teach EAP classes?

- Are they the same or different from your TOEFL preparation classes?

- In what ways?

- How are the materials/class activities/your role in the classroom/skills covered different?

Remember to:

- Get permission form signed.

- <u>Record</u> initial impressions.

- Note postinterview notes.

Appendix H

Structured Interview With TOEFL Students

a) Name of interviewer: _____

b) Name of institution: _____

c) City: _____

d) Date of interview: _____

e) Time of interview: _____

f) Location of interview: _____

g) Type of course followed: TOEFL only   ☐

                               General English plus TOEFL ☐

                               Other:_____

                                                            previous

h) Names of students: From   TOEFL?

_____ male ☐ female ☐ / _____ Y/N

_____ male ☐ female ☐ / _____ Y/N

_____ male ☐ female ☐ / _____ Y/N

_____

Notes to interviewer:

- (Anything in parentheses are notes to the interviewer—not to be read out loud.)
- If so & if not questions: Only ask if relevant.
- *Please mark on this sheet any questions NOT asked.*
- *Check students have labels/name cards and use their names frequently for sake of clearer recording.*

_____

To complete section h above/Warm up:

1. What are your names?
2. What should I call you?
3. Where are you from?

4. Have you taken TOEFL before? *(check they understand this means the real TOEFL—not just practice/mock tests, etc.)*

Background

5. Why are you taking TOEFL?
6. *If it's for future studies*: What do you plan to study?
7. Where?
8. What is the date of your TOEFL exam?
9. Which version will you take, computer based or paper & pencil? *(explain the difference briefly, if necessary)*
10. Did you have a choice?
11. If so, why did you make that choice?

Aims

12. Do you see the main aim of the course as above all to (SHOW PROMPT # 5)
    a) work on the right things for passing TOEFL (e.g., to get tips and learn TOEFL tricks)
    b) improve your general English
    c) prepare you for working in an academic environment?
13. Is that what you wanted when you registered for this course?

Your TOEFL preparation lessons—course content

14. What areas of **language** does your teacher focus on most in the lessons (SEE PROMPT #3)?
15. Are all **four skills** practiced in lessons? *(to double check)*
16. Do you do any activities *(gloss to "exercise," if necessary)* working on **two 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 what you heard?
17. What activities do you do to develop your **listening?**
18. What activities do you do to develop your **reading?**
19. What kinds of texts do you usually **read** in the lessons?
20. How long are they on average?
21. What topics do they generally cover?

22. Do you do activities to improve your **speaking**?

23. *If so:* What kinds of activities do you do?

24. What activities do you do to develop your **writing?**

The TOEFL preparation lessons—materials and resources

25. What materials does your teacher use in class (e.g., books, worksheets)? *(Get book titles. Don't use prompt #4—for Teachers only)*

26. What other materials are available for you at this school for independent TOEFL preparation? *(Get titles. Don't use prompt #4—for Teachers only)*

27. Are these materials useful? (refers to both of above two questions)

28. Are there any other materials you would like to have access to here?

29. *If not mentioned explicitly*: Do you use any ETS (i.e., official TOEFL) materials in class (e.g., TOEFL Tips, POWERPREP)? *(to double-check; show photocopies of covers)*

30. Do you usually/ever spend time during lessons on computer-based exercises? *(question depends on what's seen in observation)*

31. How much of the course, so far, has been spent on computer-based exercises/practice?

32. What's your opinion on that?

The TOEFL preparation lessons—test-taking techniques

33. Do you do test-taking techniques in your lessons?

34. What sort?

35. How much of the whole course so far has been spent on test-taking techniques?

Your TOEFL preparation lessons—methodology

36. Which language **(L1 or L2)** does your teacher use most in class?

37. What's your opinion on that?

38. Which language do you, the students, use most in class?

39. What's your opinion on that?

40. Which of these different **working arrangements** is used most in class? (SHOW candidates PROMPT #2: pair work, group work, individual work, whole class, etc.)

41. What else is used? (*from that prompt list*)

Assessment

42. Does your teacher set you essays to write for TOEFL preparation?

43. *If so*: Do you get marks from your teacher for your essays?

44. What system of marking (grading) does your teacher use?

45. Is it useful/helpful?

46. Does your teacher give you tests to check your other skills (e.g., reading or listening)?

Atmosphere

47. How would you describe the general atmosphere in your lessons—fun or serious?

48. Was today's lesson typical concerning that?

Opinion

49. First of all, was today's lesson typical?

50. What do you like about your lessons?

51. Is there anything you dislike about your lessons?

52. Is your language improving?

53. If so: How?

54. Is it improving in all four skills (*gloss if necessary*)?

55. Why did you decide to take a preparation course rather than simply preparing on your own?

Independent language development strategies

56. Do you do any study outside class to help prepare for TOEFL?

57. *If so*: What do you do?

58. How much time do you spend out of class on TOEFL preparation?

59. Do you do any computer-based practice outside class hours?

60. *If so*: How much?

61. (If not mentioned in Material & Resources section above) What materials do you use? Which packages? (Try to get specific titles)

62. Is this prompted by your teachers (e.g., by giving tips or ideas for what to do)?

63. What other strategies do you use to learn English? *(paraphrase if necessary)*

64. Are these strategies helpful just to succeed at TOEFL or to improve your English in general, in your opinion?

Computer skills

65. Do you feel confident in your computer skills?

66. *If not:* How often do you usually use a computer?

67. Has this increased due to computer -based TOEFL practice?

68. Does your confidence in using computers affect your ability to do well on the TOEFL exam?

69. How do you feel about taking a computerized test?

Test taking strategies NB: They don't have to answer this if they don't want to:

70. Have you heard of any methods for cheating which candidates try? (If not mentioned, probe the difference between types of cheating: a) substitution of the real candidate and b) preparation, i.e., getting information about what questions might come up. Beware of feeding ideas!)

71. *If relevant:* Is it common?

72. Why do you think some candidates might try to cheat?

TOEFL awareness

73. What is your main **source of information** about TOEFL? (e.g., teacher, Web site, books, people who've taken it already…) *If Web sites are mentioned, clarify which ones—ETS, institutional, language schools, and so on.*

74. *If not mentioned already:* Are you familiar with the TOEFL **Web site**?

75. *If relevant:* How would you describe it (e.g., useful/interesting/easy to navigate/hard…)?

76. As far as you know, what official preparation materials are available from TOEFL?

Attitudes to TOEFL—make clear we're discussing current CBT TOEFL

77. In your opinion, is TOEFL a good test?

78. Why?/Why not?

79. Do you believe TOEFL scores reflect a student's real language ability?

80. Can you explain your views on this?

81. What knowledge or skills, apart from language, does a candidate need to get good TOEFL scores, in your opinion?

82. In your view, what are good methods for getting a good result on TOEFL—either on your own or on class? (Show PROMPT #1)

83. Which section or aspect of the TOEFL exam is hardest?

84. Why?

85. Which section or aspect of the TOEFL exam is easiest?

86. Why?

87. Does TOEFL motivate you to learn more English?

88. Has TOEFL been stressful in any way?

89. Do you feel TOEFL is expensive?

90. Would you be prepared to pay more to take it?

Attitudes to tests

91. Do you agree or disagree with the following statements? And why?

- 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 you study how to take tests not how to develop your language skills

92. As far as you know, what does TOEFL test? (if queried—anything is acceptable here to see what students think—do not lead)

93. In your opinion does it test the following things? (Yes/No answers required, plus brief expansion if they wish)

       a) Your ability to use grammar correctly

       b) Your ability to use a wide range of vocabulary appropriately

       c) Your ability to use idioms correctly *(paraphrase if necessary)*

       d) Your ability to understand a wide range of texts

       e) Your ability to express original ideas in writing

       f) Your ability to translate from your native language to English and vice-versa

       g) Your ability to take an active part in an academic discussion or seminar

       h) Your ability to understand lectures

       i) Your ability to understand someone's opinion, when it is not stated clearly

       j) Your ability to understand the organization of a text

k) Your ability to write formal letters

l) Your ability to analyze information from several texts

m) Your ability to give a presentation

n) Your ability to understand unfamiliar vocabulary from context clues

o) Your ability to state your opinion on a given topic and support it

p) Your ability to understand language used in everyday situations and conversations

q) Your ability to speak for an extended period on a familiar topic

r) Your ability to write an academic style article

NT awareness

94. Have you heard about the **new version** of TOEFL, which will be introduced next year?

95. *If so*: What do you know about it?

96. How did you find out about it?

97. As far as you are aware, how does it differ from the current computer-based TOEFL?

98. What's your reaction to the <u>new</u> TOEFL?

Comparison

99. Have you ever attended an **English for Academic Purposes** lesson? *(paraphrase if necessary)*

100.   *If so*: How was it similar or different from your TOEFL lessons?

Other

101.   If you could make any changes to TOEFL what would they be and why?

_____

Remember

- to get permission forms signed

- to check recording

- to record first impressions ASAP

- to write up postinterview notes

Prompts To Be Used With Teacher Interview and Students Interview

(as indicated in interview schedules)

1. **In your view, what are good methods for getting a good result on TOEFL—either on your own or on class?**

    a) Memorizing lists of vocabulary

    b) Listening to discussions on the radio

    c) Studying grammar rules

    d) Memorizing lists of idioms

    e) Reading English literature (novels, stories, poetry, etc.)

    f) Watching documentaries on TV

    g) Taking TOEFL practice tests

    h) Practicing expressing your ideas and opinions when you speak

    i) Practicing expressing your ideas and opinions in writing

    j) Reading newspapers

    k) Reading academic journals

    l) Reading general interest magazines

    m) Working carefully through a TOEFL preparation text book

    n) Memorizing model essays

2. Types of interaction

- students working in pairs or small groups
- whole class involved at once
- students working independently
- students using computers
- something else

3. **Skills covered**

- Reading
- Writing
- Listening
- Speaking
- Grammar
- Vocabulary
- Something else

4. What materials are available at your institution for teaching TOEFL classes?

    a) Materials such as books, CDs and cassettes produced by ETS, who produce TOEFL (e.g., POWERPREP)

    b) Practice materials downloaded from the TOEFL Web site

    c) Other (non-ETS) commercial publications locally published

    d) Other (non-ETS) commercial publications published abroad

    e) Unpublished materials produced by your institution

    f) Materials produced by yourself

    g) Past exam papers

    h) Something else

5.

    a) work on the right things for passing TOEFL (e.g., to get tips and learn TOEFL tricks)

    b) improve your general English

    c) prepare you for working in an academic environment

Appendix I

Structured Interview With Directors of Studies

Name of interviewer: _____

Name of director: _____ male ☐ female ☐

Name of institution: _____

City: _____

Date of interview: _____

Time of interview: _____

Location of interview: _____

(Age—approx.)

_____

Notes to interviewer:

- (Anything in parentheses are notes to the interviewer—not to be read out loud.)
- If so & if not questions: only ask if relevant.
- *Please mark on this sheet any questions NOT asked.*

_____

Courses offered

1. How many TOEFL courses do you run in a year on average?
2. Do you offer students courses in preparing for the CBT version, the P&P version of TOEFL or a combined TOEFL course?
3. Do you offer various TOEFL courses, with different formats?—(expand if necessary, e.g., general English plus intensive TOEFL, super intensive TOEFL)
4. How long are the TOEFL courses?
5. How many days per week do the students have a lesson?
6. How long is each lesson?
7. What times of year do you run the courses?
8. Why at those times?
9. What times of year do you get most students enrolling for TOEFL courses?

10. How many candidates from your institution take TOEFL (yearly)?

11. How long is the current course? (number of days/weeks)

12. What stage of the current course are you at now? (beginning/mid/end?)

13. What is the date of the next TOEFL exam?

The TOEFL preparation classes—selection

14. Are students **screened** (preselected) in any way before they can join this TOEFL class?

15. If so: How?

16. Why do you think students choose to come to this particular school?

Class size

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

18. What decides this?

19. In your opinion, does class size affect the teaching & learning?

Selection

20. What criteria do you use to select teachers for TOEFL classes?

21. What criteria do you use to select teachers for general EAP classes?

22. Are computer skills seen as important for either course?

23. Do you think teaching TOEFL has any effect on the teaching of other classes?

Training

24. Is any training offered by your institution on teaching EAP classes?

25. Is any training offered by your institution on teaching TOEFL classes?

26. Is any training in how to teach computer-based classes offered to teachers?

27. *If so: B*y whom?

28. Do teachers take up the training offered?

Resources

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

30. Are they heavily used?

31. What resources do you have which are specific to TOEFL preparation?

32. Are any computers available for students to use <u>in class</u>? How many?

33. Are any computer made available for students to use outside class hours?

Test-taking strategies **NB: They don't have to answer this if they don't want to:**

34. Have you heard of any methods for cheating which candidates try?

35. Is it common?

36. Why do you think some candidates might try to cheat?

NT awareness

37. Have you heard about the **new version of TOEFL**, which will be introduced next year? *(if not, finish here)*

38. *If so:* What do you know about it?

39. *If so: H*ow did you find out about it?

40. *If so:* As far as you are aware, how does it differ from the current CBT TOEFL?

41. What's your reaction to the <u>new</u> TOEFL?

42. Do you think you'll have to make changes?

43. How do you feel about that?

Appendix J

Observation Schedule

Teacher:		Date:	
Observer:		Time:	
Institution:		City:	

What are the planned objectives of this lesson? To be completed with the teacher BEFORE the lesson
Remember to ask for copies of teaching materials.

How long before the students take the exam?	_____ days/ _____ months/ not known	Are they also taking TSE?	Yes/No
Is the institution a TOEFL test center?	Yes/No	Length of course:	(# of hours)
Stage of course	beginning/middle/end	Frequency of lessons:	(times per week)
Intended length of lesson:	(minutes)	Class length:	(minutes)
How many students in the class will take TOEFL?		Number of students registered:	
How many students in the class have taken TOEFL before?		Number of students present:	

KEY

S work mode: Individual Pair Group Class	Medium: Computer Paper & Pen	Language: L1 L2	Atmosphere: ☺ etc.	Skill focus: Integrated Single

Start time:	End time:	S work mode: I P G C	Medium: Com PP	Language: L1 L2	Atmosphere:	Skill focus: Int Sing
LISTENING					✓	
Activities:			**Notes:**			
						Identifying main ideas (gist)
						Identifying supporting details
						Identifying order in a process
						Making inferences
						Categorizing topics/objects
						Identifying pragmatic function (NT)
						Identifying pragmatic stance (NT)
						Linking content (NT)
						Identifying the meaning of specific lexis
						Identifying the meaning of colloquialisms
						Multiple choice
						Selecting a visual
						Selecting more than one answer
						Matching (objects/text to categories, etc.)
						Short answer
						Selecting a negative response
						Translating L1L2
						Translating L2 L1
						Summarizing
						Both reading and hearing directions
						Both reading and hearing questions
						Both reading and hearing response options
						Looking at visuals while listening

Start time:	End time:	S work mode: I P G C	Medium: Com PP	Language: L1 L2	Atmosphere:	Skill focus: Int Sing
			READING		✓	
Activities:				**Notes:**		
						Identifying main ideas (gist)
						Identifying supporting details
						Identifying factual information
						Making inferences
						Identifying paraphrasing in the text
						Identifying synonyms (words or phrases)
						Identifying vocabulary from context
						Examining text organization— coherence & cohesion
						Linking info between texts (NT)
						Looking at informal and formal registers (NT)
						Looking for author purpose (NT)
						Scanning
						Identifying reference
						Multiple choice
						Selecting a negative response
						short answer
						Inserting text in correct place
						Selecting specific words/phrases, etc. to respond
						Making/completing charts from info in text
						Making/completing timelines from info in text
						Translating L1L2
						Translating L2 L1
						Summarizing
						Reading texts of 250—35- words
						Reading texts on US culture

172

Start time:	End time:	S work mode: I P G C	Medium: Com PP	Language: L1 L	Atmosphere:	Skill focus: Int Sing
WRITING					✓	
Activities:			Notes:			
						Generating ideas
						Organizing ideas
						Developing ideas
						Supporting ideas with examples/evidence
						Selecting appropriate vocabulary
						Developing sentence structure
						Writing essays
						Writing an essay—no time limit
						Writing an essay in time limit
						Writing essays—no word limits
						Writing essays with word limits (NT)
						Writing essay based on a listening (NT)
						Writing an essay based on reading (NT)
						Organizing ideas from listening/reading before writing (NT)
						Writing on topics from ETS pool
						Writing on topics selected by teacher
						Writing on topics selected by student(s)
						Examining ETS scoring scale
						Synthesizing data from 2 or more texts

173

Start time:	End time:	S work mode: I P G C	Medium: Com PP	Language: L1 L2	Atmosphere:		Skill focus: Int Sing
STRUCTURE					✓		
Activities:			Notes:				
							Grammar taught explicitly
							Work on spelling
							Work on plural forms
							Work on word order
							Work on tenses
							Work on grammar points:
							Multiple choice grammar exercises
							Completing sentences from choices provided
							Completing sentences (no choices provided)
							Identifying incorrect words/phrases from choices provided
							Identifying incorrect words/ phrases in a passage (no choices of proofreading)

174

Start time:	End time:	S work mode: I P G C	Medium: Com PP	Language: L1 L2	Atmosphere:		Skill focus: Int Sing
		SPEAKING			✓		
Activities:			Notes:				
							Organizing ideas before speaking
							Discussing material from listening passage
							Discussing material from reading passage
							Narrating (TSE)
							Recommending (TSE)
							Persuading (TSE)
							Giving an opinion (TSE)
							Supporting an opinion (TSE)
							Responding to cues (TSE)
							Narration based on a story (TSE)
							Giving directions (based on a map)
							Giving information based on a graph/chart
							Speaking activities
							TAST type tasks

To be completed by the observer AFTER the lesson

What text types were used?

READING Put topic on boxes	Topic:	Auth?	Acad?	LISTENING Put topic on boxes	Topic:	Auth?	Acad?
Exposition				Mini-lecture			
Argumentation				Academic discussions			
Historical							
Auto-/ Biographical							

What teaching materials & resources were used? Please attach copies

What % of the teacher talk was in English?	%
Explain in what circumstances the teacher did not speak in English	

Did the students use English?	No	Yes		
		rarely	sometimes	Mostly

What percentage of the lesson time was spent on the following?			
Listening	%	Writing	%
Reading	%	Speaking	%
Other:	%		

Did the teacher set any homework?	Yes/No
If so, please give a description:	

Do you think you observed the influence of the TOEFL exam on the methodology in this lesson (i.e., <u>how</u> the teacher taught)?	Yes	No
Please explain		

Do you think you observed the influence of the TOEFL exam on the content in this lesson (i.e., _what_ the teacher taught)?	Yes	No
Please explain		
Please note here anything else, which is not covered above, that you think might be of interest regarding TOEFL washback		
Please note here any documents attached to this document		
Teaching materials/lesson plan/marking guidelines/assessment guidelines/tests class handouts/other _____		
Please note below any problems you had using this instrument.		

PLEASE DRAW A CLASS MAP HERE:

Appendix K

Data Analysis Codes

Table K1

Antecedents

Traditional pedagogic practices re. current TOEFL		Characteristics of the user	
Aim	Course aims	TAEd	Teacher's attitude towards education
Typ	Course type	TACrmT	Teacher's attitude towards classroom teaching
CAss	Classroom assessment	TALT	Teacher's attitude towards language teaching
Ct	Content—general	TAEx	Teacher's attitude towards exams
CMat	Re. materials	TAEng	Teacher's attitude towards English
CLang	Re. language areas general	TAIds	Teacher's attitude towards new ideas
CRd	Reading	TLEd	Teacher's level of education
CWr	Writing	TAb	Teacher's abilities
CLs	Listening	TAbT	Teacher's technological abilities
CSp	Speaking	TPL	Teacher's personal life
CVo	Vocabulary	TEcon	Teacher's economic situation
CGr	Grammar	TInt	Teacher's interests
CTTT	Re. test-taking techniques	TPsG	Teacher's personal goals
		TClG	Teacher's goals for class
Mth	Methodology—general	TAw	Teacher's awareness of current exam
MMat	re. materials	TAwN	Teacher's awareness of New TOEFL
MLang	re. language areas general	TAtt	Teacher's attitude towards TOEFL
MRd	Reading	TAttN	Teacher's attitude towards New TOEFL
MWr	Writing	TATC	Teacher's attitude towards TOEFL classrooms
MLs	Listening		
MSp	Speaking	SAEd	Student's attitude towards education
MVo	Vocabulary	SACrmT	Student's attitude towards classroom teaching

(Table continues)

Table K1 (continued)

Traditional pedagogic practices re. current TOEFL		Characteristics of the user	
MGr	Grammar	SALT	Student's attitude towards language teaching
MTTT	Re. test-taking techniques	SAEx	Student's attitude towards exams
MRol	Re. role of teacher	SAEng	Student's attitude towards English
MMan	Re. Classroom management	SAIds	Student's attitude towards new ideas
MInt	Integrated skills	SLEd	Student's level of education
		SAb	Student's abilities

Characteristics of the user system		Characteristics of the user	
EdAd	Education administration	SAbT	Student's technological abilities
Sch	School factors	SPL	Student's personal life
SchT	Technology in school	SEcon	Student's economic situation
Crm	Classroom factors	SInt	Student's interests
Pol	Political factors	SPsG	Student's personal goals
Geo	Geographical factors	SClG	Student's goals for class
Econ	Economic factors	SAw	Student's awareness of current exam
Cult	Cultural factors	SAwN	Student's s awareness of New TOEFL
Sch Tr	School-based training	SAtt	Student's attitude towards TOEFL
		SAttN	Student's attitude towards New TOEFL
		SOOC	Student's out of class preparation activities
		SATC	Student's attitude towards TOEFL classrooms
		DAwN	DOS' s awareness of New TOEFL
		DAttN	DOS' s attitude towards New TOEFL
		DALT	DOS's attitude towards language teaching
		DAw	DOS' s awareness of TOEFL
		DACrmT	DOS's attitude towards classroom teaching
		DAIds	DOS's attitude towards new ideas
		DAbT	DOS's technological abilities
		DATC	DOS's attitude towards TOEFL classrooms
		DAEx	DOS's attitude towards exams
		Vers	Version of TOEFL taken

Table K2

Process

Characteristics of the resource system	
Cap	Capacity
St	Structure
Op	Openness
Hy	Harmony

Communication of the innovation	
Comm	Communication
Orig	Originality

Characteristics of the innovation	
Comx	Complexity
Expl	Explicitness
RelAd	Relative Advantage
Tri	Trialability
Obs	Observability
Stat	Status
Pra	Practicality
Flex	Flexibility
Prim	Primacy
Fm	Form

Appendix L

Observation Summaries: Examples of TOEFL and Non-TOEFL Classes

Teacher TL2: TOEFL Class

Table L1

TOEFL Class Information

Teacher ID:	TL2	
Institution:	xxx Institution	
Town & country:	xxx	Lithuania
Date:	11/28/03	
Class timing:	18.00–20.00	
Teacher profile:	NNS V. Exp	Other: EAP training and exp.
Number of students:	5	
TOEFL/Non-TOEFL (+ level):	TOEFL	
Materials used:	*Master the TOEFL CBT 2004* (Sullivan, Brenner, & Zhong, 2003)	
Stage of the course:	5	(Beginning = 1 End= 5)
Length of course (hours):	36 hrs	
Frequency of classes (per week):	2/wk	
TOEFL only:		
Any students taking TSE:	**Yes/No**	**No**
Are all students planning to take TOEFL:	**Yes/No**	**Yes**

Table L2

TOEFL Class Activities

Time	Activity	Focus	Work mode	Inte-grated?
18.05	Intro: will use a test (from the coursebook) plus extra grammar-based materials today. Ss do test (p. 119).	St	I–C	No
18.25	T stopped class—some had finished. Checked answers as a group. Some expansion of problem items.	St	C	No
18.32	Look at ex. on word order (pp. 168–169)—focus on key words that signal abnormal word order.	St	C	No
18.43	Students do ex. (p. 172) and then check responses—discussion of problems.	St	I–C	No
18.52	Focus on conjunctions (p. 175)—look at sample items, explanation, and translation by T; Ss do ex. 28 (p. 176); checked answers by T giving correct answers (no S input).	St	C–I–C	No
	Revision of *do* and *make*—T elicited common collocations.	St	C	No
19.03	Focus on comparisons (p. 178)—checks farther/ further (even tho' book notes this will not be tested)—review by T. No exercises done.	St		No
19.08	Test (p. 123) 20 minutes. Set to do q. 46–55 (on p. 132). Check and discuss.	Rd	I–C	No
19.27	Q. 24-35 set (p. 128)—15 minute limit. Check and discuss.	Rd	I–C	No
19.51	T returned marked essays written at home. Gave tem to read through them briefly. General comments from T (no grades given).	W	I	No
20.00	Close			
HW:	None?			

Questions

- If TOEFL class: Influence of TOEFL observed?

  Yes—very much. Exercise types, tips/advice all TOEFL-oriented, L1 only used—no focus on developing speaking skills.

- Were computers used in this class?

  No

- What was the atmosphere like in general? Fun/Neutral/Serious:

  Neutral

- Did it change throughout the class?

  No

- When was L1 used/when was L2 used?

  L1 throughout

- Use of textbook?

  Yes—throughout

- What was the class layout like? Rows/Circular/Other:

  Around a large table—T at front

- Other points:

  Ss very focused—little S-initiated talk—mostly T-led

Teacher TL2: Non-TOEFL Class

Table L3

Non-TOEFL Class Information

Teacher ID:	TL2	
Institution:	xxx Institution	
Town & country:	xxx	Lithuania
Date:	11/28/03	
Class timing:	14.15–15.30	(usually 2 academic hours—extraordinary class)
Teacher profile:	NNS V. Exp	Other: EAP training and exp.
Number of students:	6	
TOEFL/Non-TOEFL (+ level):	Non-TOEFL	Public administration group
Materials used:	Ts' materials from Web	
Stage of the course:	5	(Beginning = 1 End= 5)
Length of course (hours):	15 wks (1 semester)	
Frequency of classes (per week):	2/wk	
TOEFL only:		
Any Ss taking TSE:	**Yes/No**	**NA**
Are all Ss planning to take TOEFL:	**Yes/No**	**NA**

Table L4

Non-TOEFL Class Activities

Time	Activity	Focus	Work mode	Inte-grated?
14.15	Ss read text for main points. Ss asked to pick out sentences they found most informative—and then any they had difficulty with. Analyzed difficulty in understanding final para.: a gen. problem.	R (+V, S)	C	No
	Ss worked in pairs to analyze other paras.— suggested use of a diagram.	R	P	No
14.50	Ss asked to reformulate text based on the outlines without referring back to text.	W	P?	No
15.03	Vocabulary check—including some idioms—T elicited ideas from Ss.	Voc	C	No
15.08	T gave headings of paras. From original text and Ss had to reorder them using their summaries.	R/W?	P	No
15.26	Ss given another long reading—Ss chose which subsection to read. To complete at home?	R		No
	T summarized lesson—what covered today and what will come next.	-	C	No
15.35	End			
HW:	Using headings from text to form a summary paragraph. To complete reading chosen subsection of text	W		

Questions

- If TOEFL class: Influence of TOEFL observed?

  No

- Were computers used in this class?

  No

- What was the atmosphere like in general? Fun/Neutral/Serious:

  Neutral

- Did it change throughout the class?

  No

- When was L1 used/when was L2 used?

  L2 throughout

- Use of textbook?

  No

- What was the class layout like? Rows/Circular/Other:

  Around a large table

- Other points:

  Pace of class seemed slow but students enthusiastic. One focus only (reading).

Appendix M

Skills and Activities Observed in TOEFL Preparation Classes

	TB1[a]	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1[b]
Listening	✓	✓[c]	✓			✓	✓		✓
Identifying main ideas (gist)						✓			
Identifying supporting details						✓			
Identifying order in a process									
Making inferences							✓		
Categorizing topics/objects									
Identifying pragmatic function (NT)							✓		
Identifying pragmatic stance (NT)									
Linking content (NT)									
Identifying the meaning of specific lexis			✓?						
Identifying the meaning of colloquialisms							✓		
Multiple-choice			✓			✓	✓		
Selecting a visual									
Selecting more than one answer									
Matching (objects/text to categories, etc.)									
Short answer							✓		
Selecting a negative response									
Translating L1–L2			✓						
Translating L2–L1									
Summarizing									

(Table continues)

Table (continued)

	TB1[a]	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1[b]
Both reading and hearing directions									
Both reading and hearing questions									
Both reading and hearing response options									
Looking at visuals while listening									
Reading just the responses			✓						
Brainstorming			✓						
Reading	✓		HW		✓	✓			✓
Identifying main ideas (gist)					✓				
Identifying supporting details					✓				
Identifying factual information				✓	✓				
Making inferences					✓				
Identifying paraphrasing in the text				✓					
Identifying synonyms (words or phrases)					✓				
Identifying vocabulary from context									
Examining text organization—coherence & cohesion									
Linking info between texts (NT)									
Looking at informal and formal registers (NT)									
Looking for author purpose (NT)					✓				
Scanning					✓				

(Table continues)

Table (continued)

	TB1[a]	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1[b]
Identifying reference					✓				
Multiple choice					✓				
Selecting a negative response					✓				
short answer									
Inserting text in correct place					✓				
Selecting specific words/phrases/sentences/paragraphs, etc. to respond (e.g., finding synonyms/antonyms)	✓								
Making/completing charts from info in text									
Making/completing timelines from info in text									
Translating L1–L2									
Translating L2–L1					✓				
Summarizing									
Selecting a visual	✓								
Reading texts of 250–350 words					✓				
Reading texts on U.S. culture									
Writing			✓	✓		✓+HW		✓	✓
Generating ideas				✓		✓		✓	
Organizing ideas			✓						
Developing ideas				✓					

(Table continues)

Table (continued)

	TB1[a]	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1[b]
Supporting ideas with examples/evidence				✓					
Selecting appropriate vocabulary									
Developing sentence structure									
Writing essays			✓						
Writing an essay—no time limit				✓					
Writing an essay in time limit									
Writing essays—no word limits									
Writing essays with word limits (NT)									
Writing essay based on a listening (NT)									
Writing an essay based on reading (NT)									
Organizing ideas from listening/reading before writing (NT)									
Writing on topics from ETS pool			✓						
Writing on topics selected by teacher				✓					
Writing on topics selected by student(s)									
Examining ETS scoring scale									
Synthesizing data from 2-or more texts									
Revising possible structure for essay				✓					
Examining model essays								✓	
Asking Ss to say what they recall						✓			

191

(Table continues)

Table (continued)

	TB1[a]	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1[b]
Reading essay aloud—to partner/to group						✓			
Writing introduction only						✓			
Structure	✓		✓	✓+HW	✓			✓	✓+HW
Grammar taught explicitly									
Grammar revised explicitly					✓				
Work on spelling									
Work on plural forms			✓						
Work on word order					✓				
Work on tenses								✓	
Work on grammar points:									
Word forms—parts of speech			✓						
Verb/noun agreement				✓					
Subject/auxiliary verb inversion				✓					
Verb forms				✓					
Conjunctions					✓				
Comparisons					✓				
Adjective forms								✓	
Conditionals								✓	
Set phrases taking gerund/ infinitive								✓	
Multiple-choice grammar exercises				✓	✓				

(*Table continues*)

192

Table (continued)

	TB1[a]	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1[b]
Completing sentences from choices provided					✓				
Completing sentences (no choices provided)								✓	
Identifying incorrect words/phrases from choices provided					✓			✓	
Identifying incorrect words/phrases in a passage (no choices cf. proofreading)									
Identifying errors in a sentence (no choices provided)				✓					
Identifying errors in a sentence (no choices provided) and correcting			✓	✓					
Translating L2 -> L1					✓				
Transforming verbs to correct form								✓	
Speaking									
Vocabulary				✓				✓?	✓

Note. ✓ = Observed**.** HW = Homework task set. NT = New TOEFL.

[a] TB1 = Hard to analyze. Students did practice test. Observer not able to go through the test and check item types, etc. Did not want to watch any one student too closely—too off-putting. [b] TS1 = No TOEFL-like activities. [c] In the second half of the lesson students worked independently at computers on practice test or exercises—therefore various foci.

Appendix N

Teachers' Views of What Is Tested on TOEFL

(Teacher Interview: Question 95 A to S: What Does TOEFL Test?)

	TB1	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1	Our answer	ETS answer
A Use grammar correctly	yes if they can recognize it then they can use it	yes	yes	yes but sometimes tests things they may not know (e.g., inverting subject/ verb)	yes to some extent	yes	yes	yes	no	yes?	yes
B Use wide range of vocabulary correctly	yes	yes	yes	yes	no	no (yes—to understand it)	yes	yes	yes	yes?	no
C Use idioms correctly	yes the more popular ones and to recognize rather than use	yes in the listening section and to in reading to some extent	no	yes tests their nonability to use them!	no to recognize, not use—in the listening section	no	yes	yes	no	no	no
D Understand wide range of texts	yes	yes	yes?	yes	yes	yes	[misunder-stood questions?]	yes	yes	yes?	no (if = text types) Topics= yes

(*Table continues*)

Table (continued)

	TB1	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1	Our answer	ETS answer
E Express original ideas in writing	yes	yes	no good written English more important	yes	no	no	not really (depends on definition of original ideas)	no	It depends: not all do TWE. If not—no. In TWE flow of though tested more than original ideas.	no?	yes
F Translate, L1-L2 and L2-L1	no	yes	no	(response referred to prep classes, not test)	no	no	no	no	no	no	no
G Take active part in academic discussion	no	no	no	yes? indirectly— test exposes them to broad range of topics	no	yes	no	no	no	no	no
H Understand lectures	yes	yes	no specifically —general under- standing	yes	yes	yes	yes	yes and no 50/50	yes	yes	yes
I Infer opinion	yes	yes	yes? [not clear]	yes [but admits he's guessing]	yes	yes	yes	yes	yes	yes	yes
J Understand organization of text	yes	yes	yes in organizing the essay	yes? [not clear]	yes	yes	no	no	yes	yes	yes

195

(*Table continues*)

Table (continued)

	TB1	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1	Our answer	ETS answer
K Write formal letters	no	no	no	no	no	no	no	no	no	no	no
L Analyze information from several texts	no	no	no [misunder-stood question?]	yes	no	yes	no	no	no	no (was proposed for new TOEFL— not now included)	no
M Make inferences from text	yes	yes	yes in reading and some in listening	yes [but again seems to be guessing]	no but in some reading questions [so=yes?]	yes	yes	yes	yes	yes	yes
N Give a presentation	no	no	no	no (but the necessary skills— vocabulary and organization —are tested	no	no	no	no	no	no	no
O Understand unfamiliar vocabulary from context	yes	yes	yes	yes	yes	yes	no	yes	no	yes	yes

(*Table continues*)

Table (continued)

	TB1	TB2	TC1	TL1	TL2	TP1A	TP1B	TR1	TS1	Our answer	ETS answer
P State opinion and	yes	yes in the essay	yes in the essay	yes in the essay	yes	yes	yes	yes	yes in written part sometimes, depending on the question	yes	yes writing only
Q Understand language in everyday situations	yes	yes	yes in Listening Part A	yes	yes	no	yes	yes	yes	yes	no
R Speak for extended period on familiar topic	no	no	no	no	no	yes	no	no	no	no (yes in new TOEFL)	no
S Write academic article	no	no	yes formality and academic level required [but this not necessarily same]	yes [seems to be guessing again]	no	yes	yes	no	yes	no	yes no directly but correlates highly [evidence?]

Table continued

197

Note. [ ] = our comments; ( ) = respondent's comments.

Allocation of Time in TOEFL Classrooms
(Teacher Interview Questions 21 and 31, Student Interview Question 14)

Most time spent on:	Grammar	Vocabulary	Listening	Writing	Reading	Speaking	Equal on all sections
TB1	most				most		
SB1	most S3	most S1&S3	most S3		most S1		
TB2			most		least		
SB2			most				
TC1						least	equal, but not speaking
SC1							
TC2 (No Ss)			most	least			
TL1	most	most					
SL1	most						
TL2	most		least	most			
SL2	most S1		most S2	most S2	most S1 & S2		
TP1A						least	equal, but not speaking
SP1A				most	most		
TP1B	least		most[a]				
SP1B				most S2 & S3	most S2	most S1	
TR1	most				least		
SR1	most	most					
TS1	least	most					
SS1		most S1 & S2	most S1 & S3	most S1 & S3			

Note. No entry means no data found. most = most time spent on this area; least = least time spent on this area, teacher's view only (students not asked about this). S1 = Student 1; S2 = Student 2; S3 = Student 3 (numbers given if students within the group hold different opinions).

[a] This teacher was teaching the course for the first time so did not have a clear idea of what she would be spending most time, but predicted it would be listening. Her students would not have known this was her prediction.

Appendix P

Preparation Coursebooks Used in TOEFL Classrooms

De Souza, D. (1996). *Penguin practice book for TOEFL*. New York: Penguin.

Duffy, C., & Mahnke, K. (1996*). Heinemann TOEFL preparation course*. New York: Macmillan.

Gear, J., & Gear, R. (2001). *Cambridge preparation for the TOEFL test* (3rd ed.). Cambridge, UK: Cambridge University Press.

Kaplan. (2002). *TOEFL CBT with CD-ROM*. Lawrenceville, NJ: Author.

Phillips, D. (2000). *Longman complete course for the TOEFL test*. Reading, MA: Addison-Wesley.

Rogers, B. (2000). *Heinle & Heinle's complete guide to the TOEFL test, CBT edition*. Boston: Heinle & Heinle.

Rogers, B. (2002). *TOEFL CBT success 2003*. Lawrenceville, NJ: Peterson's.

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

Sharpe, P. J. (1996). *Barron's how to prepare for the TOEFL test* (7th ed.). Hauppauge, NY: Barron's.

Sharpe, P. J. (1999). *Barron's how to prepare for the TOEFL test* (8th ed.). Hauppauge, NY: Barron's.

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

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