TOEFL®

Monograph Series

MS - 10
MAY 1997

Communicative Language Proficiency: Definition and Implications for TOEFL 2000

Carol Chapelle

William Grabe

Margie Berns

Ⓔ Ⓣ Ⓢ
**Educational
Testing Service**

Communicative Language Proficiency:
Definition and Implications for TOEFL 2000

Carol Chapelle, William Grabe, and Margie Berns

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

E-mail: toefl@ets.org

Web Site: http://www.toefl.org

Foreword

The TOEFL® Monograph Series features commissioned papers and reports for TOEFL 2000 and other Test of English as a Foreign Language program development efforts. As part of the foundation for the TOEFL 2000 project, a number of papers and reports were commissioned from experts within the fields of measurement and language teaching and testing. The resulting critical reviews and expert opinions were invited to inform TOEFL program development efforts with respect to test construct, test user needs, and test delivery. Opinions expressed in these papers are those of the authors and do not necessarily reflect the views or intentions of the TOEFL program.

These monographs are also of general scholarly interest, and the TOEFL program is pleased to make them available to colleagues in the fields of language teaching and testing and international student admissions in higher education.

The TOEFL 2000 project is a broad effort under which language testing at ETS will evolve into the 21st century. As a first step in the evolution of TOEFL language testing, the TOEFL program recently revised the Test of Spoken English (TSE®) and announced plans to introduce a TOEFL computer-based test (TOEFL CBT) in 1998. The revised TSE test, introduced in July 1995, is based on an underlying construct of communicative language ability and represents a process approach to test validation. The TOEFL CBT will take advantage of the new forms of assessments and improved services made possible by computer-based testing while also moving the program toward its longer-range goals, which include

- the development of a conceptual framework that takes into account models of communicative competence
- a research agenda that informs and supports this emerging framework
- a better understanding of the kinds of information test users need and want from the TOEFL test
- a better understanding of the technological capabilities for delivery of TOEFL tests into the next century

It is expected that the TOEFL 2000 efforts will continue to produce a set of improved language tests that recognize the dynamic, evolutionary nature of assessment practices and that promote responsiveness to test user needs. As future papers and projects are completed, monographs will continue to be released to the public in this new TOEFL research publication series.

TOEFL Program Office
Educational Testing Service

Abstract

Discussion of TOEFL 2000 in the TOEFL Committee of Examiners' (COE) meetings resulted in a framework representing components believed to be relevant to defining language use in an academic context. The framework, called the COE Model, is comprised of aspects of the context of language use as well as hypothesized capacities of the language user. The COE Model suggests that test development should begin by examining the types of academic contexts in which language is used in order to hypothesize what those abilities may be for any specific context of interest. COE discussions of TOEFL 2000 were motivated by a broad range of validity concerns (e.g., content validity, construct validity, and the social consequences of test use), and the Model may have implications for how validation of TOEFL 2000 is conceived. The COE model is described to serve as a record of past discussion which can inform future work.

Acknowledgment

Table of Contents

List of Figures

List of Appendices

1. Introduction

Over the past several years, the TOEFL Committee of Examiners (COE) has discussed TOEFL 2000, a test whose tentative purpose is the following:

> TOEFL 2000 is a measure of communicative language proficiency in English and focuses on academic language and the language of university life. It is designed to be used as one criterion in decision making for undergraduate and graduate admissions.

Because the intended purpose of TOEFL 2000 is to test communicative language proficiency for academic life, the COE discussions of TOEFL 2000 have focused primarily on how to define "communicative language proficiency for academic life." These discussions have produced a framework for such a definition that has been codified as a schematic diagram representing components believed to be relevant, as well as hypothesized relations among the components. This diagram, called the "COE Model," has been useful within the COE meetings to focus discussion on how to define what TOEFL 2000 is intended to measure, and it may be useful for discussion beyond the COE.

The purpose of this paper is to introduce the COE Model. We first present the background of the COE Model and the assumptions committee members brought to these discussions. The major portion of the paper explains the COE Model, defining its components and how they are hypothesized to work together; it also addresses many unresolved issues. We then suggest implications of the model for test development and for validation of TOEFL 2000. We conclude by restating the Model's purposes, which should continue to motivate and direct its evolution.

2. Background and Assumptions

2.1 Why a Model of Language Use in Context?

From the first discussions of TOEFL 2000 at COE meetings, all those involved have eagerly anticipated possible solutions to questions about the new test. What would it look like? Would it test the "four skills"? Would it provide different versions for students in different subject areas? What would the item formats look like? How could technology be used? How could score meanings at different levels be defined? These questions have consistently directed discussion to two fundamental questions:

- What is the intended use of TOEFL 2000?
- What is TOEFL 2000 intended to measure?

The COE addressed the first question by drafting the tentative statement of intended use for TOEFL 2000 (see Section 1). This statement, in turn, became a guiding assumption for considering the second question. Discussions of the second question have resulted in a framework for defining communicative language proficiency in academic contexts, called the "COE Model." The COE Model explained in this paper is a descriptive sketch that reflects discussions at several meetings, as well as additional minor elaborations by individual COE members. (Appendix A contains more details about this process.) The Model attempts to summarize existing research and current assumptions by researchers in cognitive psychology, applied linguistics, and language testing. The COE does not view the Model described in this paper as a definitive or final version of the framework that will prove most useful for test development and validation. Instead, it represents an interpretation of communicative language use in a form that should facilitate future discussion.

2.2 COE Assumptions About a Definition of Language Ability

Language can be discussed from a variety of perspectives. For TOEFL 2000, we believe it is essential to define language in a way that is consistent with the views of professionals in applied linguistics and in a way that will be useful for test development and validation. The theory of language informing the COE's goals for TOEFL 2000 draws on the work of many applied linguists and language specialists who have been concerned at least since the 1970s with the interactive nature of language — interaction between and among speakers of a language, as well as between the speakers and the context in which they use language (e.g., Hymes, 1971; Halliday, 1978). The approach to language that these scholars take has been termed "functional" to capture their focus on language use (as opposed to focus on form, exemplified by linguists such as Chomsky [1965] and his followers).

Although several interpretations of "communicative competence" have been offered (see e.g., Campbell & Wales, 1970; Habermas, 1970), it is Hymes' (1971) interpretation that has been most commonly used in the United States. Coined to describe language use from an ethnographer's perspective, the term subsequently was interpreted for pedagogical purposes by many language teaching specialists (e.g., Canale & Swain, 1980; Canale, 1983; Munby, 1978; Savignon, 1983). The more familiar description, outlined in the seminal work of Canale and Swain (1980) and in a paper by Canale (1983) defines four components of communicative competence: (1) sociolinguistic competence, referring to knowledge required for understanding the social context in which language is used — the roles of the participants, the information they share, and the function of the interaction; (2) grammatical (linguistic) competence, including

2

knowledge of grammatical well-formedness; (3) strategic competence, referring to the strategies one uses to compensate for imperfect language knowledge or other limiting factors (such as fatigue, distraction, and inattention); and (4) discourse competence, comprising knowledge of the connections among utterances in a text to form a meaningful whole. A useful extension of this work for language testing is Bachman's (1990) description of a more specific model of language ability, which hypothesizes how Canale and Swain's four competencies work together in language use and which expresses an explicit relationship between "context" and the competencies. The COE Model, presented in Section 3, follows directly from the Hymes, Canale and Swain, and Bachman conceptions of language.

2.3 Assumptions About Testing Language Ability

Throughout discussions of TOEFL 2000, the COE has focused on the test's validity as the primary concern. In concert with other applied linguists (e.g., Bachman, 1990; Shohamy, 1993), the COE has viewed "validity" in the broad sense as referring not only to construct validity but also to evidence about relevance and utility, the value implications TOEFL 2000 will reflect, and the social consequences of TOEFL 2000 use. The fact that the initial steps in test design have been occupied with construct definition (i.e., elaborating the COE Model) reflects the COE's conviction that decisions at this stage will provide the essential foundation for test development, construct validation, and other validity justifications. In Sections 4 and 5, we speculate on implications of the COE Model for test development and for validity inquiry.

3

3. COE Model

The schematic diagram in Figure 1 identifies significant variables that affect language use (both comprehension and production) in academic contexts. This model distinguishes the context (above the line) from the individual language user (below the line). The context (3.1, in the nonshaded area above the line) includes those elements of language use external to the language user, many of which are observable to others in the act of communication (e.g., the setting in which communication takes place and the language that the individual contributes to that setting). Below the line are the individual's capacities (3.2, internal operations) which work in concert to interpret and produce language in context. We will describe the model and how it works by beginning with the features of the context that we believe call on specific capacities defined within the internal operations.

FIGURE 1

Working Model of Communicative Language Use in an Academic Context

CONTEXT (3.1)	

SITUATION (3.1.1)
 	
SETTING (3.1.1.1)
PARTICIPANTS (3.1.1.2)
TASK (3.1.1.3)
TEXT (3.1.1.4)
  - key
  - act sequence
  - norms of interaction & interpretation
  - instrumentalities
  - genre
TOPIC (3.1.1.5)
	PERFORMANCE (3.1.2)

INTERNAL GOAL-SETTING 3.2.1

INTERNAL PROCESSING OUTPUT 3.2.6

INTERNAL OPERATIONS (3.2)

VERBAL PROCESSING COMPONENT

- metacognitive processing
- on-line processing
3.2.3

VERBAL WORKING MEMORY 3.2.2

LANGUAGE COMPETENCE

- linguistic
- discourse
- sociolinguistic
3.2.4

WORLD KNOWLEDGE 3.2.5

3.1 Context

Throughout applied linguists' discussion of communicative competence, the notion of context has been emphasized as an essential element for understanding language and language ability. Reflecting this perspective, the COE Model specifies that all language processing is initiated in some way by the context. The individual is in a given situation and must use language to communicate, whether the communication is due to a conversational partner, a task assigned by another person, a need to respond to a remote event or topic, or a need to inform or entertain oneself. The primary assumption is that all language use is, at least remotely, based on a need to communicate (even if with oneself; cf. Crystal, 1987).

The context in which communication takes place is crucial throughout language development. Native speakers of a language develop their communicative competence through participation in the social and cultural life of their family, friends, teachers, and neighbors. Speakers develop the ability, referred to here as communicative competence, to communicate appropriately and grammatically correct in a variety of situations. This ability is complex, consisting of many interacting components or abilities. These interrelated abilities are activated by various features of the environment surrounding a language user. The users, whether interpreting discourse through reading and listening, or expressing themselves through writing or speaking, are engaged in an ongoing and dynamic process of assessing relevant information available in the environment or in negotiating the meanings expressed. Given the crucial role of context in communication, in defining communicative language proficiency we must address these questions: What do we mean by the word "context"? What features of context are relevant to language use?

The term "context" refers to the environment of a text. Both concrete and abstract aspects of context are relevant to communicative competence. Concrete aspects of context include the physical setting, the specific place where communication occurs, and those observable features that represent a "concrete" sense of context. Abstract aspects of context refer to such features as the status and roles of the participants (e.g., the instructor and student), knowledge that the participants share, the verbal and nonverbal actions of the participants (e.g., listening to a lecture, writing answers to a quiz, carrying out an experiment), and the effects of the verbal actions, or the changes they bring about as a result of a participant having said a particular thing (e.g., a certain step being taken in an experiment, clearer understanding as a result of an instructor's answer to a student's question). Features of the concrete context (such as test tubes and beakers in a chemistry lab, an outline on a blackboard, or the chairs that have been moved into a circle for a class discussion) also may be part of the abstract context, but only if they directly influence the activity the participants are involved in. These features of a speech event — both the abstract and the relevant concrete features — are referred to as the "context of situation". This abstract sense of context is associated with a Firthian approach to linguistics (Firth, 1957; Halliday, 1978; Halliday & Hasan, 1989; see also Malinowski, 1923).

The abstract sense of context is important because the performance of a language user may not necessarily be tied to a physical setting. A context consists of more than the observable. Moreover, the physical setting of a situation is not always relevant to communication. For example, a student's performance may be represented in a letter of complaint to a car-rental agency, although the physical setting may be a lecture in an auditorium where all other students are taking notes. Here, the product and

6

its effectiveness as a letter of complaint are independent of the physical setting. A further example of the possible irrelevance of physical setting is an exchange between a professor and student about an aspect of the day's lecture. The exchange can take place anywhere — in a cafe, on the street, over the phone, or via electronic mail. Here the precise physical setting may be irrelevant to the participants' communication. More relevant may be their status vis a vis one another and the goals each establishes for communication. The abstract sense of context covers the possible independence of the physical setting from performance.

Because of the importance of context in communicative language proficiency, the COE Model identifies specific features of context that allow us to define context and, subsequently, to analyze specific academic contexts of interest to TOEFL 2000 users. The features in the model are based primarily on those identified by Hymes (1972): (1) setting; (2) participants; (3) ends; (4) act sequence; (5) key; (6) instrumentalities; (7) norms of interaction and interpretation; and (8) genre. These eight categories (which can be remembered with the mnemonic "SPEAKING", remain the most useful analysis of context and have been elaborated only slightly (Saville-Troike, 1989; Kramsch, 1993). The use of the features is illustrated and discussed in Subsection 3.1.1.

In the COE Model, the context of interest is the academic context. Academic contexts can be seen as of two types: (1) those relating to university life; and (2) those of scholarship/the classroom. Those of university life are comparable in many respects to situations of daily life off as well as on campus. For example, students meet and converse with others, establish and maintain relationships, and get and give information. One salient feature about the use of language on campus may be the use of vocabulary generally associated with campus and student life (for example, dorm living, registering for classes, dropping a class, flunking an exam, or getting an "A"). This vocabulary marks the interaction as belonging to the campus context. Other features (such as a familiar and friendly tone between students) also may mark the discourse in this way, but it is the use of this vocabulary that is most salient.

The other kind of academic context, the classroom/scholarship context, is marked in a variety of forms, as the contrasting examples of a lecture and a faculty office appointment show. These two examples do not, of course, represent all situations in which students find themselves, especially since the two illustrations primarily involve oral language. A great deal of student use of language and language ability is involved in the interpretation and expression of meaning through written texts. Furthermore, classrooms and faculty offices are not the only settings for linguistic interaction, nor are listening and notetaking. Neither is a request for help and assurance on an assignment the only norm of interaction and interpretation in which students and instructors engage. Because academic contexts differ from one another in important ways, the COE Models specifies a set of features which are important for defining "context."

3.1.1. Situation

In the "Context" section of the model, the left-hand side is labeled "situation." Situation is defined here as including those aspects of the academic situation that are likely to influence academic language use: "setting," "participants," "task," "text," and "topic." The situations "lecture" and "office appointment" are used to illustrate these features.

3.1.1.1. Setting

Setting describes the physical location where communication takes place, where participants are located. The setting for the lecture is typically a classroom or lecture hall; the lecturer delivers the lecture in front of the audience, who may be seated in rows of chairs or at desks. The lecturer may use any of a variety of visual aids (blackboard, overhead projector, or slides). The time devoted to the lecture may be more or less than the class period.

The office appointment takes place in a room in an office building or complex. The room usually has a desk and at least two chairs, bookshelves, books, and other standard faculty office items. The instructor is seated at or behind the desk, the student may be seated beside or facing the instructor. They may be looking together at a textbook or piece of paper with an assignment or quiz on it.

3.1.1.2. Participants

Participants are the individuals involved in the language event. In academic contexts, participants are generally some combination of instructors (professors or teaching assistants) and students (either graduate or undergraduate). Each participant is associated with institutional status and role characteristics. Moreover, these institutionally defined characteristics may be colored by personal features such as age, gender, level of experience, nationality, and familiarity with the other participant(s).

3.1.1.3. Task

A task is a piece of work or an activity with a specified goal (see Long and Crookes, 1992). The definition of "task" in most applied linguistics work refers to getting something done, although some applied linguists are working to refine definitions of "task" (e.g., Duff, 1993; Skehan, 1992; Bachman & Palmer, 1996). The goal, or "ends" in Hymes' terminology, of the lecture is to transmit/receive information on a range of points to be used by the students for a future assignment. The goal of the office appointment is to provide/obtain individual attention that will help the student to understand material needed to write a paper, present an oral report, or take a test.

3.1.1.4. Text

The term "text" refers to the type of language used to complete a task. A task might be completed, for example, through a formal or an informal conversation, a written or orally presented story, or an interview or debate. Text types (e.g., engineering reports, letters of complaint, question-answer exchange sequences, academic advising sequences) can be analyzed using the following of Hymes' features:

Key. The key, or tone, of the lecture is likely to be scholarly, serious, and formal, and perhaps even humorous at times. The consultation may be less formal and scholarly, but is likely to be mostly serious. It also may have a sympathetic tone if the student is concerned and/or upset (for example, about performance in the class). This feature is often referred to in the literature as "register" or "style" (Halliday, 1978; Joos, 1962).

8

Act Sequence. This refers to the form and content of the speech event. Here the lecture and office hour differ; the lecture is likely to contain a large number of facts, illustrations, and examples, as well as the general points being made. The office exchange may also contain facts, but they are likely to be ones repeated from the lecture; new illustrations, however, may be given. In addition, more questions are likely to be raised in the office exchange.

Norms of Interaction and Interpretation. These refer to the rules for language use that apply in this particular event and the information about the speech community and its culture that is necessary for the participants to understand the event. Norms for the lecture in North American culture are likely to be polite listening on the part of the audience, with note taking and perhaps some hands raised for questions. The lecture is generally one-way — from instructor to student. In the office, the participants generally take turns speaking, and either the student or the instructor may initiate the exchange. Standards set by the surrounding speech community determine what is and is not appropriate and acceptable behavior for the particular event.

Instrumentalities. Code and channel are the considerations here. The language of the lecture and the office visit will be oral in channel but may differ in code. The lecture may be delivered in a more formal language or dialect of the community, while the office appointment may be accomplished in the less formal code.

Genre. The genres represented by each of the example speech events are "lecture" and "consultation." Other genres include editorials, scientific abstracts, book reports, business letters, and talk show interviews.

3.1.1.5 Topic

Topic refers to the specific content information that is being addressed by various participants, and to various tasks and texts in the situation. Different topics impact a student's performance on many types of tasks, reflecting different levels of linguistic competence and language-processing abilities. For example, a student who is asked to critique a text, and who has relatively little knowledge of the topic, may rely more heavily on linguistic and textual strategies to compensate for weaker topical knowledge.

3.1.2. Performance

The other element in Figure 1 that is part of "Context" is labeled "Performance," or linguistic and behavioral output. This is the contribution that the language user makes to the context. This contribution may be verbal and in the form of a text (writing an essay or asking a question) or nonverbal (turning to a designated page or following along on a map). The broken line between situation and performance indicates that, while these two elements of context can be analyzed separately, they are interrelated notions. Performance occurs within a situation; a situation can be described in part by linguistic and nonlinguistic performance, or behavior.

3.1.3. Conclusion

This understanding of context is important in constructing a theoretical foundation for TOEFL 2000 because of applied linguists' view of the role of context in defining language use. Each feature of context described previously is important in understanding why language performance is as it is and why a particular text or discourse takes the form it does, has the intent it has, and performs one function and not another. Each aspect of context also illustrates ways in which sentences out of context are not typical of language use. Meaning in language use is derived from the complex of features that describe a situation. The whole of a discourse is more than the sum of the parts. It is insufficient to look at the parts of the discourse and to decode the meaning of the words and the syntax in order to determine the meaning. The meaning is dependent on each sense of context; meaning can be determined from the physical setting and from the relationship of the participants to the situation, to one another, and to the task and text relevant to the situation.

3.2. Internal Operations

"Internal operations" refer to the processing that goes on in the mind during communicative language use. All of the space below the line in Figure 1, representing the internal processing of the individual, is set within some mental space (referred to as "verbal working memory") that includes internal goal setting, verbal processing, and internal processing output. The subcomponents of verbal processing, shown in boxes partly in the circle, represent those aspects of the components used for the specific processing task.

The internal operations component does not presume either a strong modularity position or a strong distributed processing position. While the schematic diagram suggests some modularity with the boxes and circles, the notion of modularity is both too complex and too unclear to be a controlling metaphor. A more realistic position is the notion that certain aspects of language processing tend to be modular, or encapsulated (cf. Bereiter, 1990; Oakhill & Garnham, 1988; Perfetti, 1989; 1991; Rayner, Garrod, & Perfetti, 1992; Singer, 1990; Stanovich, 1990, 1991, 1992; Walczyk & Royer, 1993). Current evidence argues that word recognition and initial parsing are encapsulated processes for language comprehension. Lexical access and syntactic processing are also likely to be encapsulated to some extent in L1 speech production. The role of modularity in writing or in second-language learning is less clear. Perfetti and McCutchen (1987) argue that the main point of attaining sophisticated writing skills (and a major difference from reading) is that more of the processing becomes open to reflection and manipulation as one becomes more skilled. For second-language learning, the need to attend to the language for learning and the need for proceduralization (Anderson, 1993; Schmidt, 1992, 1993) would argue for a gradual acquired modularity but not necessarily complete modularity.

Beginning with internal goal setting (Subsection 3.2.1), the individual interprets the features of the context and then sets a goal specifically for that situation. The goal setting then activates the appropriate resources in verbal working memory (Subsection 3.2.2), which includes the relevant aspects of the verbal-processing component (Subsection 3.2.3), language competence (Subsection 3.2.4), and world knowledge (Subsection 3.2.5). Within the verbal-processing component, the on-line processing mechanism and metacognitive strategies call on relevant world knowledge and language competence to produce internal-processing output (Subsection 3.2.6) and (sometimes) performance (Subsection 3.1.2). In the following

10

section, we define each of these parts in greater detail and discuss unresolved issues associated with specific components.

3.2.1. Internal Goal Setting

Language use is motivated by an individual's perceptions of and responses to context. Therefore, the COE Model includes an internal goal-setting component responsible for interpreting the context (Douglas, forthcoming), setting the language user's goals, and activating associated plans for achieving those goals (Færch & Kasper, 1983). For example, in an academic lecture, most students set goals such as, "get the important information down in my notes." Several factors influence goal setting. One factor is the language user's ability to interpret the salient features of the context. Another refers to the attitudes, emotions, motivations, attributions, and social relations associated with the language user's perception of the context (e.g., Mathewson, 1994; McKenna, 1994). A third factor concerns the familiarity of the context and its goals. The goal-setting component activates initial processing routines (perhaps used in the past for similar tasks), which in turn activate the scripts and begin a processing cycle in verbal working memory.

An unresolved issue is the relationship between the language user's goal setting and language processing. Relatively little research exists on this issue in language learning, even though it is important for models of language use and for language testing (cf. van Dijk & Kintsch, 1983; Dweck, 1989; Renninger et al., 1992). At issue is how the individual sets goals and plans to reach a goal once it is set. Some evidence exists that variation in goal setting and awareness of purpose influences reading, although this line of research is not extensive (Dweck, 1989; Mathewson, 1994; McKenna, 1994; Myers & Paris, 1978; Paris, Lipson & Wixson, 1983; Renninger et al., 1992; Rothkopf, 1982). Research on writing and the composing process also sheds a little light on this relationship. A person can indirectly observe planning behavior, and a person can induce planning behavior. Induced planning behavior does seem to have some influence on the outcome of the language task (Bereiter & Scardamalia, 1987; Flower & Hayes, 1980, 1981; Hayes et al., 1987). An individual's attitudes will affect the motivation for carrying out a task, as well as the care and efficiency of task performance (Crookes & Schmidt, 1991).

Given this view, it is likely that the internal goal-setting mechanism in communicative language use would need to account not only for the influences from the academic context and the efficiency of translation from intention to language processing (e.g., specific task training, language knowledge), but also intentions to perform a task as affected by attitudes, anxieties, motivations, emotions, expected outcomes/purposes, involvement, prior experiences on similar tasks, interpersonal relations, and similar task-specific planning routines (awareness of how to proceed) (Gardner & MacIntyre, 1992; Hidi, 1990; MacIntyre & Gardner 1991; Ortony, Clore, & Collins, 1988; Reed & Schallert, 1993; Sadoski, Goetz, & Fritz, 1993). It may be useful to explore further this set of issues related to the goal-setting component.

3.2.2. Verbal Working Memory

Verbal working memory is defined as those aspects of world knowledge, language competence, and verbal processing used to accomplish a particular goal. The purpose of this model is not to specify how working memory as a whole might operate, but only to suggest how a specific language-based task would

11

be carried out within working memory. Working memory, in this model, is represented by the entire internal-processing unit. This follows the assumption that working memory is situated within long-term memory, involving processing mechanisms, metacognitive processes (available throughout verbal working memory space), and resources activated from long-term memory networks.

The term "verbal working memory" was chosen on the basis of arguments given by Barsalou (1992), who argues that the traditional autonomous multistore model of short-term memory encounters a number of problems in explaining language-processing results (Barsalou, 1992: pp. 92-115). Language processing (as a limited-capacity activity) is more likely to be constrained in activating information and procedures by the limitations of the central processor operating within long-term memory than by the limits of a separate processing component called short-term memory (cf. Cowan, 1993; Kintsch, 1993; Shiffrin, 1993). Alternatively, the preference of working memory lies with its primary emphasis on activation rather than on retrieval and storage, its preference for coordinating both storage and computation, and its preference for allowing parallel processes rather than a purely serial processing (Anderson, 1990; Harrington & Sawyer 1992; Kintsch, 1993; Just & Carpenter, 1992). Just and Carpenter (1992) explained the requirements of working memory as follows:

> A somewhat more modern view of working memory takes into account not just the storage of items for later retrieval, but also the storage of partial results in complex sequential computations, such as language comprehension. The storage requirements at the lexical level during comprehension are intuitively obvious.... But storage demands also occur at several other levels of processing. The comprehender must also store the theme of the text, the representation of the situation to which it refers, the major propositions from preceding sentences, and a running, multilevel representation of the sentence that is currently being read (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). Thus, language comprehension is an excellent example of a task that demands extensive storage of partial and final products in the service of complex information processing. Most recent conceptions of working memory extend its function beyond storage to encompass the actual computations themselves.... These processes, in combination with the storage resources, constitute working memory for language.... We present a computational theory in which both storage and processing are fueled by the same commodity: activation. In this framework, capacity can be expressed as the maximum amount of activation available in working memory to support either of the two functions.

> In our theory, each representational element has an associated activation level. An element can represent a word, phrase, proposition, grammatical structure, thematic structure, object in the external world, and so on. The use of the activation level construct here is similar to its widespread use in other cognitive models, both symbolic (e.g., Anderson, 1983) and connectionist (e.g., McClelland & Rumelhart, 1986). During comprehension, information becomes activated by being encoded from written or spoken text, generated by a computation, or retrieved from long-term memory. As long as an element's activation level is above some minimum threshold value, that element is considered part of working memory; it is available to be operated on by various processes (Just & Carpenter, 1992:121-122).

3.2.3. Verbal-Processing Component

The verbal-processing component includes metacognitive processing, on-line processing, language knowledge, and world knowledge. Other aspects, such as the text model or the mental model in working memory, are left unspecified here, although they might be thought of in terms of the verbal-processing output. This section will address issues relating to metacognitive processing in the Model.

Metacognitive processing includes strategic processes that are directed by goal setting, problem solving, and multiple (and sometimes conflicting) informational sources. For example, the need to adjust speech production to conform to a superior's new expectations requires a balancing of elements from sociolinguistic and discourse competence, along with various outcome scenarios that are potentially available from world knowledge. This situation will require the individual to direct strategic attention to the speech output to carefully monitor the context and make adjustments. Metacognitive processing will also include the strategies associated with strategic competence in Canale and Swain's (1980) framework (i.e., processes that enhance the message and repair perceived miscommunication). Metacognitive processing is typically seen as requiring extensive demands on working memory capacity. The more complex the task (or the more unfamiliar the topic, the more difficult the vocabulary, the more unusual the setting, the more anxiety-provoking the context), the more demands are placed on metacognitive processing in working memory.

Further issues associated with metacognitive strategies include the debatable value of distinctions such as cognitive strategies versus metacognitive strategies and strategies versus skills. These distinctions may not be useful to maintain in any strict sense and are not assumed by the Model. Moreover, following Baker (1991) and (Paris, Wasik & Turner, 1991), the distinction between cognitive strategies and metacognitive strategies is argued to be variable by topic, task, and individual. For example, the need to read five pages of Chomsky's latest article will impose severe demands on an individual's processing; many processes that might otherwise be on-line (such as proposition integration with new vocabulary) will require directed attention and problem-solving routines as part of comprehension. An inverse example is that of summarization. This ability is typically nominated as a metacognitive process, yet an individual's regular updating of the plot to a mystery novel does not require the directed attentional processing that Chomsky's article would. Thus, what might be a metacognitive strategy in one situation will only invoke minimal on-line processing demands (a procedural routine) in another situation. It is therefore difficult to specify a universal set of skills versus strategies or a set of cognitive versus metacognitive strategies.

On-line processing refers to the basic skilled processing that (for native speakers) does not require extensive attentional resources, such as word recognition, initial parsing, and nondemanding processing related to propositional formation and integration into a text model. It also reflects those aspects of mental model processing that are not "directed" for any particular purpose or goal. Thus, on-line processing represents not only potentially encapsulated activities, but also those activities not placing serious demands on metacognitive processing or attentional resources. Generally, this view of on-line processing conforms with the sketch of working memory noted by Just and Carpenter in Subsection 3.2.2.

In much the same way that a task may not overwhelm resources for a native speaker's on-line processing, the advanced learner of a second language may be sufficiently skilled and have efficient

processing routines so on-line processing works well. For less skilled second-language learners, such as those taking the TOEFL 2000 test, the on-line processing may not be sufficiently skilled to process the information without great resource demands (which are also dependent on topic, tasks, internal goal-setting, etc.). For these individuals, aspects of on-line processing will not be much different from the resource-intensive strategic processing typical in metacognitive processing. Thus, one major source of L2 test-taking variation may well be the limits of on-line processing in working memory (again a prediction of Just and Carpenter's Capacity Theory).

3.2.4. Language Competence

Language competence in the COE model refers to the language user's grammatical, discourse, and sociolinguistic knowledge. It is important to note that this component simply defines the types of language knowledge that might be required in a given context. What is done with that knowledge (e.g., whether it is used for interpreting linguistic input or for producing output) is defined in the verbal-processing component (explained in Subsections 3.2.3 and 3.2.4). Each of the language subcomponents is defined here, but these are only general definitions. The specific elements of language knowledge required (activated) in a given context depend on the features of the context described in Subsection 3.1 and the internal goal setting (Subsection 3.2.1).

Grammatical competence includes phonological/orthographic, morphological, lexical, structural, and semantic knowledge. It includes knowledge of possible structures, word orders, and words. The specific grammatical knowledge required in a given context depends on the grammatical features that the language user must comprehend and produce to accomplish the goals he or she sets.

Many issues remain concerning how best to represent grammatical knowledge, but the most difficult aspect of what is defined here as grammatical knowledge is the nature of the lexicon. The model includes lexical knowledge as a part of grammatical knowledge, even though the lexicon most likely contains more than formal linguistic features. The problem is that the lexicon's relation to any other language-processing component is not simple or straightforward. While everyone can agree that a lexicon is necessary, it is not entirely clear where it should be located, what it should encompass, and how it should interact with other processing components. For example, it is not clear to what extent the lexicon is linked with knowledge of the world — to what extent is knowledge of the world simply knowledge of the terms and concepts primarily stored in the lexicon itself (cf. Paivio, 1986)? From this, many other questions arise. To what extent is procedural knowledge linked to the lexicon (perhaps as generic script entries)? To what extent are schemas and knowledge frames represented in the lexicon as some set of generic defaults for declarative knowledge concepts? To what extent does the lexicon obviate the need for an independent syntactic processing component? To what extent are sociolinguistic knowledge and discourse structural knowledge keyed to terms and concepts of the lexicon? To what extent are intentions, purposes, and plans keyed to lexical concepts and terms? To what extent is the L2 lexicon distinct from the L1 lexicon? All of these questions point to the undefined nature of the lexicon in relation to other processing components and the need for additional work in this area.

Discourse competence refers to the language user's knowledge of how language is sequenced and how it is organized above the syntactic level. This component includes knowledge of exchange sequences in

14

interaction, genre and register markers, coherence markers and coherence relations, topic development, links between informational units, and the structuring of informational flow. Discourse competence also includes knowledge of genre structure to account for the fact that people recognize whole genre forms in many instances. As with respect to grammatical competence, the specific discourse knowledge needs will depend on the features of the context (particularly the features of texts defined in Subsection 3.1.1.4.).

Sociolinguistic competence includes knowledge of language functions and language variation. Functions include, for example, knowledge of language for greeting, convincing, apologizing, criticizing, and complaining. In any given setting, the language user will need to know some combination of functions to participate. This component of language knowledge is activated by goal setting (Subsection 3.2.1) since the functions follow directly from goals, attitudes, and purposes. Functional knowledge, in turn, activates the specific linguistic knowledge needed to produce or interpret the relevant functions. For example, a student who disturbs a small class by entering late must perceive the situation as one that requires an apology so his or her "goal-setting" component can set the goal of apologizing. To actually apologize, however, the students functional knowledge must know how to make an apology in English, (which specific words and syntactic patterns to use, as well as how much of an excuse to provide and how much detail to include).

Knowledge of language variation consists of knowledge of dialect diversity (e.g., regional differences such as midwestern versus southern), of naturalness (e.g., archaic forms and vocabulary versus contemporary colloquial speech), of cultural references (e.g., "to meet one's Waterloo"), and of figures of speech (e.g., "to have been around the block"), as well as knowledge of numerous configurations of register variation. Register variation is defined as knowledge of the language appropriate for the following contextual situations: (1) one or many in the intended audience; (2) familiar or distant relationship among participants; (3) informal or formal occasions; (4) subordinate or superordinate relation to participant(s); (5) general or topical content; and (6) relative background knowledge of participants.

Each dimension of register variation defines an element of context that influences language use; therefore, knowledge of the language associated with the combinations of dimensions is an important component of language competence. For example, the student entering the class late would choose different language to express apology in a small class than in a large class. The student's language would be different in a class comprised entirely of friends than in a class of strangers. It would be different if the instructor were there than if the instructor were absent. Our understanding that language varies across these dimensions of register is the result of empirical research in sociolinguistics, but the nature of native speakers' linguistic variations across the contexts of interest remains an important research area.

The Model states that the types of knowledge defined in these three major subcomponents of language competence — grammatical, discourse, and sociolinguistic — are in combination, the major components of language necessary for communicative language use in context. The answers to questions about how each of these general areas of language competence can be specified for the academic contexts of interest to TOEFL 2000 test users await further research. Also of interest would be developmental definitions of each of these areas of language knowledge. For example, is it generally the case that learners know how to use greetings before they learn to complain? A third issue involves the question of the level of sociolinguistic knowledge a learner must obtain to be able to work effectively in academic contexts.

3.2.5. World Knowledge

The language user's world knowledge refers to the store of information that the individual has from past learning and experience in life. The Model indicates that world knowledge works together with language competence to comprehend and produce language in context. The Model suggests a relationship between the two components that is similar to that proposed by Bereiter and Scardamalia (1987) for the knowledge-transforming model of composing. In their model, problem-solving situations will activate new information. The new information will then create rhetorical/linguistic choices and problems as the writer must decide how best to use the new information. The solution to the rhetorical problems may then require additional world knowledge information. Thus, the informational needs cycle back and forth as more information is needed to satisfy the task requirements (cf. the role of the lexicon). Tasks that are simple and follow an established routine will, of course, require less of this interactive communication between the two informational components. Both will simply activate the relevant information typically needed for the routine task, and on-line processing will execute the routine.

There are many unresolved issues concerning world knowledge. Little is established aside from the general agreement that such a component is needed, that it is likely to have many default concepts, and that it is organized in network like pathways activated under various conditions. The often-cited concept of schema theory is not without controversy, and some researchers have suggested that schemas are only temporary constructs assembled from memory exemplars rather than stable generic frames stored in memory (Kintsch, 1988; Rayner & Pollatsek, 1989; Singer, 1990; cf. Barsalou, 1992). Others have argued that schemas stored as propositional networks do not account for the range of an individual's world knowledge (Paivio, 1986; Sadoski, Paivio, & Goetz, 1991). The extent to which world knowledge is verbal and the extent to which it is nonverbal is also an open question. For the part that is verbally encoded, it is not clear how independent the world knowledge component is from the lexicon. In spite of these issues, the world knowledge component is presented as a separate interacting component of the Model, a component that contributes information to working memory and that is important for a language user's interpretation and construction of verbal messages.

3.2.6. Internal-Processing Output

During verbal processing, the language user constructs a representation of "comprehension so far," which the Model calls internal-processing output (i.e., the output from the verbal working memory). This output is likely to include copies of the "text model" and "mental model" (see Subsection 3.3.1). In the COE Model, this internal processing output can then refer to both the "text model" and "mental model" of the reader and listener and the representation of "what I've produced so far" of the speaker/writer.

For any goal-directed language activity, there must be a mechanism for monitoring the output and assessing its similarity to the internal goal setting. During verbal processing, the language user will monitor the internal-processing output and compare it with the internal goal setting. At this time, metacognitive processing may invoke changes in the processing, require the activation of additional or different information, or invoke specific processing strategies that will enhance the potential output or address mismatches between the output and the goal setting. As the processing cycle produces/recognizes additional information, the output continues to be matched to the goal setting. As processing nears

16

completion, the monitoring either rejects the output and tries again through another processing cycle or is satisfied with the match to internal goal setting and ends the iterative cycle for that particular task or subtask. The monitoring could also respond to a nonmatch with frustration and end the processing cycle, even though the output does not match the goal setting.

The comparison of goals with output, or "monitoring," is not discussed extensively in cognitive psychology and psycholinguistics but is important in applied linguistics (cf. Buck, 1991; Krashen, 1985; Morrison & Low, 1983; Pawley & Syder, 1983; Schmidt, 1992), and composition (Bereiter & Scardamalia, 1987; Hayes et al., 1987). Monitoring is an essential process in language use and in specific task performance. In the sense discussed here, the "monitor" is basic to language processing. This description should not be confused with Krashen's (See, for example, Krashen, 1985) use of the term in his discussion of his input theory.

3.3. Model of Communicative Language Applied

Having defined the components of the COE Model and how they work together, we can consider how this perspective on language use relates to the more familiar "skills" perspective and how it can be used to describe specific instances of language use in an academic context. As the COE Model evolved, one of the most revealing heuristics participants used to understand the Model was to see the extent to which the familiar "four skills" could be translated into and described by the Model's constructs. The result of these "Model testing" sessions was a set of working lists (see Appendix C) that distorted the Model's definitions with a skill-based orientation. For example, constructing the skills list required us to ask the following types of questions: In what settings, and for what tasks, are listening skills required in academic contexts? What grammatical, discourse, and sociolinguistic knowledge is required in all those (listening) settings identified by the first question?

The Model, in contrast, directs us to ask questions such as the following: What are the academic settings about which we want to infer our test takers language ability? What language abilities (e.g., grammatical, discourse, sociolinguistic knowledge) are required to succeed in each of those settings? The answer to the first question will not be divided into skill areas (as the examples in Subsection 3.3.2 show); it will be divided by settings or tasks. The answer to the second question (for each setting) usually will include the abilities associated with more than one skill. Thus, use of the skills as an organizing principle is inconsistent with the Model and appears too limiting to be the guiding metaphor for TOEFL 2000 in its early stage. Nevertheless, the vast majority of research, as well as much of the knowledge in our field, is organized around the four skills. Therefore, it is useful to consider how the skills fit within the framework of the COE Model and how skills come into play when we use the model to define the language ability required in a specific context.

3.3.1. The Skills Described Through the Model

With respect to reading and listening skills, the Model includes components responsible for text processing. Text processing in the Model can be described in a manner consistent with a number of recent psycholinguistic and cognitive psychology approaches (Anderson, 1990; Barsalou, 1992; Just & Carpenter, 1987, 1992; Oakhill & Garnham, 1988; Singer, 1990). For written language comprehension, the

17

individual visually perceives the text and engages in word recognition. For aural language, lexical activation is keyed by auditory perception and, perhaps, other cues in the context. Written language production typically is initiated through goal setting and initial activation of plans. Spoken language production will, at times, be initiated by goal setting and planning, although spoken language will often also often be initiated by relatively automatic response patterns.

Beginning with a discussion of reading comprehension, the skills and processes the Model assumes will be outlined. As a reader begins to read, and the first word of a sentence is activated for working memory, the semantic and syntactic information from the word is used to begin parsing the incoming sentence. The additional incoming words are accessed and combined in terms of general parsing principles, relying on semantic and syntactic information attached to each word and, at some point, pragmatic and contextual information. The word and the growing syntactic structure are also interpreted as a propositional structure, representing the meaning of the sentence. The proposition is integrated as the reader reaches the end of the sentence. This proposition is then "sent" to be incorporated in a text model (within working memory), which synthesizes the incoming proposition with an existing (or created) propositional network.

At the same time that the new propositional structure is being integrated into the text model, the words from the next sentence are being activated and assembled in a new parsing representation. Meanwhile, the proposition being integrated into the text model probably will require one or more bridging inferences to assist the coherent and thematic restructuring of the text model (van Dijk & Kintsch, 1983; Kintsch, 1993; Perfetti, 1993; Singer, 1993; cf. Graesser & Kreuz, 1993). Thus, by the end of the proposition construction and integration, inferential processes are being used to fit the proposition into the text model. The text model (as a summary of the information in the text) and necessary bridging inferences will be constrained to represent consistently nominated information more strongly, as well as information which has been marked in one of several ways as thematic.

While the text model is being constructed and reconstructed, an interpretive model of the text is also being constructed. This mental model, or situation model (Barsalou, 1992; van Dijk & Kintsch, 1983; Fincher-Kiefer, 1993; Kintsch, 1988; Singer, 1990) represents the reader's interpretation of the text (beyond the comprehension of the text model). As an interpretation of the text information, it will include additional processing information, such as explanations for the information, evaluations of the information, connections to other sources of information, emotional responses, adequacy of information assessments, and appropriate purpose assessment (Graesser & Kreuz, 1993). This interpretation may not be a complete or accurate representation of the text, but it is the reader's individual interpretation. While this latter stage of comprehension and interpretation is going on, the other processes of word recognition, parsing, semantic interpretation, and text-model building continue.

Speaking and writing must be explained somewhat differently because they follow partially developed plans that reflect an internal text model; in the COE Model, this process would be associated with the goal-setting component. The initial procedures for lexical activation are different for speaking because they are internally driven, instead of driven by an external language source. Speaking is also likely to involve a different set of demands on the processing output in terms of a heightened monitoring, if nothing else.

Writing processes combine the internally driven activation of information with the more reflective interpretive analysis used to create a more elaborate mental model. While both speaking and reading usually require relatively fast on-line processing in terms of lexical access and parsing, writing almost always demands more reflective operation of both. That is, writing, as it improves, requires the penetration of automatic production with concerns for appropriate word and structure choices, as well as concerns for organization, reader expectations, writing purposes, emotional signaling, and attitude to task and topic (Bereiter & Scardamalia, 1987).

It is possible to see the Bereiter and Scardamalia models of knowledge telling and knowledge transforming in writing processes as types of inverses of the text model and mental model created by reading processes. While the text model for the reader and knowledge telling for the writer represent relatively automatic processing and thus operate similarly, the more reflective and problem-solving counterparts move in different directions, at least in goal setting and output matching. The processing difficulty for the writer is that the text produced, to the extent that it is slanted to a writer's particular purpose or reflects attitudes, emotions, or evaluations, must be tempered because of its potentially public nature. In contrast, readers, in constructing their mental models, can be much more aggressive in their interpretations, evaluations, and emotions toward a text and not have to be judged. Other important differences exist between reading and writing, as between writing and speaking, but a discussion of these would take us beyond the scope of the present report.

Recognizing the differing demands that each language skill makes for language processing, the goal of the TOEFL 2000 model, particularly the internal operations, is to take into account these differences and construct a simple model that can be shaped to language processing in any of the four skill areas. When we look at academic contexts, we see that skills work together to accomplish goals. Two brief examples in the next subsection illustrate this integration.

3.3.2. Using the Model for Describing Language Use

In the previous discussion of skills, there was no mention of the contexts in which the skills were used or any specific knowledge required to perform them. Skills can be defined abstractly in terms of processes without reference to context. The COE Model, in contrast, specifies that language ability must be defined in view of a particular type of context. In the two examples that follow, the COE Model allows us to specify the critical components of communicative language use (both external and internal) in an academic context.

The first example is a task that has its setting specified as a classroom. The text type is news feature reporting from *Newsweek*. The task is to read a high-interest news article to report on it to the class. The context specification could be extended to include aspects of the context using the features outlined in Subsection 3.1. The internal goal setting will involve the individual's interpretation of what would be a high-interest article; this might include consideration of topic and text length. The individual will also consider previous experiences with similar assignments, consider what the teacher thinks is appropriate/good performance, and choose an article that appears to be understandable and interesting. A planning routine, based on prior experiences and academic training, will be activated for working memory.

As the student begins to read, he or she will be trying to understand in general what the article is about and to jot down some specifics that may be important to include in the class report. As the student reads, lexical access processing will activate linguistic knowledge and world knowledge. Both of these sets of knowledge will already have been activated to some extent during preliminary goal setting. The linguistic knowledge activated here will include the specific morphological, lexical, structural, and semantic knowledge. Discourse knowledge will include knowledge of how *Newsweek* articles typically are organized and knowledge of how local coherence relations are established. Sociolinguistic knowledge will include the knowledge of the dialect and cultural referents in the article and knowledge of the text type of the article in terms of the dimensions of register variation. As the on-line and metacognitive processing continues, new world knowledge and linguistic knowledge will be cycled into the processing component. Thus, the three subcomponents are seen as operating in tandem, without presupposing a strictly linear set of operations.

At certain points, the individual will want to monitor the progress of the processing and the match between the potential output and the internal goal setting (completion of the article). When the individual gets tired of processing, or when the results develop toward completion, the internal-processing output will match the results to the goal setting. If the match is satisfactory, the results will be sent to the production output or (as in this example) will be stored for later retrieval. The individual may also stop the processing, because the internal goal setting is not a strong model for comparison or because the individual is incapable of making a good match and stops trying.

The second example also occurs in a classroom setting. The text type is a question/response sequence. The task is to respond to the teacher's question on the basis of lecture notes from the previous week. The internal goal setting is based on previous experiences with this routine, academic training, and the likelihood of retrieving the relevant information. The specific goal will be either to respond appropriately and to be recognized as knowing past information, or to respond in a sufficiently vague manner so as to satisfy the teacher and not betray total ignorance. The goal setting begins to activate whatever world and linguistic knowledge can be recalled from the past week's lectures. This information is to be combined with whatever world and linguistic knowledge can be usefully inferred from the teacher's question.

To formulate a response, the activated world knowledge and linguistic knowledge are processed. The language knowledge will include consideration of the functional purpose of the response and appropriate register information. Because the response time will be relatively rapid, a fair amount of the response processing will depend on set routines for assembling an answer. Delays due to metacognitive strategy feedback may lead the responder to produce filler sounds and phrases until some linguistically well-formed response is sent to the internal-processing output. This then may or may not be matched to the goal setting component. The response is then generated as performance, which others in the context can observe.

3.4. Conclusion

TOEFL 2000 discussions have so far focused on the Model itself: what components are essential for describing language use, how to express them effectively, and how to translate our current understanding of skills to applied linguists' perspectives on communication. The ultimate success of these efforts will depend on their usefulness in test development and validation. Anticipating the need to examine the Model in light of its intended uses, the following two sections speculate on some of the model's implications for test development and validation.

4. Implications for Test Development

The COE Model suggests that test development should begin by examining the nature of the academic context in which the language ability of interest is used. It should be underscored again here that the COE Model is not itself a definition of the specific language abilities used in academic contexts; instead, it provides a way of hypothesizing what those abilities may be for a specific context of interest. This, in turn, allows us to use the Model to develop some test items/tasks, as outlined in Subsection 4.1. This process raises issues associated with use of the Model for test development (explained in Subsection 4.2).

4.1. Using the COE Model for Test Development

Most testing specialists say that the first step in writing a test is to define what is to be measured, but this is easier said than done. The nature of a construct definition is not clear-cut. Should it consist of lists of pieces of knowledge? Should the construct be understood in terms of performance criteria? Should it include a description of the mental processes we want to test? Where do we find the relevant information? In linguistics books? In English as a Second Language (ESL), and English as a Foreign Language (EFL) textbooks? The COE Model suggests both the types of information that should go into a construct definition (for language proficiency) and a starting point for the test developer to compose such a definition. In the example that follows, a scenario is provided for testing the language abilities required for performance in a science lab. The example works through the four-step process that the model implies.

4.1.1. Identify and Analyze the Academic Context of Interest

The academic context of interest is the science lab, so the first step is to find some university science labs to visit. Leaving aside important sampling issues of deciding which labs count as "science," how to choose the ones to visit, and how long to stay, the researchers might begin by identifying several science labs at a university, getting permission, and visiting them to document the nature of language use in these contexts. The following is one brief example of the type of data the researchers might collect: an excerpt from a conversation between a Teacher's Assistant (T) and a student (S) who is attempting to complete a required lab report (which apparently consists of questions on an experiment about light refraction that the student has set up in front of him).

| 1 | S: | Can you explain why it's doing this? The ray entering the |
| 2 | | block refracts away from the normal to the surface. |

| 3 | T: | What was your question? |

| 4 | S: | I s- we see why it does that but why — I mean we see that it |
| 5 | | does that, but why does it do that? |

| 6 | T: | Why — why does it bend toward the normal? |

| 7 | S: | Yeah. |

8 T: Uh, that's because of the difference of the index of
9   refraction...

10 S: OK

11 T: ...between air and...

12 S: Is that the same thing for why this does that away from it?

13 T: Yes, that's the same reason but, depending on from where to
14   where the beam goes, there's a change in angle...

15 S: OK

16 T: ...direction of the angle

17   (writes the response down as the answer to the question on the
18   lab report)

19 S: All right.

20 T: So to bend toward the normal to the surface, does the index
21   of refraction should be greater than one, or smaller than
22   one?

(pause)

23 S: What'd you say?

24 T: To bend to the normal to the surface, um is the index of
25   refraction of this plastic...should...

26 S: Is greater, isn't it?

27 T: Greater than one? Should it greater than one?

28 S: Is that right or not? (clatter from object dropped on neighboring table) Huh?

29 T: Yes, that's right?

30 S: OK

(Searls, 1991, pp. 45-46)

The researchers return from their fieldwork with plenty of classroom data. Examining all the data the researchers have collected, the test developers frequently see the type of conversation illustrated here. The discussion consists of questions and answers centered around a lab report that must be completed. For the purpose of analysis, completion of the lab report would be the "task." When the test developers and researchers look through a lot of conversations in labs centered around completing a lab report, they see that sometimes a Teacher's Assistant (TA) is present (as in this example). Sometimes the dialogue is between lab partners, and the difference in participants changes the nature of the discourse. A lab partner, the test developers see in other data, never uses the long, thought-provoking questions that the TA attempts in this example (lines 20-25). Much of the language in the data refers to the lab equipment present in the setting — a physics lab, which the researchers have documented in great detail. The text is oral, consisting of short Q-A turns between two participants of unequal status and knowledge. The researcher and test developer could do a more thorough discourse analysis of the text, citing the expressions and syntax used and the functions and sociolinguistic characteristics of the exchanges. This is the type of data that would be used for hypothesizing the language abilities required for performance in science labs.

4.1.2. Hypothesize the Abilities Required in the Context

On the basis of this fieldwork, the test developer can work on the necessary construct definition by hypothesizing the goal setting, language competence, and verbal processing required of a student to participate in this situation. Test developers would want to look at a large amount of data, extracting typical settings, participants, tasks, and texts to produce the construct definition. But until that field work and analysis have been done, for the purpose of this example, we will hypothesize the language abilities required of the student in the situation given here. The Model implies that the following be included in the definition:

Goal Setting: The student needs to understand the instructor as he explains the goal of setting up the experiment and completing the lab report. Understanding requires language abilities to read and listen to instructions. The student must also set a goal of producing written answers to questions in the lab report.

Language Processing: Metacognitive processes keep the student working toward the large goal, constructing subgoals as needed during the course of the conversation. On-line processes retrieve specific aspects of language knowledge as needed to accomplish goals, which include both production (asking questions and writing responses in the lab report) and comprehension (listening to the TA and reading the lab questions).

Language Competence: Linguistic competence includes knowledge of the following: syntax of questions (both yes/no and why), statements in the present tense, a limited range of morphological forms associated with present, concrete language, and phrases used in response to questions. Vocabulary knowledge includes knowledge of simple and concrete words (e.g., explain, enter, block, difference, thing, reason), as well as technical words directly associated with the materials and the basic concepts of light refraction (e.g., ray, index, refraction).

23

Discourse Competence: This includes knowledge of question/answer turn taking (including interruption sequences), associated cohesive devices, and knowledge of language referring to the physical setting. Discourse competence also includes knowledge of the genre, "lab report."

Sociolinguistic Competence: This includes knowledge of the language associated with a variety of spoken language functions used in questioning. The student has to know standard English (for reading the lab materials) and a nonnative variety of English to communicate with the international TA. To engage in the conversation with the TA, he has to know the language used for one-to-one conversation on a specific topic with a specialist on the topic with whom he has a distant and subordinate relationship, in a somewhat formal setting. Also relevant is the fact that the TA is "primary knower" (Berry, 1987), meaning that he knows the answers to the student's questions. The language used to complete the lab report would require different sociolinguistic knowledge.

4.1.3. Construct Relevant Item/Task Formats

Having defined what should be measured, the test developer can construct a test task (or tasks) to measure it. For example, using the fieldwork from the science lab to look at the tasks used in context, the test developer can get ideas for test tasks, even though the resulting test will not be exactly like having the learner perform in the context. Again, one would want to consult a large set of science lab data to look for potential item/task formats, but for the purpose of the example, the item/task format "filling in a lab report" can be used. From the abilities analyzed on the basis of the transcript, the test developer would recognize the need to test the student's ability to use language to understand a goal. A good item/task would require the test taker to activate metacognitive processes to update goals and use on-line processing efficiently to retrieve the necessary language knowledge for both production and comprehension. Moreover, the student should have to call on the linguistic knowledge identified: questions and answers with simple present, concrete forms, knowledge of turn taking, and language referring to objects in the physical context. On the basis of this analysis, a simple lab report format can be designed, one that visuals of an experiment would support. However, how can this format be extended to test the oral interaction ability so important to this task?

Test item/tasks may have difficulty assessing all of the abilities included in a construct definition. At this point, one has to make compromises — while noting where and what kind of compromises these are. If it is not possible to produce a format requiring the test taker to complete her lab report through oral interaction with a TA, the test developer might consider other interactive formats through which the test taker could gain information (e.g., a computer program that would allow the student to query a database). The reason for hypothesizing abilities required in the academic context of interest (as illustrated in Subsection 4.1.2) is that potential constraints and compromises of the test can be recognized and understood. These constraints and compromises will influence the abilities we can actually test in relation to the abilities we want to measure and are therefore a relevant object of validity inquiry.

4.1.4. Establish a Scoring Rubric

In concert with planning item/task formats, the test developer must decide how to evaluate performance on the task. An obvious method of scoring a "lab report" item/task is to assess the amount and/or quality

of the responses on the actual report. If the report has seven blanks that must be completed with short answers, it might most simply be scored by accepting only the completely correct answers. However, one purpose of the explicit construct definition should be to create a more meaningful scoring method. For example, if the test developer had test takers gain information to include in the lab report by querying a database, the ability to engage in this process should be assessed, in addition to the final product. The scoring rubric should distinguish the student who asked a lot of good questions (but never the right one) from the student who was unable to form a question at all. Both students would end up with a "product" score of "0," but their "process" scores would be different. The current COE Model provides little guidance on this important issue. Moreover, the fieldwork the Model suggests would only reveal the nature of the academic context, describing the setting, participants, task, text, and topic. These aspects of the context will affect what it means to be successful, but the Model does not provide guidelines on how to analyze success (or levels of success) in a given situation. This issue obviously requires more work to attempt a principled method of evaluating performance (cf. Subsection 4.2.4). The data in the example of the student in the lab did include the language of a student who was successfully filling in his lab report. To do it he was extracting information from the TA, despite the TA's desire to make him think about what he was doing. To obtain that information successfully, he had to use the interactive oral/aural ability defined previously. For the student in this setting, the analysis of the abilities required to successfully complete the lab report should help in evaluating the abilities important in this context. This type of construct definition provides some guidance about what might be scored; however, much work remains to be done on the question of what should be scored and how to define levels of abilities in academic contexts. This is only one of the test development issues the Model raises.

4.2. Issues Raised by the COE Model for Test Development

Because one of the primary purposes of the Model is to suggest directions for test development, future discussions will have to focus on how it might be improved for this purpose. The COE has not yet discussed these issues in depth, but it may be useful to identify some starting points for discussion. This section presents four of the many issues that have been mentioned and that are apparent from the above example: (1) authenticity; (2) the four skills; (3) the definition of situation; and (4) correctness.

4.2.1. Why Not Just Give Them "Authentic" Academic Tasks?

The process of test development outlined in Subsection 4.1 might be — and indeed has been by some test developers — boiled down to the principle. "Just make tests that look like authentic language tasks." The "Make them look authentic" principle has been useful in some contexts (e.g., perhaps for some teacher-made classroom tests). The COE, however, in concert with many language-testing researchers, has rejected it as the guiding principle for test development. As Spolsky (1985) pointed out, a language test is not authentic as something else: "Any language test is by its very nature inauthentic, abnormal language behavior, for the test taker is being asked not to answer a question giving information but to display knowledge or skill" (p. 39). Moreover, "authenticity" is an attribute that cannot be defined adequately from the perspective of the test developer; it is instead a quality affected by the participants' perception (Lantolf & Frawley, 1988; Bachman & Palmer, forthcoming).

Because test situations are inherently different from the contexts about which we want to infer test takers' ability, students' performance on a test is likely to offer a distorted picture of the ability they would use in "authentic" contexts. The issue is, then, how we can use the picture of ability obtained by test performance to make inferences about abilities in other contexts. In order for tests to be used appropriately, it is the responsibility of test developers to demonstrate, and test users to consider, evidence concerning the meaning of test scores (i.e., construct validity evidence). To investigate construct validity, it is necessary to hypothesize the construct that the test is intended to measure. Developing a test whose validity can be justified is the primary objective of TOEFL 2000; therefore, the COE has devoted its time to articulating a model to be used for construct definition, which is necessary for construct validation. We will elaborate on this point in Subsection 5.1.

4.2.2. What About the "Four Skills"?

Reflecting applied linguists' perspectives, the COE Model uses the situation (e.g., a segment in a science lab) as the organizing unit for hypothesizing language abilities. Test developers and many test users, on the other hand, are accustomed to thinking of skills (i.e., reading, writing, listening, and speaking) as the organizing principle for understanding learners' abilities. In other words, test developers are accustomed to hypothesizing the abilities that define reading comprehension. How can we use the collective "skills" knowledge and experience of language testers to inform test development from a situation perspective?

The collective "skills work" provides useful explanations of some aspects of language ability used in situations. In attempting to define reading comprehension, for example, skill theorists provide us with well-conceived lists of the types of processing (both on-line processing and metacognitive processing) that can be required in a given situation. Where the skills approach falls short, in the view of applied linguists, is its failure to account for the impact of the context of language use.

Situation theorists emphasize that reading, along with other language "skills," takes place for a purpose (determined by other aspects of the context). In a homework assignment, for example, instructions are given orally (requiring listening comprehension), reading is done at home (reading comprehension), and answers to questions at the end of the chapter are written (writing ability). The material is referred to in the class lecture (listening comprehension), where the student must take notes (writing ability) to use to study for an exam. What the "situation" perspective offers is an understanding of the complete context in which the reading skills take place. This context is essential for hypothesizing the abilities required. For example, the fact that the situation is reading for answering questions in writing and preparing for tomorrow's lecture affects the reader's goals and therefore his or her metacognitive processing. The fact that the reading is a history textbook affects the specific language competence required. In short, the situation-based approach that the COE Model suggests acknowledges the complexities and interrelatedness of features of language and contexts in communicative language proficiency.

4.2.3. What Is a "Situation"?

The theoretical Model attributes an important position to the situation (Subsection 3.1.1). Elements of the situation (setting, participants, task, text) are hypothesized to affect the specific language abilities

required for performance in that situation. In theoretical terms, this is a useful conceptualization for defining language ability. When it comes to empirical research (the fieldwork mentioned in Subsection 4.1) and test development, however, it is apparent that the "situation" will require an operational definition. To take the example of the science lab, when does the situation begin and end? The goal of the task was identified as completing the lab report, but was that the only relevant goal in that lab situation? Would the TA agree on the goal identified for that session? The science lab situation consisted of more than one text: the oral text of the dialogue between TA and student and the written language of the lab report. How many texts can a situation contain? How many tasks? Should we attempt to define situations by the numbers of other elements they contain?

These questions point to the need for theoretically informed empirical work whose objective is to construct an operational definition of "situation." Such research would investigate the nature of academic situations and their associated language. It would be similar to, and informed by, research that investigates reading or writing in academic contexts. It would differ from such research by not starting with predefined categories of "reading" and "writing," but would instead begin without preconceived units into which data would be placed. It would use multiple perspectives, including those of instructors and students, to help obtain a realistic account of the data. In these senses, such research would draw on ethnographic methods. The objectives, however, would have to include constructing a definition of situation that test development staff could use for additional research on academic situations and for guidance in selecting task/item formats.

4.2.4. What Is Correct?

As indicated in Subsection 4.1.4, the COE Model provided little guidance on establishing a scoring rubric. Moreover, the Model, with its context-based definition of language ability, raises difficult issues for judging correctness. The Model suggests that language users establish the norms for acceptable and appropriate language behavior and, therefore, decisions about what is correct depend on language users within specific contexts. The meanings of language are socially constructed; meanings do not exist except by consensus of the community of language users. A certain linguistic behavior is not "wrong" or "inappropriate" unless a group of individuals has determined that it is. (See here Hymes', 1972 features "norms for interpretation and norms for interaction.") Language academies (such as the Academie Francaise) or writing guides and usage manuals are extreme institutionalized examples of bodies that prescribe correctness. Teenagers represent a social group that has its own norms for speaking, illustrated by such forms as "awesome," "like" or "you know," and "so, I go" (for "so, I said"). These are not forms considered acceptable by all groups of speakers. Nor would teenagers consider them appropriate in all settings and situations. It would depend on the context.

The idea of a context-specific correctness may be problematic for constructing scoring criteria that accurately reflect perspectives on success in completing academic tasks. Discussion is needed to identify relevant research and to weigh options for appropriate response evaluation. Research and discussion of this issue should help to refine the definition of the sociolinguistic competence (Subsection 3.2.5) required of students to function successfully in an academic context.

4.3. Conclusion

The four steps in test development (Subsection 4.1) and the four issues raised (Subsection 4.2) are intended to provide a starting point for discussions about implications of the Model for test development. Much remains to be said and researched as the practical concerns of test development are viewed from the perspective of a model reflecting theory in applied linguistics. This discussion and research will, in turn, help in modifying, specifying, and better understanding subsequent versions of the Model. As the test development perspective will highlight some aspects of the Model, validity research will also make requirements of and provide contributions to the Model's evolution. In the next section, we initiate discussion of the Model's role in validity research.
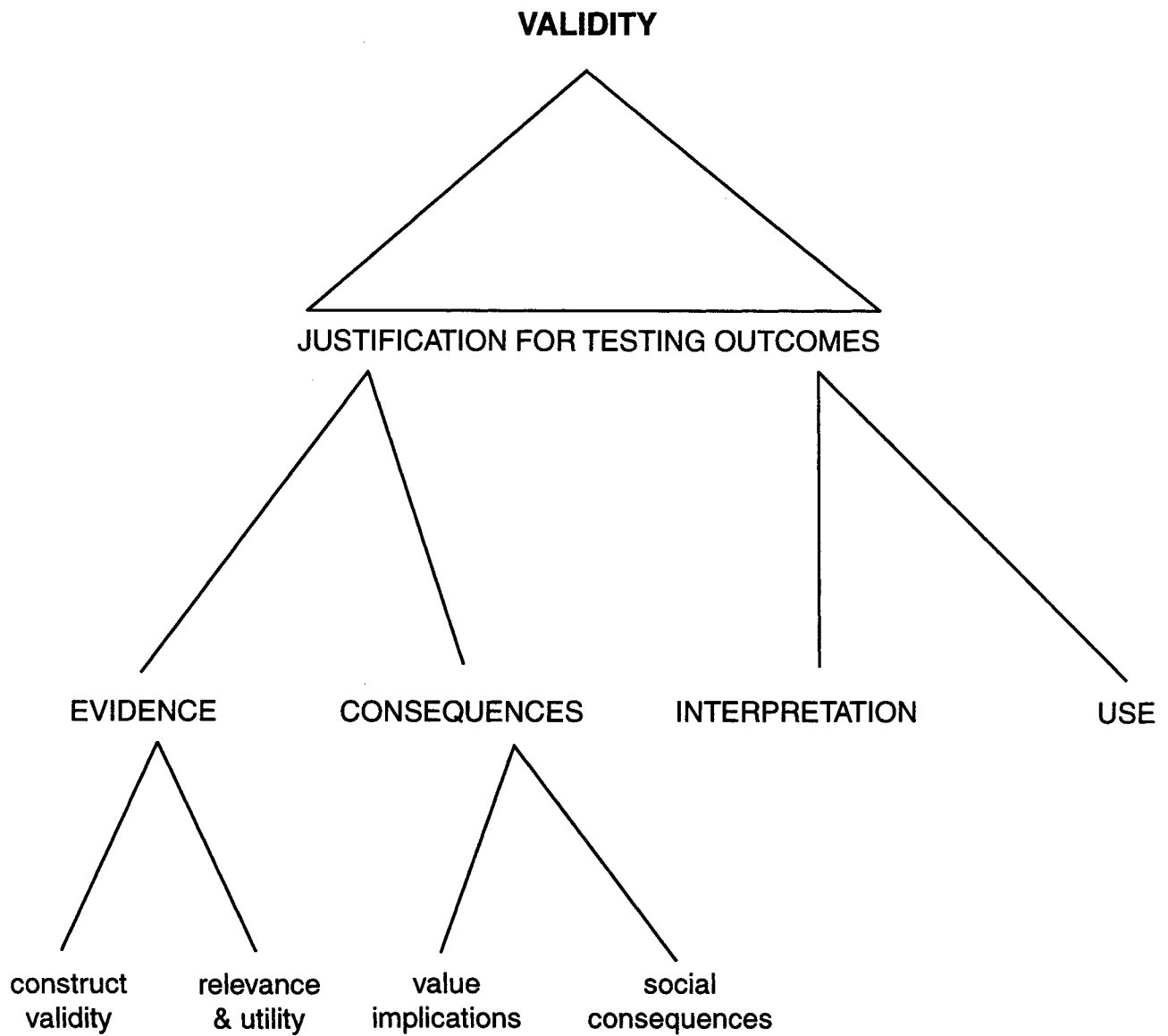
5. Implications for Validation

To understand the role of the COE Model in TOEFL 2000 validation, it is necessary to define "validation." As mentioned in Subsection 2.3, the COE conception of validity encompasses a broad range of concerns. Discussions of the Model have been motivated by, for example, content validity, construct validity, and the social consequences of TOEFL 2000 use. These broad validity concerns are consistent with those of other applied linguists who have argued for multiple approaches to construct validation (e.g., Grotjahn, 1986; Alderson, 1993; Bachman, 1990; Cohen, forthcoming; Anderson, Bachman, Perkins, & Cohen, 1991) in addition to responsibility for the social consequences of testing (Canale, 1988; Alderson, 1993). Applied linguists (e.g., Bachman, 1990; Chapelle & Douglas, 1993) are beginning to discuss these validity perspectives systematically within a framework defined by Messick in the third edition of *Educational Measurement*. Messick's validity definition acts as the emerging frame of reference for validity in language testing (as well as in other areas of educational measurement). Therefore, it is used here to systematize the validity issues raised by the COE and to begin to explore the relationship of the COE Model to TOEFL 2000 validity research.

According to Messick, validity refers to "the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores" (p. 13). This definition is comprised of two major parts (see Figure 2). "Empirical evidence and theoretical rationales" refer to justifications for testing. Justifications for TOEFL 2000 will include evidence that test takers' performance reflects communicative language proficiency in an academic context and evidence that TOEFL 2000 is useful for making decisions about language readiness for academic study in North American universities. In other words, "evidence" refers to the data and arguments we can gather concerning construct validity and pertaining to relevance and utility. A second source of justification will include the positive consequences we can cite for TOEFL 2000's interpretations and uses. Arguments concerning consequences are drawn from construct validity evidence as it pertains to test interpretation and use: value implications associated with test interpretations and the social consequences of test use. The second part of Messick's definition, "interpretations and actions" refers to the outcomes of testing. Outcomes for TOEFL 2000 include the inferences made on the basis of test performance about learners' levels of communicative language proficiency in an academic context, the decisions about admissions to North American universities, and the impact of TOEFL 2000 on the structure of curricula in ESL/EFL programs.

FIGURE 2

Definition of Validity in Testing Based on Messick (1989)

(from Chapelle, 1994)

VALIDITY

JUSTIFICATION FOR TESTING OUTCOMES

EVIDENCE   CONSEQUENCES   INTERPRETATION   USE

construct   relevance   value   social
validity   & utility   implications   consequences

30

Messick's definition offers a coherent perspective for conceptualizing validity. To move from perspective to research, however, it is necessary to identify specific types of evidence and consequences that will allow us to investigate TOEFL 2000 interpretation and use. In addition, we need to suggest research methods for identifying the relevant evidence and consequences. Some researchers have begun to explore a range of criteria and approaches for validity inquiry (e.g., Linn, Baker, & Dunbar, 1991). The TOEFL 2000 research program will eventually be in a good position to contribute to these explorations. We begin by considering Messick's suggestions for validation research. This report does not attempt to cover these issues comprehensively, but only to outline some of the specific types of justifications that TOEFL 2000 researchers may investigate and to speculate how the COE Model might apply to each.

5.1. Construct Validity Evidence

Earlier views of construct validity treated it as one of three types of validity alongside content-related validity and criterion-related validity. In Messick's definition, however, content- and criterion-related considerations are subordinate to construct validity; each is considered among the admissible types of evidence for construct validity. This view emphasizes that the "varieties of evidence [e.g., criterion-related evidence] are not alternatives but rather supplements to one another" (p. 16) and that construct validation is a process through which evidence pertaining to the meaning of test scores is accrued from a variety of sources. Because construct evidence assists in interpreting score meaning (i.e., what TOEFL 2000 really measures), it is viewed as the cornerstone for all other validity inquiry.

Investigation of construct validity for TOEFL 2000 requires a definition of what TOEFL 2000 should measure; as explained previously, it is intended to measure communicative language proficiency in an academic context. The COE Model is intended to support construct validity inquiry because it is a general framework describing communicative language use, which, for a given academic context, can be used to create a specific construct definition. But what is construct validity research? Messick (1989) names the following approaches:

> We can look at the content of the test in relation to the content of the domain of reference [i.e., content evidence]. We can probe the ways in which individuals respond to the items or tasks [i.e., empirical item and task analysis]. We can examine relationships among responses to the tasks, items, or parts of the test, that is, the internal structure of test responses. We can survey relationships of test scores with other measures and background variables, that is, the test's external structure. We can investigate differences in these test processes and structures over time, across groups and settings, and in response to experimental interventions — such as instructional or therapeutic treatment and manipulation of content, task requirements, or motivational conditions (p. 16).

The relevance to TOEFL 2000 of each of these five sources of construct validity evidence will be addressed in turn.

5.1.1. Content Evidence

Content evidence consists of experts' judgements concerning the relevance and representativeness of the abilities measured by TOEFL 2000 to the language abilities required in academic contexts. "Judgements of the relevance of test items or tasks to the intended score interpretation should take into account all aspects of the testing procedure that significantly affect test performance...what is judged to be relevant...is not the surface content of test items or tasks but the knowledge, skills, or other pertinent attributes measured by the items or tasks" (Messick, 1989, p. 38-39). Because content evidence requires analysis of the "knowledge, skills and other relevant attributes" measured by the items/tasks on TOEFL 2000, a principled method is needed for making these judgements.

The COE Model provides a framework for making such judgements by defining the types of knowledge (linguistic competence and world knowledge), processes (verbal-processing component), and other relevant attributes (goal setting) that TOEFL 2000 items/tasks may measure. For any given item/task, then, experts can judge what specific aspects of communicative language proficiency are measured. The analysis of a situation in the science lab (Subsection 4.1.2.) illustrates such judgements. To use the model for content analysis of test items/tasks, the "situation" (Subsection 3.1.1) becomes "taking the test," the "task" becomes whatever the test items/tasks require, the "text" becomes the language that the student must interpret and produce, the "setting" becomes the location where the test is being taken, and the "topic" refers to the propositional content of the tasks. The COE Model allows the same framework to be used for discussing language use in academic settings and in a test setting. This consistency of perspective between the context of interest to test score users and the context of the test should be useful for producing and interpreting content evidence for validity.

5.1.2. Empirical Item and Task Analysis

Empirical item and task analysis requires examination of statistical item difficulty (and other item statistics), as well as observation and analysis of learners' performance on the test and qualitative response analysis. Quantitative research compares item difficulties with theoretically derived difficulty predictions (e.g., Abraham & Chapelle, 1992; Perkins & Linnville, 1987). Qualitative research investigates the problem-solving processes of learners as they take a test (e.g., Buck, 1991; Cohen, 1984; Grotjahn, 1987; Feldmann & Stemmer, 1987). The fundamental question in both types of research is: To what extent do the empirical (item-difficulty or processing evidence) results support the theoretical predictions that are based on the definition of what the test is supposed to measure? In order to interpret this research, it is necessary to have defined what the test is supposed to measure.

With respect to empirical item-difficulty research, the COE Model provides a framework in which the necessary definition(s) can be developed. However, to construct a definition in terms that will be useful for such research, it would be necessary to specify further stages of development within the relevant components (i.e., language competence, verbal-processing component, internal goal setting). To imagine how this would work for some aspects of language competence, we might look at research on second-language acquisition (Larsen-Freeman & Long, 1991) to hypothesize stages of syntactic development. It is less clear how developmental stages might be hypothesized for other components (such as goal setting), but one might speculate. The problem of using second-language acquisition research to define language ability

32

developmentally is an important unresolved issue in language testing (e.g., Canale, 1988; Brindley, 1991; Pienemann, Johnson, & Brindley, 1988).

With respect to understanding learners' problem-solving processes, the COE Model — with an explicit "goal-setting" and "verbal-processing component" — can be used to guide research. The question will be: To what extent do the strategies/processes required for TOEFL 2000 fall within the definition of communicative language proficiency in academic contexts? This question becomes important when we define language proficiency as consisting of strategies/processes, as the COE Model does. The COE Model suggests that the language strategies required for success in academic contexts are an important part of the proficiency to be measured. Moreover, the Model suggests that the strategies required for performance depend on the situation. Because the test situation is not the same as the academic situation, investigation of the strategies required in both becomes an important validity question.

5.1.3. Internal Structure of Tests

Validity evidence concerning the internal structure of tests refers to demonstrating empirically that test items adhere to the same structural relations as those hypothesized for the construct they represent. In other words, each item/task should be theorized to measure some aspect(s) of an overall construct, and empirical data should support the presumed item structure. For example, research investigating internal structure of language tests has used factor analysis (Bachman, 1982) and item response theory (Hudson, 1993). Fundamental to this type of research is explicit specification of elements that comprise a construct. Ideally, the study of internal item structure might ultimately provide empirical justification for theoretically motivated procedures for combining item/task scores. But what kind of item/task structures are implied by the COE Model?

The COE Model defines proficiency as consisting of components; however, the components are hypothesized to work together, making predictions about isolated performance effects from individual components difficult. Within the "language competence" component, distinct pieces are also defined; again, however, the distinctions are made to allow us to define them. Components of language knowledge are hypothesized to work together in communicative language performance. The COE Model suggests that hypotheses about item/task similarity should begin by identifying similar "situations." The similarities in situations should require similarities in the abilities measured. For example, if we have items/tasks intended to measure ability to use oral language in a lab setting, ability to listen and take effective lecture notes, and ability to write lab reports (on the basis of given data), our analysis of each of these settings would have defined the abilities required in each. On the basis of overlap in abilities, we might hypothesize an item structure of items/tasks within each domain. The prediction of structural relations from this type of model is not as straightforward as it is when there are lists of separate pieces of knowledge or skills; therefore, the question of how to conduct this type of research will require theoretically guided deliberation.

5.1.4. External Structure of Tests (Correlational Evidence)

Correlational evidence is used to investigate whether theorized convergent and discriminate relationships among tests can be observed empirically. Messick (1989) distinguishes two types of external structure questions that correlational research can address. One is trait evidence for validity, "the extent to

which a measure relates more highly to different methods for assessing the same construct than it does to measures of different constructs assessed by the same method" (p. 46). Results from this kind of study in language testing have provided validity data for particular tests by identifying the influence of test methods (Stevenson, 1981; Bachman & Palmer, 1982). To design such a study, the researcher must begin by defining the trait that is to be measured by multiple methods. How does the COE Model pertain to the design of this research?

The COE Model provides a means for developing a definition of language ability in a given setting, but can we call "language ability in a given setting" a trait? In the strict sense of the term, and in the sense used by multi-trait, multi-method researchers, no. In the strict sense, a trait is defined independently from the method used to measure it; in the multi-trait, multi-method research design, trait effects on performance are good, and method effects on performance are bad. The COE Model, in contrast, defines "language ability in a given setting." In other words, setting (i.e., method) effects are not bad; they are expected. The COE Model reflects an interactionalist (rather than trait) perspective (Messick, 1989, p. 15; Zuroff, 1986) of construct definition, attributing performance to three sources: (1) the context; (2) the capacities of the individual, and (3) an interaction between the two. Trait-oriented correlational research on TOEFL 2000 should be cautious to take into account the interactionalist orientation of the COE Model.

The second type of correlational research investigates nomothetic span, "the network of relationships of a test to other measures" (Embretson, 1983, p. 180). This type of study requires the researcher to hypothesize strengths of correlations (based on distances in the nomothetic span of constructs) with other measures. Fundamental to this type of research is a construct definition of what is measured by the test under investigation and the hypothesized strengths of relationships expected between that construct and the others in the study. The COE Model, which includes multiple components, would make this type of research interesting to consider. Theoretically guided discussion of this type of research (i.e., asking which constructs we expect the abilities we define for TOEFL to relate to and to what degree) may help inform evolution of the Model.

5.1.5. Experimental Manipulations

Experimental manipulations of subjects or test methods enable the researcher to examine hypotheses about test performance. This is done by systematically modifying test conditions to verify that observed test performance behaves in concert with theory-based predictions. The COE Model should provide a useful framework for the design and interpretation of such research because it illustrates observable output (data) existing in hypothesized relation to unobservable learner capacities, which are affected by features of the situation. To use this basic framework for constructing experiments, test method facets (Bachman, 1990) can be modified to affect performance data obtained in predictable ways. The COE Model should help researchers to choose research questions and make such predictions. For example, the researcher might hypothesize that learners with differing levels of sociolinguistic competence might respond differently to item formats that systematically vary contextual features pertaining to the aspects of register variation. The following is an example of the type of empirical question that might be investigated: How does performance vary across ability levels when the format calls for formal versus informal language?

34

Not reflected in the Model, but a subject of discussion by the COE, is the need for experimental research to focus on different subject populations who would take TOEFL 2000. In particular, native speakers should perform very well, and there should be no significant difference between the performance of graduate and undergraduate students if the use of the same test for both groups is to be justified. Planning and conducting a variety of experimental research will help to improve both the COE Model and TOEFL 2000.

5.2. The Consequences of Testing

Although the Model is most directly relevant to how construct validity evidence is conceived and interpreted, the COE's validity concerns extend beyond these construct-related issues. We therefore briefly outline other important modes of validity inquiry pertaining to test use and the consequences of testing and speculate how the Model relates to them.

5.2.1. Evidence Concerning Relevance and Utility

In the COE's discussions of how to define communicative language ability, the primary concern has been for TOEFL 2000 to provide the information necessary for university admissions people at North American universities to make accurate decisions about language proficiency of candidates. As a part of the TOEFL 2000 research program, then, it seems essential to obtain systematic feedback from these test users as well as from others affected by the admissions decisions (e.g., university professors who work with the admitted students). The construction and implementation of such a feedback system should be a part of TOEFL 2000 validation research, and the information it obtains should be considered for test revision decisions. The need for such research in the North American context is clear; other issues concerning evaluation of TOEFL 2000 relevance and utility are still open to debate.

Questions about "unintended" uses of TOEFL 2000 have not been resolved. Therefore, it is difficult to define the scope of validity inquiry focusing on relevance and utility. It might be argued that test developers should not be obligated to investigate the utility of unintended test uses. In contrast, some have argued that testers' responsibilities extend beyond constructing and administering tests: "Once one has been involved in gathering information, one becomes responsible in some way to see that it is used [appropriately]" (Canale, 1988, p. 75). For TOEFL 2000, we can predict that "unintended uses" (e.g., admissions to English-medium universities outside North America; placement decisions in intensive English programs in the United States), will be made, and therefore the question of "relevance and utility for what" must be discussed as research in this area is planned.

5.2.2. Value Implications

As we noted in Subsection 2.2, the COE Model attempts to represent a perspective of language consistent with current views in applied linguistics. The COE believes that it is important to use such a model as the basis for construct definition because the test's foundation ultimately will be reflected in its form and content. These, in turn, will portray applied linguists' beliefs about language. Spolsky's (1978) summary of "traditions" in language testing illustrates the way in which testing practices can be seen in light of their linguistic and measurement foundations — foundations rooted in contemporary academic

beliefs. Nominations for a "tradition" consistent with applied linguists' perspectives include "naturalistic-ethical," emphasizing test methods requiring natural language use in context and the test developer's social responsibility (Canale, 1988) and "communicative-psychometric," emphasizing the need for linguists and psychometricians to work together to develop and evaluate communicative language tests (Bachman, 1990).

Beyond a test's value-laden foundations, however, we can investigate the values that a test expresses to those it affects. A test can be thought of as a social event that conveys messages about applied linguists' views of language to test takers, instructors, applied linguists, academics in other disciplines, and other members of societies (Canale, 1987). One such message that the COE has discussed is the value TOEFL 2000 might communicate about the privileged status of particular varieties of English. This concern revisits and adds a validity-relevant dimension to the difficult issue of the relative nature of correctness, discussed in Subsection 4.2.4.

Language that is correct in one context may be considered nonstandard in North America. Varieties of English exist in numerous English-speaking contexts, such as India and Singapore, and speakers of these other Englishes will be among those taking the TOEFL test. Moreover, the future language-use situations of the test takers and the purposes of attending North American universities are also relevant concerns. Many students will spend their university lives and future professional lives interacting with the international community of scholars and professionals, not all of whom are speakers of American English. The issue underlying these situations is whether Englishes that do not conform to American or Canadian English standards can be defined as acceptable. Can the English norms that the TOEFL test assumes represent an international English rather than exclusively a North American variety, thereby not penalizing educated speakers of English whose competence is not a perfect match to educated American varieties? TOEFL 2000's answer to this question will display values of test developers toward different varieties of English.

5.2.3. Social Consequences

Applied linguists' most pressing validity concern is the consequence TOEFL 2000 will have on English teaching and testing throughout the world. This issue has become a focal point for our profession, members of which refer to the consequences of testing using the following terms: washback, backwash, impact, face validity, and systemic validity. Whatever we call it, it refers to the fact that TOEFL 2000 test design will shape English language curricula, methods, materials, and tests throughout the would. Messick suggests that for one line of validity inquiry "we can trace the social consequences of interpreting and using the test scores in particular ways, scrutinizing not only the intended outcomes but also unintended side effects (p. 16)." Although methods for tracing social consequences are not clear-cut (Alderson & Wall, 1993), research documenting TOEFL 2000's effects should be an essential facet of validity inquiry.

5.3. Conclusion

As we begin to look at the implications of applied linguists' views of communicative language proficiency for validity justifications, we can foresee some of the dilemmas Moss (1992) points out with respect to performance assessment across disciplines:

Performance assessments present a number of validity problems not easily handled with traditional approaches and criteria for validity research. These assessments typically present students substantial latitude in interpreting, responding to, and perhaps designing tasks; they result in fewer independent responses, each of which is complex, reflecting integration of multiple skills and knowledge; and they require expert judgement for evaluation. Consequently, meeting criteria related to such validity issues as reliability, generalizability, and comparability of assessments — at least as they are typically defined and operationalized — becomes problematic (Moss, 1992, p. 230).

As we suggested previously, the COE expects such problems to be encountered as validity issues for TOEFL 2000 continue to be explored. However, we also believe that TOEFL 2000 provides a unique and valuable opportunity in applied measurement for realizing ideals for validity inquiry, as well as for pioneering efforts to establish alternative and practical validity criteria for the profession.

6. Evaluation and Evolution of the COE Model

Any psychologist, psycholinguist, educational measurement researcher, or applied linguist will have questions and suggestions for improving the COE Model. The needs of test development clearly point to areas that require additional work, some of which we point out in Subsection 4.2. In addition, planning and conducting theoretically motivated construct validity research will provide input for the Model, as pointed out in Subsection 5.1. As the Model continues to evolve, it will be important to keep in mind its purposes:

- Informing test development (see Section 4)
- Supporting content analyses of item/task formats (see Subsection 5.1.1)
- Guiding and interpreting empirical validity research (see Subsections 5.1.2 through 5.1.5)
- Informing inquiry pertaining to utility, values, and consequences of testing (Subsection 5.2)

References

Abraham, R., & Chapelle, C. (1992). The meaning of cloze test scores: An item difficulty perspective. *Modern Language Journal, 76(4)*, 468-479.

Alderson, J. C. (1993). The relationship between grammar and reading in an EAP test battery. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research*, (pp. 203-219). Arlington, VA: TESOL Publications.

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

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *Cognitive psychology and its implications, 3rd ed.* New York: W. H. Freeman and Co.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*, 35-44.

Anderson, N. J., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing, 8(1)*, 41-66.

Bachman, L. F. (1982). The trait structure of cloze test scores. *TESOL Quarterly, 16(1)*, 61-70.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative competence. *TESOL Quarterly, 16*, 449-465.

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

Baker, L. (1991). Metacognition, reading, and science education. In C. Santa & D. Alvermann (Eds.), *Science learning: Processes and applications* (pp. 2-13). Newark, DE: International Reading Association.

Barsalou, L. (1992). *Cognitive psychology.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Bereiter, C. (1990). Aspects of an educational learning theory. *Review of Educational Research, 60*, 603-624.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Berns, M. (1990). *Contexts of competence: Social and cultural considerations in communicative language teaching.* New York: Plenum Press.

39

Berry, M. (1987). Is teacher an unanalyzed concept? In M. A. K. Halliday & R. P. Fawcett (Eds.), *New developments in systemic linguistics. Volume 1: Theory and description* (pp. 41-63). London: Pinter Publishers.

Brindley, G. (1991). Defining language ability: the criteria for criteria. In Anivan, S. (Ed.), *Current developments in language testing (pp. 139-164)*. Singapore: SEAMEO Regional Language Center.

Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing, 8(1)*, 67-91.

Campbell, R., & Wales, R. (1970). The study of language acquisition. In J. Lyons (Ed.), *New horizons in linguistics* (pp. 242-260). London: Penguin.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 2-27). London: Longman.

Canale, M. (1987). Language assessment: The method is the message. In D. Tannen & J. E. Alatis (Eds.), *The interdependence of theory, data, and application* (pp. 249-262). Washington, DC: Georgetown University Press.

Canale, M. (1988). The measurement of communicative competence. *Annual Review of Applied Linguistics, 8*, 67-84.

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

Carroll, J. B. (1989). Intellectual abilities and aptitudes. In A. Lesgold & R. Glaser (Eds.), *Foundations for a psychology of education* (pp. 137-197). Hillsdale, NJ: Lawrence Erlbaum.

Carver, R. (1992). Effect of prediction activities, prior knowledge, and text type on amount comprehended: Using Rauding theory to critique schema theory research. *Reading Research Quarterly, 27*, 164-174.

Chapelle, C. A. (1994). Is a C-test valid for L2 vocabulary research? *Second Language Research, 10(2)*, 157-187.

Chapelle, C., & Douglas, D. (1993). Foundations and directions for a new decade of language testing research. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 1-22). Alexandria, VA: TESOL Publications.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cohen, A. (1984). On taking language tests: What the students report. *Language Testing, 1(1)*, 70-81.

Cohen, A. (1993). The role of instructions in testing summarizing ability. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 132-160). Alexandria, VA: TESOL Publications.

Cohen, A. (forthcoming). Strategies and processes in test taking and SLA. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research.*

Cowan, N. (1993). Activation, attention, and short-term memory. *Memory and Cognition, 21,* 162-167.

Crookes, G., & Schmidt, R. (1991). Motivation: Reopening the research agenda. *Language Learning, 41,* 469-512.

Crystal, D. (1987). *The Cambridge encyclopedia of language.* New York: Cambridge University Press.

Douglas, D. (forthcoming). Testing methods in context-based second language acquisition research. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research.*

Duff, P. (1993). Tasks and interlanguage performance: An SLA perspective. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 57-95). Philadelphia: Multilingual Matters.

Dweck, C. (1989). Motivation. In A. Lesgold and R. Glaser (Eds.), *Foundations for a psychology of education* (pp. 87-136). Hillsdale, NJ: Lawrence Erlbaum Associates.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93(1),* 179-197.

Færch, C., & Kasper, G. (Eds.). (1983). *Strategies in interlanguage communication.* London: Longman.

Feldmann, U. & Stemmer, B. (1987). Thin__ aloud a__ retrosprctive da__ in C-te__taking: Diffe__ languages--diff__ learners--sa__ approaches? In C. Færch & G. Kasper (Eds.), *Introspection in Second Language Research,* (pp. 251-267). Philadelphia, PA: Multilingual Matters.

Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes, 16,* 99-124.

Firth, J. R. (1957). *Papers in linguistics 1934-1951.* London: Oxford University Press.

Flower, L., & Hayes, J. (1980). The dynamics of composing: Making plans and juggling constraints. In L. Gregg and E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31-50). Hillsdale, NJ: Lawrence Erlbaum Associates.

41

Flower, L., & Hayes, J. (1981). The pregnant pause: An inquiry into the nature of planning. *Research in the teaching of English, 15,* 229-244.

Gardner, R. C., & MacIntyre, P. (1992). On the measurement of affective variables in second language learning. *Language Learning, 43,* 157-194.

Graesser, A., & Kreuz, R. (1993). A theory of inference generation during text comprehension. *Discourse Processes, 16,* 145-160.

Grotjahn, R. (1986). Test validation and cognitive psychology: Some methodological considerations. *Language Testing, 3(2),* 159-185.

Grotjahn, R. (1987). On the methodological basis of introspective methods. In C. Færch & G. Kasper (Eds.), *Introspection in Second Language Research,* (pp. 54-81). Clevedon, Avon: Multilingual Matters.

Guthrie, J. (1988). Locating information in documents: Examination of a cognitive model. *Reading Research Quarterly, 23,* 178-199.

Guthrie, J., Britten, T., & Barker, K. G. (1991). Roles of document structure, cognitive strategy, and awareness in searching for information. *Reading Research Quarterly, 26,* 300-324.

Habermas, J. (1970). Toward a theory of communicative competence. *Inquiry, 13,* 360-375.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning.* London: Edward Arnold.

Halliday, M. A. K., & Ruquiya, H. (1989). *Language, context and text: Aspects of language in a social semiotic perspective.* Oxford: Oxford University Press.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14,* 25-38.

Hayes, J., Flower, L., Schriver, K. A., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics, Volume 2,* (pp. 176-240). New York: Cambridge University Press.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research, 60,* 549-571.

Hudson, T. (1993). Testing the specificity of ESP reading skills. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 58-82). Alexandria VA: TESOL Publications.

Hymes, D. (1971). Competence and performance in linguistic theory. In R. Huxley & E. Ingram (Eds.), *Language acquisition: Models and methods*. London: Academic Press.

Hymes, D. (1972). *Towards communicative competence*. Philadelphia: Pennsylvania University Press.

Kramsch, C. (1993). *Context and culture in language teaching*. Oxford: Oxford University Press.

Joos, M. (1962). *The five clocks*. Bloomington, Indiana: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological review, 99*, 122-149.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston, MA: Allyn & Bacon.

Kachru, B. B. (Ed.). (1992). *The other tongue: English across cultures. Second revised edition*. Urbana, IL: University of Illinois Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review, 95*, 163-182.

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes, 16*, 193-202.

Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and reproduction. *Psychological Review, 85*, 363-394.

Krashen, S. (1985). *The input hypothesis*. New York: Longman.

Lantolf, J. P., & Frawley, W. (1988). Proficiency: Understanding the construct. *Studies in Second Language Acquisition, 10*, 181-195.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. London: Longman.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, Nov.*, 15-21.

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

Lowenberg, P. H. (1992). Testing English as a world language: Issues in assessing nonnative proficiency. In B. B. Kachru (Ed.), *The Other tongue: English across cultures,* (2nd ed., pp. 108-121). Urbana, IL: University of Illinois Press.

MacIntyre, P., & Gardner, R. C. (1991). Language anxiety: Its relationship to other anxieties and to processing in native and second languages. *Language Learning, 41,* 513-534.

Malinowski, B. (1923). The problem of meaning in primitive languages. In C. Ogden & I. A. Richards (Eds.), *The meaning of meaning* (pp. 296-336). London: Trubner and Co.

Mathewson, G. (1994). Model of attitude influence upon reading and learning to read. In R. Ruddell, M. Ruddell, and H. Singer (Eds.), *Theoretical models and processes of reading. 4th ed.* (pp. 1131-1161). Newark, DE: International Reading Association.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1.* Cambridge, MA: MIT Press.

McKenna, M. (1994). Toward a model of reading attitude acquisition. In E. Cramer and M. Castle (Eds.), *Fostering the love of reading* (pp. 18-39). Newark, DE: International Reading Association.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement (3rd ed.)* (pp. 13-103). NY: Macmillan Publishing Co.

Morrison, D. M., & Low, G. (1983). Monitoring and the second language learner. In J. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 228-250). New York: Longman.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62(3),* 229-258.

Munby, John. (1978). *Communicative syllabus design.* Cambridge: Cambridge University Press.

Myers, M., & Paris, S. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology, 70,* 680-690.

Nelson, C. L. (1985). My language, your culture: Whose communicative competence? *World Englishes, 4(2),* 243-50.

Oakhill, J., & Garnham, A. (1988). *Becoming a skilled reader.* New York: Basil Blackwell.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotion.* New York: Cambridge University Press.

Paivio, A. (1986). *Mental representations.* New York: Oxford University Press.

Paris, S., Lipson, M., & Wixson, K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8,* 293-336.

Paris, S., Wasik, B., & Turner, J. (1991). The development of strategic readers. In R. Barr, et al. (Eds.), *Handbook of reading research, Volume 2* (pp. 609-640). New York: Longman.

Pawley, A. (1992). Formulaic speech. In B. Bright, et al. (Eds.), *Oxford international encyclopedia of linguistics, Volume 2* (pp. 22-25). New York: Oxford University Press.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191-226). New York: Longman.

Perfetti, C. (1989). There are generalized abilities and one of them is reading. In L. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glazer* (pp. 307-334). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. (1991). Representations and awareness in the acquisition of reading competence. In L. Rieben and C. Perfetti (Eds.) *Learning to read: Basic research and its implications* (pp. 33-44). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. (1993). Why inferences might be restricted. *Discourse Processes, 16,* 181-192.

Perfetti, C., & McCutchen, D. (1987). Schooled language competence: Linguistic abilities in reading and writing. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics, Volume 2* (pp. 105-141). New York: Cambridge University Press.

Perkins, K., & Linnville, S. (1987). A construct definition study of a standardized ESL vocabulary test. *Language Testing, 4(2),* 125-141.

Pienemann, M., Johnson, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition, 10,* 217-243.

Rayner, K., Garrod, S., & Perfetti, C. (1992). Discourse influences during parsing are delayed. *Cognition, 45,* 109-139.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Reed, J., & Schallert, D. (1993). The nature of involvement in academic discourse tasks. *Journal of Educational Psychology, 85,* 253-266.

Renninger, K., Hidi, S., & Krapp, A. (Eds.). (1992). *The role of interest in learning and development.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Sadoski, M., Goetz, E., & Fritz, J. (1993). Impact of concreteness on comprehensibility, interest, and memory for text: Implications for dual coding theory and text design. *Journal of Educational Psychology, 85,* 291-304.

Sadoski, M., Paivio, A., & Goetz, E. (1991). A critique of schema theory in reading and a dual coding alternative. *Reading Research Quarterly, 26,* 463-484.

Savignon, S. (1983). *Communicative competence: Theory and classroom practice.* Reading MA: Addison-Wesley.

Saville-Troike, Muriel. (1989). *The ethnography of communication: An introduction.* (2nd ed.). New York: Basil Blackwell.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 219-258.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14,* 357-385.

Schmidt, R. (1993). Awareness and second language acquisition. In W. Grabe, et al. (Eds.), *Annual review of Applied Linguistics, 13. Issues in second language teaching and learning* (pp. 206-226). New York: Cambridge University Press.

Searls, J. (1991). *The role of questions in the physics laboratory classes of two nonnative speaking teaching assistants.* Unpublished master's thesis, Iowa State University, Department of English. Ames: Iowa State University.

Shallice, T. (1988). *From neuropsychology to mental structure.* New York: Cambridge University Press.

Shohamy, E. (1993). A collaborative/diagnostic feedback model for testing foreign languages. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research.* Alexandria, VA: TESOL Publications.

Shiffrin, R. (1993). Short-term memory: A brief commentary. *Memory and Cognition, 21,* 193-197.

Singer, M. (1990). *Psychology of language.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Singer, M. (1993). Global inferences of text situations. *Discourse Processes, 16,* 161-168.

Skehan, P. (1992). Second language acquisition strategies and task-based learning. *Thames Valley University Working Papers in English Language Teaching, 1,* 178-208.

Smith, L. E., & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World Englishes, 4(3),* 333-342.

Spolsky, B. (1978). Introduction: Linguists and language testers. In B. Spolsky (Ed.), *Advances in language testing research: Approaches to language testing 2* (pp. v-x). Washington, DC: Center for Applied Linguistics.

Spolsky, B. (1985). The limits of authenticity in language testing. *Language Testing, 2(1)*, 31-40.

Stanovich, K. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review, 10*, 72-100.

Stanovich, K. (1991). Changing models of reading and reading acquisition. In L. Rieben & C. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 19-31). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stanovich, K. (1992). The psychology of reading: Evolutionary and revolutionary developments. In W. Grabe, et al. (Eds.), *Annual review of Applied Linguistics, 12. Literacy* (pp. 3-30). New York: Cambridge University Press.

Stevenson, D. K. (1981). Beyond faith and face validity: The multitrait-multimethod matrix and the convergent and discriminant validity of oral proficiency tests. In A. S. Palmer, P. J. M. Groot, & G. A. Trosper (Eds.), *The construct validation of tests of communicative competence* (pp. 37-61). Washington, DC: TESOL Publications

Stillings, N. A., Feinstein, M. H., Garfield, J. L., Rissland, E. L., Rosenbaum, D. A., Weisler, S. E., & Baker-Ward, L. (1987). *Cognitive science: An introduction.* Cambridge, MA: MIT Press.

van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Walczyk, J., & Royer, J. (1993). Does activation of a script facilitate lexical access? *Contemporary Educational Psychology, 17*, 301-311.

Weaver, C., & Kintsch, W. (1991). Expository text. In R. Barr, et al. (Eds.), *Handbook of reading research, Volume 2* (pp. 230-245). New York: Longman.

Whitney, P., Ritchie, B., & Clark, M. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes, 14*, 133-146.

Zuroff, D. C. (1986). Was Gordon Allport a trait theorist? *Journal of Personality and Social Psychology, 51(5)*, 993-1000.

Appendix A

Chronological Development of the TOEFL 2000 Model
at COE Meetings

1. April 1990, San Juan, Puerto Rico

At the spring 1990 Committee of Examiners (COE) meeting, the topic of planning for TOEFL 2000 was raised. The first serious discussion of this topic with the COE was a part of the agenda.

2. January 1991, San Diego

At the first TOEFL 2000 meeting, we began the process with ETS test-development staff presenting position papers on the current status and future directions for each section of the TOEFL test. At this meeting, we recognized the need for a statement of purpose for the TOEFL 2000 test and a definition of what the test would intend to measure. We wanted that definition to be consistent with current applied linguistics theory, so we began talking about how Canale and Swain's work might inform the definition.

3. January 1992, Los Angeles

This meeting began by developing a Statement of Purpose (see Section 1, page 1), which we reaffirmed in Quebec. We continued to discuss how we could define communicative language proficiency in a way that would be useful for test development and validation. We examined Bachman's model. We attempted to examine the implications of Bachman's model by attempting to use it to redefine the familiar four skills. We began to consider how language functions and forms could be listed for each of the four language skills as a starting point — even though we recognized that the test was likely to evolve into some variation on integrated skills modules.

On the basis of a second day's discussion, we decided to define "context" in the model as texts, tasks, and settings and attempted to understand what grammatical, sociolinguistic, and discourse knowledge would mean in contexts associated with each of the four language skills. We agreed that strategic competence, as it is defined in much of the applied linguistics literature, was not what we wanted; so, instead of adopting Bachman's strategic competence, we discussed the need to have some procedural component that included basic notions of processing.

4. May 1992, Quebec

During this COE meeting, we revised the model categories and elements within categories. We also developed the schematic model that appears in the (May 1992) ETS draft document. At this meeting, we made many useful decisions about the development of the model:

> 1. We agreed that the academic setting did not mean the test would include any type of language that might be spoken on a campus; instead, it reflected language that would be relevant and needed for students to succeed at a university.

2. We recognized the need to distinguish the components of academic language that were observable in the academic context versus those processes/components that were internal to the language user (Academic Contexts versus Internal Student operations).

3. The Academic Context included the Situation (task, text, setting) and the Production Output.

4. The Internal Operations had to be keyed to some kind of language user intentions, which then translated the situation into internal verbal processing in Verbal Working Memory.

5. The Internal-Processing Output had to be monitored in some way to match the internal output with the language user's intentions to his or her satisfaction. When this match was made sufficiently, the Internal Output became the Production Output observable in the Academic Context.

6. The major verbal processing done by the language user was within some complex language component that included access to world knowledge, some type of verbal-processing component, and subcomponents of language knowledge (linguistic competence, sociolinguistic competence, discourse competence).

7. The committee agreed that the model being sketched was simply a descriptive schematic model. Specific questions, such as how Verbal Working Memory relates to Internal Goal Setting (if the latter also requires access to the language component), what exists outside verbal working memory in the shaded box, and how world knowledge fits in, were issues that could be considered later. This first model was not an attempt at theory building based on an extensive review of the cognitive psychology research literature.

8. The language competence component was to include linguistic, sociolinguistic, and discourse competence, although its specification was not elaborated far beyond what was discussed in the LA meeting in terms of language skills and language uses. However, some of these concepts were discussed and better clarified.

From these discussions, and some additional notes sent to ETS staff by Carol Chapelle and Bill Grabe, ETS staff put together the *Working Model of Communicative Language Use in an Academic Context* (draft May 1992; see *Appendix C*). This ETS model included (1) a schematic diagram; and (2) elements of the model broken down into each of the four skill areas for ease of discussion.

49

5. May 1992, Charleston

At the May Policy Council meeting two weeks after Quebec, ETS staff presented the May 1992 draft of the working model for feedback from the Policy Council. This draft model was accepted as a tentative draft model to be used as a foundation document for further discussion.

6. October 1992, Princeton

At a joint COE/Research Committee meeting, the May 1992 draft documents were discussed in some detail. In addition, a large group of ETS officers and staff joined in the discussion and assured both committees that this project was a high priority and would receive support.

7. May 1993, Sedona

ETS produced a TOEFL 2000 document that outlined planning background, issues, and next steps. The COE reviewed this document for a day and made a number of recommendations for changing the document (and the priorities for test development). Specifically, the committee recommended that the test-planning documents be organized from a validity-driven standpoint. Computer and technology issues should not be ignored but should be treated in a separate document so that technology is not seen as the driving design.

Other recommendations included position papers parallel to the one being developed for the Test of Spoken English on speaking, each of which would cover a language skill (reading, writing, listening). These papers would cover current research, and particularly research related to assessing communicative competence through each skill (and across skills). The committee also recommended that a prose description of the model of communicative language use be drafted as a foundation document for future planning.

Finally, the committee recommended that the reports and documents use a standardized set of linguistic terminology. This standardization will assist future users of the documents and clarify misconceptions among groups involved in the project. It was agreed that an overview of the model of communicative language use would be the place to present a set of terms and their definitions; this set could then be revised and used in further documents.

Appendix B

Working List of Definitions of Terms for Language Testing

academic context – a university environment where classes are offered on a variety of topics, and students attempt to learn about these topics by attending lectures, reading books, writing papers, and participating in conversations with professors and other students.

academic task – a job that students must accomplish as part of the learning process in connection with a class. Academic tasks include completing lab reports, writing about things they have read, and taking tests covering material learned in class.

applied linguistics – a field of study concerned with language use in context, including teaching and testing contexts.

authentic language tasks – tasks that occur in contexts of language use for purposes other than teaching or testing language. This definition reflects the view that you can only have an authentic task in the contexts in which it normally occurs.

COE Model – the framework that the Committee of Examiners has constructed, on the basis of previous work in applied linguistics, for discussing what TOEFL 2000 is intended to measure.

communicative competence – the ability to use language to express oneself in context. Hymes is responsible for this term and its original definition. Canale and Swain worked on specifying a definition of communicative competence that would be useful for second language teaching and testing.

communicative language ability – the ability to use language to express oneself in context. This term evolved with Bachman's model (which developed Canale & Swain's) to avoid the confusion associated with the multiple meanings of "competence."

communicative language proficiency – the term used to indicate what TOEFL 2000 is intended to measure. This term evolved as a combination of "proficiency testing" (in contrast to placement, achievement, or diagnostic), "communicative competence" (Hymes, 1972), and "communicative language ability" (Bachman, 1990). The purpose of the COE Model is to elaborate what is meant by this term.

construct definition – a theoretical description of the capacity that a test is supposed to measure. A construct definition is formulated on the basis of judgments of experts in the field — judgments that may be informed by a variety of evidence.

construct validity evidence – judgmental and empirical evidence pertaining to what a test measures. This evidence is interpreted relative to a construct definition.

content evidence – judgmental analyses of the knowledge and processes measured by test items/tasks.

discourse analysis – the study of oral or written texts focusing on the elements (e.g., vocabulary and syntactic patterns) contributing to the text, its overall structure, and its context.

discourse competence – knowledge of how language is sequenced and connected appropriately above the sentence level in terms of coherence, information flow, and cohesion.

functional approach to language – an approach to the study of language that focuses on the use (or functions) of language in context rather than on linguistic forms.

grammatical competence (see linguistic competence)

item/task – a unit on a test that requires a test taker to respond. This term is intended to be ambiguous about the form of the units.

language competence – a language user's knowledge of three aspects of language: grammatical, discourse, and sociolinguistic.

linguistic competence – knowledge of formal aspects of language code, such as the formal features of sounds, words, syntactic patterns, and semantic interpretations.

mental model – the reader/listener's interpretation of a text, which is more variable from person to person that the text model and incorporates nonverbal representations into the interpretation.

skills – processing abilities required for performance in four discrete areas: reading, writing, listening, and speaking. We talk about "the four skills."

social consequences – the effects, impact, or washback of tests on the perceptions and practices of those affected. Applied linguists view social consequences as an important aspect of validity.

sociolinguistic competence – knowledge of how sociological phenomena govern linguistic choices and knowledge of a variety of linguistic options appropriate for a range of sociological situations.

strategic competence – the strategies or processes used to put language knowledge to work in context. Strategies include assessing the context, setting goals, etc. Strategic competence was defined by Canale and Swain as a part of communicative competence and included by Bachman as a part of communicative language ability. In the COE Model, the functions performed by strategic competence are represented by "goal setting" and the "language-processing component."

text model – the reader/listener's representation of what a text is about.

validity – justification for test interpretation and use. Justifications include construct validity evidence, evidence about relevance and utility, value implications, and social consequences.

value implications – the academic and social values that underlie testing practices and that are conveyed through testing practices.

Working Documents from TOEFL 2000 Discussions

**TOEFL 2000
PURPOSE AND DOMAIN**

STATEMENT OF PURPOSE

TOEFL 2000 is a measure of communicative language proficiency in English and focuses on academic language and the language of university life. It is designed to be used as one criterion in decision-making for undergraduate and graduate admissions. It is assumed that independent validations will be carried out for other uses of the test.

COMPONENTS OF LANGUAGE USE IN AN ACADEMIC CONTEXT

As a preliminary step in TOEFL 2000 planning and development, the TOEFL Committee of Examiners has attempted to identify the academic domain of language use. The significant components of both comprehension and production of language include Settings, Text Types, Tasks, Procedural Competence, Linguistic Competence, Discourse Competence, Sociolinguistic Competence, and Functions.

The accompanying model suggests that language use is a process involving the dynamic interaction of all these components. The model attempts to distinguish first and foremost the contributions of the academic context (above the line) and the individual student (below the line).

ELEMENTS OF THE TOEFL 2000 MODEL OF COMMUNICATIVE LANGUAGE USE

LISTENING COMPREHENSION

Settings
Lecture hall
Classroom
Laboratory
Extra-instructional settings (library, health clinic, student union, professor's office, museums)
Interactive media library

Text Types
Informal conversations
Formal discussions (having a predetermined purpose)
Interviews
Impromptu monologue or speech
Lectures (presented from outline or notes but not from written text)
Lectures and academic papers (read from written text)
Debates
Newscasts (read from written text)
Formal commentary (read from prepared text)
Orally administered instructions
Narration (read from prepared text)
Story-telling (without a written text)
Poetry or literary pieces
Scripted dialogues (stage or film, radio plays)

READING COMPREHENSION

Settings
Lecture hall
Classroom
Laboratory
Extra-instructional settings (library, health clinic, student union, professor's office, museums)
With support/outside resources (dictionaries, references)
Computer/word processor
Pencil-and-paper/manual

Text Types
Textbooks
Research reports

Summaries
Book reports, reviews
Proposals
Appeals, petitions
Recommendations
Lab reports
Theses, abstracts, dissertations
Journals
Charts, graphs, maps
Literary texts (fiction, poetry, autobiography)
Newspapers
Manuals
Memoranda
Technical and business texts
Case studies
Personal communications (letters, correspondence, professor's written comments)
Notes/outlines
Teacher's comments
Classroom readings (assignment topics, instructions, lecture outlines, test questions, syllabi, course policy)

SPEAKING

Settings
Lecture hall
Classroom
Laboratory
Extra-instructional settings (library, health clinic, student union, professor's office, museums)
Interactive media library

Text Types
Informal conversations
Formal discussions (having a predetermined purpose)
Interviews
Impromptu monologue or speech
Lectures (presented from outline or notes but not from written text)
Lectures and academic papers (read from written text)
Debates
Newscasts (read from written text)
Formal commentary (read from prepared text)
Orally administered instructions
Narration (read from prepared text)
Story-telling (without a written text)
Poetry or literary pieces
Scripted dialogues (stage or film, radio plays)

WRITING

Settings
Lecture hall
Classroom
Laboratory
Extra-instructional settings (library, health clinic, student union, professor's office, museums)
With support/outside resources (dictionaries, references)
Computer/word processor
Pencil-and-paper/manual

Text Types
Essays
Essay test questions
Term papers, research papers, report papers
Project reports
Case studies
Lab reports
Theses, abstracts, dissertations
Summaries
Summaries
Notes/outlines
Book reports/reviews
Letters
Proposals
Recommendations
Appeals/petitions

LISTENING COMPREHENSION

Tasks

Identification of aspects of the code
Orientation (tuning in, preparing to process message)
Comprehension of main idea or gist
Comprehension of details
Full comprehension (main idea plus all details)
Replication (focuses on fidelity of replication)
Extrapolation
Critical analysis
Inference

Procedural Competence

Predicting
Modifying/revising predictions based on new input
Attending to content words
Tolerating ambiguity
Guessing words from context
Checking/indicating comprehension through turn-taking
Judging relative importance of information
Using extralinguistic cues (illustrations, charts, etc.)

READING COMPREHENSION

Tasks

Identification (specific details, recognition, discrimination, focus on the code)
Orientation (author's attitude)
Comprehension of main idea
Comprehension of details
Full comprehension (main idea plus all details)
Replication (focuses on fidelity of replication)
Extrapolation
Critical analysis
Inference

**Procedural Competence
(enhancing or compensatory)**

Skimming
Scanning
Guessing words from context
Predicting
Adjusting reading speed
Re-reading (recognizing misreading)
Recognizing literal vs. nonliteral meaning
Selective reading (skipping parts)
Judging relative importance of information
Using extralinguistic cues (illustrations, charts, etc.)
Rephrasing, paraphrasing during the reading process

SPEAKING

Tasks

Explain/inform/narrate
Persuade
Critique
Synthesize
Describe
Summarize
Demonstrate knowledge
Support opinion
Hypothesize
Give directions
Transcode from charts, graphs, maps, or other text types

Procedural Competence

Circumlocute
Avoid, skip difficult language
Elaborate
Revise
Organize
Exemplify
Use resources/quote
Copy/imitate/reproduce
Paraphrase/rephrase
Use visual/graphic supports

WRITING

Tasks

Explain/inform/narrate
Persuade
Critique
Synthesize
Describe
Summarize
Demonstrate knowledge
Support opinion
Hypothesize
Give directions
Transcode from charts, graphs, maps, or other text types
Write creatively

Procedural Competence

Circumlocute
Avoid, skip difficult language
Elaborate
Revise
Organize
Exemplify
Use resources/quote
Copy/imitate/reproduce
Paraphrase/rephrase
Use visual/graphic supports
Edit/proofread

LISTENING COMPREHENSION
Sociolinguistic Competence

Register Variation
Understand/recognize variations in language
  with respect to:
- The number of listeners in the intended
  audience
- Familiar or distant relationship between
  speaker and audience
- Informal or formal requirements
- Subordinate or superordinate relationships
- General or topical content
- Lay person or specialist as intended
  audience

Recognize paralinguistic cues
Fulfill turn-taking requirements in
  conversational speech

READING COMPREHENSION
Sociolinguistic Competence

Register Variation
Understand/recognize variations in language
  with respect to:
- The number of readers in intended
  audience
- Familiar or distant relationship between
  writer and audience
- Informal or formal requirements
- Subordinate or superordinate relationships
- General or topical content
- Lay person or specialist as intended
  audience

SPEAKING
Sociolinguistic Competence

Register Variation
Produce appropriate language with respect to:
- One or many in intended audience
- Familiar or distant relationship between
  speaker and audience
- Informal or formal requirements
- Subordinate or superordinate relationships
- General or topical content
- Lay person or specialist as audience

Fulfill turn-taking requirements in
  conversational speech

WRITING
Sociolinguistic Competence

Register Variation
Produce appropriate language with respect to:
- One or many in intended audience
- Familiar or distant relationship between
  speaker and audience
- Informal or formal requirements
- Subordinate or superordinate relationships
- General or topical content
- Lay person or specialist readers

LISTENING COMPREHENSION

Linguistic Competence
Recognize phonological features of spoken language
Discriminate among forms or structures
Recognize word order pattern, syntactic patterns and devices, lexical/semantic relations, variations in meaning

Discourse Competence
Understand streams of speech
Recognize thought groups (prosodic patterns)
Infer links between events, situations, ideas
Recognize genre markings
Recognize coherence relationships
Recognize cohesive devices
Follow topic development
Analyze tone of discourse
Recognize conclusion from parts

READING COMPREHENSION

Linguistic Competence
Recognize orthographical features of written language
Discriminate among forms or structures
Recognize word order pattern, syntactic patterns and devices, lexical/semantic relations, variations in meaning

Discourse Competence
Infer links between events (situations, ideas, causes, effects)
Recognize genre markings (features of formal discourse)
Recognize coherence relationships
Recognize cohesive devices
Follow a topic of the discourse
Analyze tone from the various parts
Recognize the parts leading to the whole
Recognize conclusion from parts
Draw conclusions (using multiple bits of information)

SPEAKING

Linguistic Competence
Use appropriate pronunciation, intonation, and stress
Combine forms and structures
Use appropriate word order
Use appropriate forms of words
Use appropriate syntactic patterns and devices

Discourse Competence
Link situations and ideas
Use multiple pieces of information to support conclusions
Use appropriate genre markings (features of formal discourse)
Produce coherent speech
Develop a topic
Use appropriate tone
Use appropriate cohesive devices

WRITING

Linguistic Competence
Use distinctive features of the language
Combine forms and structures
Use appropriate word order
Use appropriate forms of words
Use appropriate syntactic patterns and devices

Discourse Competence
Link situations and ideas
Use multiple pieces of information to support conclusions
Use appropriate genre markings (features of formal discourse)
Establish coherence
Develop a topic
Create appropriate tone
Use appropriate cohesive devices

REVISIONS PROPOSED BY GRABE IN PROSE DESCRIPTION OF MODEL, 5/7/92

Alternative to a separate Text Type list for each skill:

Oral/Aural
Informal conversation
Formal discussion
Interviews/talk show
Impromptu speech
Extemporaneous speech
    [prepared but no notes]
Lectures [prepared notes, not read]
Academic paper [read]
Newscasts [prepared text]
Oral editorial
Oral instructions
Narration
Story-telling [no text]
Debates
Recitation of poetry or literature
Stage or film productions
Radio plays [scripted]

Reading/Writing
Textbooks
Research reports
Journals
Charts, graphs, tables, figures
Newspapers
Literary texts (fiction, poetry, autobiography)
Manuals
Memoranda
Technical/business texts
Case studies
Personal communication
(letters, professor's comments)
Notes/outlines
Teacher comments
Classroom reading (assignment topics, syllabi
    instructions, lecture outlines, test questions,
    blackboard notes)

Alternative to a separate list of Tasks for each skill, separating tasks into reception tasks versus production tasks:

Listening/reading
Identifying aspects of the code
Comprehension of main idea
Comprehension of detail
Full comprehension
Replication
Extrapolation
Critical analysis
Drawing inferences

Speaking/writing
Explain/inform/narrate
Persuade
Critique
Synthesize
Describe
Summarize
Demonstrate knowledge
Support opinion
Hypothesize
Give direction

Alternative to a different "Procedural Competence" list for each skill, taking into account the revised design of the model:

Processing Component

On-line Processing
Visual/Aural sensory input
Lexical Access
Propositional integration
Text modelling
Mental model interpretation

Metacognitive Processing
Skimming
Scanning
Predicting
Adjusting processing speed
Re-reading
Tolerating ambiguity
Summarizing
Paraphrasing (Exemplifying)
Awareness of organization
Use of extra-linguistic cues
Selective processing (skipping, listening to
    two conversations, communicating in
    phrases)
Judging relative importance of information
Recognize literal and non-literal meanings
Using resources (text, quotation)
Editing [production only]
Circumlocuting [production only]
Elaborating [production only]
Revising [production only]
Copying/imitating/reproducing [production
    only]

LISTENING COMPREHENSION	READING COMPREHENSION	SPEAKING	WRITING
Functions	**Functions**	**Functions**	**Functions**
Understand/Recognize:	Understand/Recognize:	Describe	Describe
Description	Description	Narrate	Narrate
Narration	Narration	Inform	Inform
Paraphrase	Paraphrase	Paraphrase	Paraphrase
Directions	Directions	Give directions	Give directions
Definition	Definition	Define	Define
Explanation	Explanation	Explain	Explain
Orders	Orders	Give orders	Give orders
Opinions	Opinions	Give/support opinion	Give/support opinion
Summary	Summary	Summarize	Summarize
Predictions	Predictions	Predict	Predict
Hypothetical language	Hypothetical language	Use hypothetical language	Use hypothetical language
Persuasive language	Persuasive language	Persuade	Persuade
Comparison/contrast	Comparison/contrast	Compare/contrast	Compare/contrast
Cause/effect relationships	Cause/effect relationships	Cause/effect	Cause/effect
Agreement and disagreement	Agreement and disagreement	Disagree/agree	Disagree/agree
Criticism	Criticism	Criticize	Criticize
Approval/disapproval	Approval/disapproval	Approve/disapprove	Approve/disapprove
Expression of feelings/moods	Expression of feelings/moods	Express feelings/moods	Express feelings/moods
Suggestions/Recommendations	Suggestions/Recommendations	Suggest/recommend	Suggest/recommend
Advice	Advice	Advise	Advise
Complaints	Complaints	Complain	Complain
Requests	Requests	Request	Request
		Give feedback	
		Elicit	
		Invite/include others	
		Negotiate	
		Convince	
		Interrupt	
		Apologize	
		Sympathize/console	
		Compliment	
		Congratulate	
		Make introductions	
		Make "small talk"	
		Express greetings/farewells	

FIGURE 3

Working Model of Communicative Language Use
in an Academic Context, April 1992

ACADEMIC CONTEXT
(observable situation)

SITUATION

- task
- text
- setting

PRODUCTION
OUTPUT

STUDENT
(internal operations)

INTERNAL
GOAL-SETTING

INTERNAL
PROCESSING
OUTPUT

VERBAL WORKING MEMORY

WORLD
KNOWLEDGE

LANGUAGE
COMPETENCE

- linguistic
- discourse
- sociolinguistic

VERBAL
PROCESSING
COMPONENT

- on-line processing
- metacognitive
  processing