



TOEFL[®]

Monograph Series

MS - 17
APRIL 2000

TOEFL 2000 Reading Framework: A Working Paper

Mary K. Enright
William Grabe
Keiko Koda
Peter Mosenthal
Patricia Mulcahy-Ernt
Mary Schedl

 *Educational
Testing Service*



**TOEFL 2000 Reading Framework:
A Working Paper**

**Mary K. Enright
William Grabe
Keiko Koda
Peter Mosenthal
Patricia Mulcahy-Ernt
Mary Schedl**

**Educational Testing Service
Princeton, New Jersey
RM-00-4**



Educational Testing Service is an Equal Opportunity/Affirmative Action Employer.

Copyright © 2000 by Educational Testing Service. All rights reserved.

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

EDUCATIONAL TESTING SERVICE, ETS, the ETS logo, GRE, NTE, TOEFL, the TOEFL logo, and TSE are registered trademarks of Educational Testing Service. The modernized ETS logo is a trademark of Educational Testing Service.

Grolier is a registered trademark of Grolier Interactive, Inc.

To obtain more information about TOEFL programs 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 TOEFL test 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 have helped 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 Educational Testing Service (ETS[®]) will evolve into the 21st century. As a first step the TOEFL program recently revised the Test of Spoken English (TSE[®]) and introduced a computer-based version of the TOEFL test. 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 computer-based TOEFL test, introduced in 1998, takes 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

Monographs 16 through 20 are the working papers that lay out the TOEFL 2000 conceptual frameworks with their accompanying research agendas. The initial framework document, Monograph 16, describes the process by which the project will move from identifying the test domain to building an empirically based interpretation of test scores. The subsequent framework documents, Monographs 17-20, extend the conceptual frameworks to the domains of reading, writing, listening, and speaking (both as independent and interdependent domains). As such, the current frameworks do not yet represent a final test model. The final test design will be refined through an iterative process of prototyping and research as the TOEFL 2000 project proceeds.

As TOEFL 2000 projects are completed, monographs and research reports will continue to be released and public review of project work invited.

TOEFL Program Office
Educational Testing Service

Abstract

The TOEFL 2000 framework monograph (Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 1999) identifies a test domain and lays out a process for the design of a new TOEFL test based on communicative language abilities. This monograph on the assessment of reading comprehension addresses the proposed TOEFL 2000 framework described in Jamieson et al. (1999) and defines how it can be realized and implemented in a test of reading comprehension. The reading framework described in this document was developed by the authors, consisting of internal ETS staff and external reading experts, who have worked together over the past two years.

This monograph documents how three broad perspectives were considered in defining the construct of reading comprehension for assessment purposes: a processing perspective, a task perspective, and a reader purpose perspective. The reader purpose perspective is recommended to guide the new test design for a number of reasons. One perceived advantage of this approach is that it is readily interpretable. It will be easier for test-score users, teachers, and examinees to understand how the construct is being defined. At the same time, the reader purpose perspective is seen to be compatible with both the processing perspective and the task perspective.

Four purposes for reading in the academic context are identified: reading to find information, reading for basic comprehension, reading to learn, and reading to integrate information across multiple texts. These four reading purposes are seen to form a natural hierarchy that can serve as a basis for describing a continuum of reading proficiency. The first two purposes are addressed in the current TOEFL reading test format. The third and fourth purposes, reading to learn and reading to integrate information across multiple texts, would expand the construct being measured. Some tasks that might be used to assess reading for different purposes are described.

Finally, technological issues specific to the delivery of the reading test are described and a detailed research agenda related to the reading construct described in this document is provided.

Key phrases: TOEFL 2000 reading, academic reading purposes, new test design, reading to learn, reading multiple texts

Acknowledgments

The authors gratefully acknowledge the support and suggestions of the TOEFL Committee of Examiners, the members of the TOEFL 2000 Committees working on the speaking, listening, and writing frameworks, and the ETS staff for their helpful comments and suggestions.

Table of Contents

	Page
1. Introduction	1
2. Conceptualizing Reading Proficiency	2
The Processing Perspective	2
The Task Perspective	3
The Reader Purpose Perspective	4
Using Reader Purpose to Guide Test Design	4
Reading to Find Information	5
Reading for Basic Comprehension	5
Reading to Learn	6
Reading to Integrate Information	6
A Difficulty Continuum	6
Differences in L1 and L2 Reading	7
Transfer of L1 Reading Skills and Strategies	8
Probable Facilitation Stemming from L1-L2 Structural Similarity	9
Cross-linguistic Interactions during L2 Reading	10
Processing Constraints Resulting from Limited Linguistic Knowledge	11
Implications for Reading Assessment among L2 Learners	12
3. Reading Framework for the TOEFL 2000 Test	14
Identifying the Test Domain	14
Organizing the Test Domain	14
Identifying Task Characteristics	14
Situation	14
Participants	15
Setting	15
Content	16
Communicative Purpose	16
Register	17
Text Material	18
Grammatical/Discourse Features	18
Pragmatic/Rhetorical Features	20
Test Rubric	28
Tasks to Assess the Four Types of Reading	31
Types of Response Formats	37
Linguistic Variables and Task Difficulty	38
4. Technological Considerations	40
The Role of Technology in Reading Comprehension	40
The Reading Comprehension Interface	41
5. Research Agenda	43
Construct Identification	45
Prototyping	46
Pilot Testing	47

	Page
Field Testing.....	47
Further Research.....	48
6. A Better Reading Test.....	49
References.....	51
Appendices	
Appendix A Linguistic and Processing Variables	63
Appendix B Text and Task Variables	67
Appendix C Other Factors that May Contribute to Test Variation.....	71
Appendix D L1 and L2 Differences.....	72
Appendix E Register Variations	74

Table

Table 1	Types of Texts and Tasks for the TOEFL 2000 Reading Test.....	30
---------	---	----

1. Introduction

This document presents the framework for the TOEFL 2000 test of reading comprehension. Part 2 discusses the construct of reading and explains how a “reader purpose” perspective will be used to guide the design of the TOEFL 2000 reading test. Four purposes for reading are highlighted: reading to find information, reading for basic comprehension, reading to learn, and reading to integrate information across multiple texts.

Part 3 reviews the domain for the test proposed in the TOEFL 2000 framework document (Jamieson et al., 1999), describes the TOEFL 2000 organizational scheme, and discusses the proposed task characteristics in terms of the reading test. This section also describes tasks that might be used to assess reading for different purposes.

Part 4 of this document considers technological issues involved in assessing reading comprehension and makes some recommendations for interface design. Part 5 presents a detailed research agenda needed to support the development of the test, and the final section, Part 6, discusses the ways in which the new TOEFL 2000 reading test will improve on the earlier versions of the test.

2. Conceptualizing Reading Proficiency

One of the major challenges for individual differences research remains the discovery of a principled set of processing explanations for individual differences, as opposed to a list of all processes that occur during reading.
(Perfetti, 1997)

Reading comprehension has been viewed from a number of perspectives. In thinking about the construct, or constructs, of reading comprehension for the purposes of the TOEFL 2000 project, we considered three perspectives:

1. The processing perspective,
2. The task perspective, and
3. The reader purpose perspective.

Each of these perspectives is briefly discussed below. Although we settled on the reader purpose perspective as the guiding principle for test design, we believe that the processing perspective and the task perspective can also both be understood from, and inform, this broad framework.

The Processing Perspective

In a recent review of reading comprehension research, Perfetti (1997) argued that research on individual differences among readers is the key to understanding the nature of reading abilities. That is, to understand reading we need to know what factors contribute consistently to differences between better and weaker readers. Perfetti himself suggests that major sources of individual proficiency differences include differences in processing efficiencies such as speed and automaticity of word recognition, thoroughness of word representation knowledge, processing efficiencies in working memory, fluency in syntactic parsing and proposition integration as part of building text comprehension, and the development of an accurate and reasonably complete text model of comprehension (see also Carpenter, Mikaye, & Just, 1994; Perfetti, 1994).

Recent efforts to explore the “simple view of reading” (Gough, Hoover, & Peterson, 1996; Chen & Vellutino, 1997) have also argued that a good part of reading abilities can be related to a combination of word recognition abilities and comprehension abilities. This view would implicate word recognition fluency, accurate word representations, processing efficiency, text model building, inferencing, and strategic processing and monitoring. Similar implications can be drawn from the recent overview by Carver (1997), who brings together a set of research studies to develop a model of components of reading. This model also argues that reading abilities are fundamentally composed of fluency, word recognition accuracy, and rate of processing (which might combine processing efficiency and reading rate).

In second language contexts, both Koda (1996, 1997) and Geva, Wade-Wooley, & Shany (1997) make strong arguments for the importance of fluent word recognition, processing efficiency, and reading rate in second-language (L2) reading.

In language testing contexts, it would seem fairly apparent that vocabulary is a key co-variate with reading, as is, to a lesser extent, some measure of grammar knowledge. The high correlation of listening test scores with reading test scores has been attributed to the importance of general comprehension abilities associated with (a) generating a text model of comprehension, (b) forming an appropriate situation model relating reader knowledge with text information, (c) inferencing of certain types, and (d) monitoring comprehension strategically (see, for example, Sticht & James, 1984; and Appendix A).

Thus, from a processing perspective, a small set of linguistic and processing variables can be said to drive the construct of reading (see Appendix A for a review of research related to linguistic and processing variables).

The Task Perspective

On the assumption that reading can be “defined” in terms of the tasks that readers are able to accomplish, or to accomplish well if they are good readers, it is possible to develop a task-based explanation of skills that readers possess, anchored to reading ability. While it may be possible to develop such explanations using authentic tasks that are carried out in the real world of reading, it is more useful for testing reading comprehension to develop a set of text and task variables that can account for the variance in performance that occurs on reading test questions. Text and task variables such as the frequency and usage of particular words involved in the task, the complexity of syntax, the amount of text that must be processed, and the amount of time allowed for completing a task may account for much of the variance in difficulty and task performance on test questions. Research has already indicated the importance of task variables such as the occurrence or absence of distracting information in the text, the degree to which the correct answer matches the wording of the information in the text itself, and the concreteness of the information requested (Kirsch & Mosenthal, 1990). To the extent that the identified task variables account for performance differences on test questions, they may be said to provide a task-based framework for interpreting the construct of reading as it is instantiated in the test design.

At one level, a degree of research interpretability is lost by using such an approach since much of the research exploring the construct of reading is processing-based. Thus, one cannot directly relate these types of task variables to processing notions such as efficiency of processing, vocabulary knowledge, word recognition, text model formation, etc. At the same time, it is possible to make these connections with a certain amount of reasonable inferencing so that task variables can be argued to “account for” the processing and linguistic variables used in the reading research literature.

At another level, the use of text- and task-based variables to “define” the construct of reading provides a strong advantage: These variables can be used to create an interpretable description of factors which cause difficulty for readers. Such descriptions can also be used to interpret more directly some of the things that readers need to do to become better readers and do so in a way that is not dependent on cognitive processing theories (see Appendix B for a review of text and task variables).

The Reader Purpose Perspective

A third way to conceptualize the construct of reading is to examine the different purposes for which people engage in the process(es) of reading. One advantage of this approach is that the defining notions are fully interpretable as concepts associated with reading comprehension. Thus, “reading for the basic idea” or “reading to learn” are concepts that certainly can be said to belong in a test of academic language abilities. We believe that it is possible to use this reader-purpose perspective as a framework for the TOEFL 2000 reading test and that both the reading-process and task-based views of reading are compatible with it.

Carver (1997) argues that a theory of reading typically needs to account for at least two types of reading: what he calls “rauding” (or basic comprehension) and “reading to learn.” It seems to us that both of these types of reading reflect important academic reading purposes that should be included in our TOEFL 2000 test framework. The purpose for the “rauding” type of reading is general comprehension or comprehension of the major points in a text. With the “reading to learn” type of reading, the purpose is to construct an organized representation of the text that includes major points and supporting details. A third type of reading, “search reading,” is also considered minimally by Carver, in part as either skimming or scanning, and in much greater detail in studies by Guthrie (1988; Guthrie & Kirsch, 1987; Guthrie & Mosenthal, 1987). The purpose for this type of reading is to find discrete pieces of information by skimming and scanning a text or a non-prose document such as a table. We believe this reading purpose should also be considered in planning for the TOEFL 2000 test, since many types of reading practiced by students in academic contexts involve search reading processes, or what we will call “reading to find information.”

One other type of reading that is not accounted for above is the reading of multiple texts. “Reading of multiple texts” represents recent work by Perfetti (1997) and Goldman (1997) on student efforts to generate “intertext models of comprehension.” This type of reading, sometimes called a “documents model” of reading comprehension, has as its purpose the integration of information across multiple texts. Perfetti (1997) has noted that the integration of information across texts depends on the same comprehension abilities involved in the comprehension of single texts, although the purpose is more complex. Interactive reading skills appear to be expected in university classes across a range of disciplines and may be an important aspect of advanced academic reading. We propose, then, that “reading to integrate information across multiple texts” be included in the TOEFL 2000 construct definition as a fourth purpose for reading.

Using Reader Purpose to Guide Test Design

For the present purpose – that is, providing a foundation for a test of academic reading abilities in English as a second language – we recommend that the guiding principle for test design be a single broad construct that includes the four academic reading purposes discussed above:

1. Reading to find information (or “search reading”),
2. Reading for basic comprehension,

-
3. Reading to learn, and
 4. Reading to integrate information across multiple texts.

Each type of reading, or “purpose for reading,” can be seen as representing a variation on one basic reading construct called “reading comprehension.” Each requires a combination of word recognition/processing efficiency and comprehension abilities, and can therefore also be related to the skills-processing perspective and the task perspective.

We believe that this purpose-driven framework will make it possible to explain the principles driving test design and test development to the general public. At the same time, it is possible to overlay frameworks that are driven by processing views of reading and task-based views of reading.

Below we briefly explain how a purpose-driven framework might work for each of the four purposes for reading. The first two types of reader purposes, reading to find information and reading for basic comprehension, are covered well in the current TOEFL reading test. The second two, reading to learn and reading to integrate information across multiple texts, are not currently tested and would expand the construct for the TOEFL 2000 test.

Reading to Find Information. One of the most basic purposes for reading is to locate and comprehend discrete pieces of information. People search text for information in order to find answers to questions that have been posed, to verify and repair any miscomprehension, and to find the most relevant parts of a text for information purposes.

From a skills processing perspective, being proficient at finding discrete information typically involves rapid, automatic identification of words, working memory efficiencies, and fluent reading rates. Scanning text is done at a relatively rapid rate. From a task perspective, item types in this category could include searching for and matching discrete pieces of information or searching a longer text for specific sections.

Reading for Basic Comprehension. We are assuming that in general the person who has the ability to read for basic comprehension also has the ability to find information in a text (Urquhart & Weir, 1998). This comprehension purpose additionally requires a reader to understand the main ideas or the main points of the text, or to form some understanding of the main theme of the text, but does not necessarily require an integrated understanding of how the supporting ideas and factual details of the text form a coherent whole. Reading for basic comprehension involves understanding a subset of individual ideas, primarily those tied to the thematic content. Comprehension of information that is not central to the main idea is unlikely to be required for this reading purpose. In particular, detailed information in the text does not need to be integrated conceptually at this level of understanding beyond the comprehension of main ideas.

From a skills-processing perspective, basic comprehension requires some ability to construct a text model representation of what is read and also the ability to form a relevant situation model.

The ability to comprehend the major ideas from a text is likely to require cycling through and integrating a range of information from various points in the text, which in turn requires a reasonably efficient reading rate. Some tasks based on simpler texts, however, may only require readers to identify a main idea statement (these are more like “locating” tasks). Types of task that might test basic comprehension include distinguishing main ideas from minor ideas or inferring the main topic. If longer passages are included in the test, it might be more appropriate to ask examinees to identify a number of important ideas. Questions testing basic comprehension of individual, discrete, and supporting ideas in the passage would also be included here.

Reading to Learn. We assume that reading to learn incorporates the ability to find information and to develop a basic comprehension of the text. However, reading to learn also requires the reader to integrate and connect the detailed information provided by the author into a coherent whole. This sort of integration requires an understanding of cause-and-effect relationships, comparisons and contrasts, classification relationships, and persuasive intent but this type of reading can be done at a slower rate.

From a skills-processing perspective, reading to learn requires that a reader form linkages between a more elaborated model of text construction and frames (such as cause/effect, compare/contrast) to organize conceptual information and to understand the author’s rhetorical intent. Conceptual knowledge that helps the reader integrate information in a text might include information derived from the text and/or from background knowledge. As such, it might represent an efficient alignment of the text model and the situation model. Tasks that require reading to learn might require the reader to cycle through a range of information, integrate that information, and at points, perhaps, evolve an appropriate rhetorical framework for interpretation (e.g., comparison/contrast, cause/effect).

Reading to Integrate Information. The ability to integrate information from multiple sources implicates all the reading purposes discussed above. This reading purpose requires a reader to integrate information from more than one source. Such tasks require a reader to work across two or more texts and generate an organizing frame that is not explicitly stated. Texts may include diagrams, charts, graphs, illustrations, and prose.

From a skills-processing perspective, an intertext model of comprehension is necessary to account for this type of advanced academic reading. Theories of learning, concept representation, and long-term memory need to be considered. This type of task can never consist of simply locating information. This reading purpose would require multiple cycles of integrating information and would require examinees to generate a conceptual frame.

A Difficulty Continuum. To some extent the four reader purposes form a kind of difficulty continuum. To be sure, easy tasks could be designed for reading to learn or reading to integrate information, and difficult tasks asking examinees to find discrete information or read for general comprehension could be designed by manipulating task and linguistic/syntactic variables. Still, we are more likely to find reading to learn and reading to integrate information across texts associated with more challenging academic tasks and to require more sophisticated processing abilities than reading to find discrete information or reading for general comprehension. The

implication is that, as the reading purpose changes from reading to find information to reading for basic comprehension to reading to learn and reading to integrate information across texts, more reading is required and more efficient strategies are necessary. We believe, therefore, that reader purpose itself may be one of the variables that can contribute to task difficulty when combined with appropriate texts and tasks. It will be important to explore this as part of the research agenda. As with any test situation, a number of other factors that may or may not be aspects of the construct may be involved (see Appendix C). Below we discuss differences in L1 and L2 reading.

Differences in L1 and L2 Reading

It is important to keep in mind that differences between first- and second-language (L1 and L2) readers may influence interpretation of the reading construct. This section discusses fundamental differences in L1 and L2 reading (see Appendix D for a list of additional differences). In this context, “L2 readers” refers specifically to individuals learning to read a second language(s) after achieving reading competence in their L1, since they represent the great majority of TOEFL examinees. With this particular subgroup of L2 readers, three fundamental differences distinguish L1 and L2 reading: (a) L2 readers build on prior L1 reading experience, (b) their reading processes are cross-linguistic, involving two or more languages, and (c) their reading instruction usually commences before adequate oral proficiency in the target language has developed. Not surprisingly, these differences tend to engender qualitatively different comprehension procedures. The uniqueness of L2 reading stems from at least four additional factors beyond those accounting for performance variability within L1 reading: (a) transfer of L1 reading skills and strategies, (b) facilitation resulting from L1–L2 structural similarity, (c) cross-linguistic interactions during L2 reading, and (d) processing constraints imposed by limited linguistic knowledge. Interestingly, these same factors also provide a basis for formulating frameworks through which L2 reading behaviors can be described and explained – as well as clarifying some of the ways in which L2 reading theory recasts accepted L1 constructs.

Such frameworks, moreover, permit several explicit predictions regarding performance diversity, emanating from the unique experiences among L2 readers. One, given that language processing skills are shaped to accommodate the structural and functional peculiarities of a particular language, and that these skills transfer across languages, it is highly conceivable that prior L1 processing experience is directly associated with procedural divergence at certain points in the development of L2 proficiency, invoking qualitative differences in L2 processing behaviors. Two, given that L2 readers have substantial L1 processing experience, it seems reasonable to expect that L2 reading can be facilitated by prior L1 experience, at least to the extent that the L1 and L2 linguistic systems share similar structural properties. This, in turn, suggests that L1-L2 linguistic distance may be responsible, at least in part, for quantitative differences in L2 reading performance between those with related and unrelated L1 backgrounds. Three, since processing competence develops through experiential exposure to a particular language, we can further predict that L1 and L2 processing experiences jointly impact L2 reading skills. And, finally, although cognitively and metacognitively mature, L2 readers are linguistically limited, and thus likely to develop coping tactics, which are, at certain points in the development of L2 proficiency, qualitatively different from those used by L1 readers of the target

language, within their restricted linguistic resources. The sections that follow elaborate upon these predictions based on empirical findings from recent L2 reading research.

Transfer of L1 Reading Skills and Strategies. L1 comprehension studies suggest that cognitive and metacognitive skills, once acquired, are transferable to other situations posing similar cognitive requirements (e.g., Palinscar & Brown, 1984; Raphael & Pearson, 1982; Guthrie, 1988). A number of L2 acquisition studies also demonstrate that various linguistic and metalinguistic elements are transferred from L1 in both oral and written forms of L2 production: e.g., morphosyntactic systems (e.g., Hakuta, 1976; Zehler, 1982; Gundel & Tarone, 1983; Yanco, 1985; Rutherford, 1983), Communicative strategies (e.g., Cohen, Olshtain, & Rosenstein, 1986; Olshtain, 1983; Scarcella, 1983), and pragmatics (e.g., Irujo, 1986). These findings suggest that some reading skills – acquired in one language – can be applied to another language.

A large number of studies have, in fact, investigated reading skills transfer across languages. Two major perspectives have dominated this research: One based on the presupposition that reading procedures are universal across languages (e.g., Goodman, 1973), and the other, on the conviction that reading involves language-specific processes. Earlier transfer studies occurred within the universal framework, focusing on two major issues: (a) the interrelationship between L1 and L2 reading competence (e.g., Skutnabb-Kangass & Toukomaa, 1976; Cummins et al., 1981; Legarreta, 1979; Troike, 1978; Cummins, 1979, 1991), and (b) the conditions inhibiting, or facilitating, reading skills transfer from L1 to L2 (e.g., Clarke, 1979; Devine, 1987, 1988). These studies, however, give little attention to the precise nature of the skills to be transferred from one language to another.

In recent times, controversy has developed among linguists, psychologists, and educators about the universality of language acquisition and processing. Experimental psychologists, for example, are challenging theories of word recognition stemming from data obtained exclusively with English-speaking subjects. Subsequent cross-linguistic investigations have been carried out, comparing skilled L1 readers with varying orthographic backgrounds. The findings generally reinforce the likelihood that different information-processing procedures are used with particular orthographies (e.g., Turvey, Feldman, & Lukatela, 1984; Navon & Shimron, 1984; Hasuike, Tzeng, & Hung, 1986; Sasanuma, 1984; Vaid, 1995). Similarly, child language studies demonstrate that children cannot deal systematically with linguistic forms which violate their perceptions of the prototypical sentence structure of their native languages (e.g., Berman, 1986; Slobin & Bever, 1982; Hakuta, 1982), thereby suggesting that learners are sensitized to the specific linguistic features very early in their language development. Such linguistic conditioning not only serves to shape cognitive strategies appropriate to individual languages, but also plays a central role in regulating the perception and interpretation of linguistic input (e.g., Slobin, 1985; Bates & MacWhinney, 1989).

The language-specific perspective of transfer has emerged from these newer conceptualizations which contradict assumptions underlying the universal perspective. Current transfer studies indeed demonstrate that L2 readers with typologically diverse L1 backgrounds utilize qualitatively different procedures at certain points in development (e.g., Akamatsu, in press; Brown & Haynes, 1985; Green & Meara, 1987), and more critically, that such procedural

diversity among L2 readers is identifiable with structural variations in their respective L1s (e.g., Koda, 1989, 1990, 1993; Ryan & Meara, 1991). A recent study (Koda, 1999) demonstrates, moreover, more subtle, yet potentially significant, L1 influence on L2 lexical development among ESL readers with non-Roman alphabetic (Korean Hangul) and non-alphabetic (Chinese logographic) L1 backgrounds. Given that logographic and alphabetic readers engage in intraword structural analysis to a differential degree during decoding, it was hypothesized that differential amounts of L1 intraword analysis experience would be causally related to the formation of L2 intraword sensitivity and subsequent decoding competence. The results compound an already complex picture. While the two groups differed neither in intraword sensitivity, nor in decoding, a clear contrast existed in the extent to which intraword sensitivity and decoding skills related to reading comprehension: i.e., while the three variables were closely interconnected in the Korean data, no such direct relationships were found in the Chinese data. These results clearly indicate that differential L1 orthographic experience does not, of necessity, result in any quantitative differences, but such L1-based variations may induce a strong preference for particular processing procedures.

Viewed collectively, reading transfer studies make it plain that L1 processing experience has a long-lasting impact on the development of L2 reading skills. Further research is desirable to clarify the extent to which transferred L1 reading skills are incorporated in L2 reading, as well as the ways in which the use of L1 skills alters the basic reading process among L2 learners.

Probable Facilitation Stemming from L1–L2 Structural Similarity. Given that reading skills transfer occurs during L2 processing even when L1 and L2 are typologically unrelated, it can be expected that the development of L2 reading skills is facilitated by L1 processing experience at least to the extent that the two languages share similar structural properties. Should this be the case, it can be further predicted that L1-L2 linguistic distance is in part responsible, at certain points in development, for efficiency differences in L2 reading performance. L2 lexical processing research, involving ESL learners with diverse L1 backgrounds, repeatedly demonstrates superior performance (faster and more accurate) among those with congruent, rather than incongruent, L1 processing experience (e.g., Green & Meara, 1987; Muljani, Koda, & Moates, 1998; Koda, in press). Logically then, the critical question in this research is precisely how L1 and L2 structural similarities facilitate L2 processing performance.

A recent study sheds substantial light on the issue by directly testing the L1-L2 distance effects on L2 processing efficiency (Muljani, Koda, & Moates, 1998). By comparing lexical decision performance among ESL learners with related (Indonesian; Roman-alphabetic) and unrelated (Chinese; logographic) L1 orthographic backgrounds, the study showed first that Indonesian subjects outperformed the Chinese across conditions; and second, that intraword structural congruence (i.e., spelling-pattern consistency) between Indonesian and English benefited the Indonesian, but not Chinese, subjects. The findings clearly suggest that the performance superiority among Indonesian ESL students is attributable to the accelerated efficiency in the precise aspects where their L1 and L2 pose identical processing requirements. L1-L2 linguistic distance, thus, not only explains overall performance differences among learners from related and unrelated L1 backgrounds, but also underscores the ways in which L1 experience facilitates L2 lexical processing.

One way of expanding linguistic distance research would be to systematically compare longitudinal changes in processing competencies among L2 learners with related and unrelated L1 backgrounds. There are at least three logical possibilities. One, once L2 readers gain specific levels of processing competence, linguistic distance should have little effect on performance efficiency or processing strategies. Linguistic distance effects, in short, will be apparent until L2 readers attain processing competence at the threshold level, but not thereafter. Two, the initial efficiency gap will never close. Should this be the case, we can expect that L2 learners from unrelated L1 backgrounds will always lag behind those with related backgrounds. Three, L2 readers from unrelated L1 backgrounds may gain processing efficiency by using qualitatively different processing mechanisms, which will result in diverse processing procedures among different L1 groups.

To date, only a handful of studies on record have directly addressed the L1-L2 distance effects in L2 reading. It would be considerably advantageous if more research could be specifically directed towards the clarification of the long-term impacts of qualitatively different L1 processing experience on L2 reading skills development.

Cross-linguistic Interactions during L2 Reading. Despite the fact that L2 reading involves at least two languages, limited attention has been given to the cognitive interplay between the two languages, as well as its resulting impacts on L2 reading behaviors. A simple comparison of the surface features of the two languages, as in the case of the Contrastive Analysis Hypothesis, only provides a limited view of the potential interaction intricacies. A simplistic analysis may not only yield an inadequate description of the interactions (Zobl, 1983, 1984; Zehler, 1982), but also may engender inaccurate predictions regarding the consequences of the interactions (e.g., Whitman & Jackson, 1972).

Through careful comparisons of intraword awareness among L2 readers, systematic investigations of cross-linguistic interactions during L2 processing are currently underway. Intraword awareness refers to readers' understanding of words' internal structure and their ability to use the structural insights during lexical processing. The role of such awareness in early literacy development has attracted considerable attention among L1 reading researchers for almost two decades. The emerging consensus is that learning to read is fundamentally metalinguistic, involving the recognition of the basic units of spoken language and the units of the writing system, and the mapping between the two (e.g., Nagy & Anderson, 1997; Fowler & Liberman, 1995; Goswami & Bryant, 1992). Recent research, moreover, consistently demonstrates that intraword awareness among young L1 readers develops primarily through cumulative print processing experience in their language (e.g., Perfetti, Beck, Bell, & Hughes, 1987; Bowey & Francis, 1991; Vellutino & Scanlon, 1987). This, in turn, implies that processing experience in the target language is likely to be a major force in shaping intraword awareness among L2 readers, regardless of their L1 backgrounds. Given that L2 readers rely upon L1 processing skills, however, it is conceivable that L2 input is filtered through the structural sensitivity developed in L1 reading (L1 intraword awareness). It can be expected, therefore, that the resulting L2 awareness is an amalgamated form of cross-linguistic interactions of L1 and L2 processing experiences.

Recent studies involving ESL readers (Koda, 1999, in press; Koda, Takahashi, & Fender, in press) repeatedly show that (a) L1-L2 processing congruity is directly related to efficiency differences among L2 readers with varying L1 backgrounds; (b) ESL learners with equivalent proficiency are sensitized, to a similar degree, to the internal structure of English words, regardless of their L1 backgrounds; and (c) variations in L1 processing experience predicted performance differences in some, but not all, aspects of L2 intraword awareness. These results indicate that the aspects of L2 processing competence, central to analyzing and manipulating L2-specific linguistic features, develop primarily through repeated experience with the lexical peculiarities in the target language and, therefore, are largely unaffected by differences in L1 processing experience. Nonetheless, the groups' response patterns systematically varied in other aspects as a consequence of their L1 processing experience. Viewed collectively, the findings suggest that L1 and L2 knowledge are both operative, in their own unique ways, in the development of L2 processing competence. Obviously, further research is necessary – and essential – before we can uncover the complexities of cross-linguistic interactions transpiring in L2 reading.

Processing Constraints Resulting from Limited Linguistic Knowledge. A fourth dimension distinguishing L1 and L2 reading is the degree of linguistic proficiency that learners have acquired prior to reading instruction. In L1 reading, children have already mastered the basic language structure before instruction begins. They are, moreover, continuously exposed to written symbols in their cultural environment (e.g., food packages, commercial logos, trademarks, billboards, etc.), enabling them to formulate visual images of words and establish strong associations between the oral and written forms of their language (e.g., Sulzby, 1986; Ferreiro, 1986; Mason, 1980). This is rarely the case with adult L2 learners. Not only are they required to read before attaining either adequate oral proficiency or “environmental literacy,” but, often, they must also deal with highly decontextualized materials from the outset.

One way of assessing the consequences of insufficient linguistic knowledge is to determine how such limitations constrain L2 reading processes. It is highly probable, for example, that limited linguistic knowledge restricts L2 readers' ability to identify the L2 morphosyntactic features which provide essential information for comprehension. Empirical data, in fact, demonstrate that information sampling patterns among L2 readers are strikingly different from those used by native readers of the target language (e.g., Bernhardt, 1987; Hatch, Polin, & Part, 1974; Saito, 1989). Given that varying aspects of morphosyntactic knowledge contribute differently to sentence processing (e.g., MacWhinney & Bates, 1989; Koda, 1990), the differential visual processing patterns among L1 and L2 readers indicate that L2 readers pay less attention to significant information — and more attention to less significant elements — than their L1 counterparts.

An alternative approach to assessing the effects of limited linguistic knowledge on L2 reading is to compare reading behaviors across proficiency levels. A notable efficiency difference exists in lower-level verbal processing. It has been reported, for example, that with increased L2 proficiency, processing speed improves (Favreau & Segalowitz, 1982; Haynes & Carr, 1990), and error rate decreases (Bernhardt, 1991). Similarly, eye-movement studies consistently demonstrate that while the number of eye fixations does not differ widely across

proficiency levels, the fixation duration among lower-proficiency learners is considerably longer than that among higher-proficiency learners (e.g., Oller, 1972; Oller & Tullius, 1973; Bernhardt, 1987; Saito, 1989).

Recent L2 studies repeatedly show that (a) oral language proficiency is not a strong predictor of either lexical processing or reading comprehension (Durgunoglu, Nagy, & Hancin, 1993; Geva & Siegle, 1999; Gholamain & Geva, 1999) and (b) inefficient word recognition reduces L2 reading performance among otherwise fluent bilinguals (Segalowitz, 1986; Segalowitz, Poulsen, & Komoda, 1991). These findings clearly suggest that linguistic knowledge alone provides a necessary — but insufficient — condition for developing L2 reading competence. Further, it has been suggested that two specific dimensions of linguistic knowledge, orthographic and phonological, independently influence word recognition (e.g., Stanovich & West, 1989; Stanovich, 1991a, 1991b; Adams, 1990; Barker, Torgesen, & Wagner, 1992). Importantly, the two may, or may not, develop concomitantly with other aspects of linguistic knowledge. Presumably, then, we must assume that first, L2 readers do not always possess these specific aspects of linguistic knowledge, and second, even if they do, they may not yet have developed the skills to use the knowledge during reading.

Importantly, inefficient lower-level verbal processing skills have two major consequences for comprehension performance. First, inasmuch as reading comprehension necessitates the construction of meaning based on textual information, the meaning construction process is seriously impaired when sufficient information is not extracted, or when the extraction is inaccurate. Second, underdeveloped processing skills strain limited capacity working memory, restricting higher-level conceptual processing. It seems likely, therefore, that L2 readers engage in text-reader information integration to a far less extent than L1 readers at least until sufficient lower-level processing skills are acquired. It has been reported, in fact, that lower-level processing predominates in the reading process among beginning L2 readers (Clarke, 1979; Horiba, 1993). Hence, the interactive mode of reading, commonly presumed in contemporary L1 reading models, may not be reflective of processing behaviors among L2 readers.

Implications for Reading Assessment among L2 Learners. In an attempt to explicate the singular characteristics of L2 reading, four conditions have been described wherein performance diversity among L2 readers is directly associated with their unique processing experiences. Given that L2 readers rely on L1 processing skills even when L1 and L2 differ typologically, there is good cause to assume that processing procedures used by L2 learners with diverse L1 backgrounds are qualitatively different at specific points in development. This, in turn, suggests that all L2 learners may not develop the exact same skills repertory, and therefore, the cognitive tasks required in L2 reading may not be accomplished in the same way. These unique characteristics of L2 readers yield at least two important implications for L2 reading assessment.

First, L2 reading research demonstrates that the degree of L1 and L2 processing congruity is closely related to performance efficiency — particularly processing speed — among L2 readers. It seems likely, therefore, that L1-L2 linguistic distance may have a definitive impact on assessment outcomes. Although superior processing performance is observable among those with similar L1 backgrounds, benefits of this sort of facilitation occur only when L1 and L2 pose

identical processing demands, and generally are not transferable to the other sub-component processes. Higher efficiency in one aspect resulting from congruent processing procedures, in short, can spawn exaggerated, misleading indicators in overall reading competence among those with typologically related L1 backgrounds. The restrictive, uni-dimensional nature of L1-linked facilitation, thus, prompts the need for caution in using timed tasks in assessing a limited range of L2 reading skills.

Second, current L1 reading models commonly assume that reading comprehension stems from reader-text interaction. L2 research, nonetheless, repeatedly shows that lower-level processing predominates in the L2 reading, thus indicating that L2 readers engage in far less interactive processing than generally assumed in L1 reading. This discrepancy clearly warrants special consideration: Namely, performance data obtained through assessment procedures based solely on L1 principles may not accurately reflect presumed reading capability, especially among beginning L2 readers.

3. Reading Framework for the TOEFL 2000 Test

Identifying the Test Domain

The framework document for the TOEFL 2000 project identified the TOEFL 2000 test domain (Jamieson et al., 1999, p. 10). Since the test is intended to “measure examinees’ English-language proficiency in situations and tasks reflective of university life in North America,” the reading component of the test will reflect the types of reading that occur in university-level academic settings. The test will be designed to improve discrimination at the upper levels of English as a second language/English as a foreign language (ESL/EFL) reading proficiency.

Organizing the Test Domain

It was decided that the most practical way to organize language tasks for the TOEFL 2000 framework is by modality (Jamieson et al., 1999). This decision was influenced by the desires of admissions officers and graduate deans. Thus, “the test will include measures of speaking, writing, listening, and reading. Within these four areas, the new test will include a variety of language features, including not only grammar and vocabulary but also discourse, pragmatics, and sociolinguistics as well as setting and task” (p. 12). It is important to note that the four modalities will be tested both independently and integratively.

Identifying Task Characteristics

Assuming the TOEFL 2000 test domain and organizational scheme described above, we will now consider reading task characteristics in terms of the three broad areas identified in the TOEFL 2000 framework document (Jamieson et al., 1999). The following sections discuss situation, text material, and test rubric respectively.

Situation. As outlined in *TOEFL 2000 Framework: A Working Paper* (Jamieson et al., 1999) situation task characteristics are defined, based on Crystal (1991, 1992) as “extralinguistic elements associated with language tasks.” Task characteristics to be considered include participants, setting, content, purpose, and register.

Most academic reading occurs without reliance on shared physical or social contexts. These aspects of situation seem to be more relevant to language modes such as speaking and listening, which require more interpersonal skills than reading does. As Cummins (1991) points out, written language is relatively decontextualized in the sense that the type of paralinguistic cues that exist in face-to-face oral communication are missing. The immediate feedback that can help correct breakdowns in communication is also missing. Durgunoglu (1997) argues that “in contrast to the current emphasis on contextualization and communicative competence, formal vocabulary knowledge is more strongly related to L2 reading proficiency than oral proficiency is.”

To some extent, therefore, we address issues related to participants and setting from a necessarily different perspective than that likely to be used in the other skill modalities. Where reading tasks are integrated with tasks in other language skill areas, however, other factors will

naturally influence these aspects of the reading situation. For example, in a combined reading and listening task, the participants and setting may influence the overall situational context.

Participants

TOEFL 2000 Framework: A Working Paper (Jamieson et al., 1999) discusses participants in terms of the people involved in the language act and the relationships among them, and suggests that this variable can be operationalized in terms of gender, ethnicity, age, and role. The identification of participants in the reading context, however, requires a more abstract view of the participant than it does, say, for the listening, speaking, or even the writing context.

One participant is clearly the reader, who constructs a text model based on the information in the text and a situation model to interpret the text that is based on his or her background knowledge, goals for reading, motivation, attitudes, and evaluation of the information in the text (Kintsch, 1995).

The authors of the reading passages or documents on the test may also be viewed as participants. In some circumstances, e.g., in an argument in which authors do not agree, characteristics of the authors as participants are likely to be relevant to understanding the text. In other cases, the text's author is fairly anonymous and is treated as unopinionated. This is the pattern in the current TOEFL test passages. The role of the author as participant is likely to be less important in some types of texts, such as expository texts, than in other types of texts, such as argumentative ones. In cases where argumentative or persuasive texts are included, we would not expect it to matter who the participants are in terms of gender, ethnicity, or age, but we believe the role of the participants in the sense of their argumentative stance, their motives, and interests in the argument, and so forth, may influence task difficulty.

Biographical/autobiographical narrative raises the author issue with respect to the relationship between the author and the person being written about. If historical biographies/autobiographies are used in the TOEFL 2000 test, this issue will need to be addressed.

Setting

In *TOEFL 2000 Framework: A Working Paper* (Jamieson et al., 1999, p. 15), setting is defined as "the place where the language act occurs." The framework document proposes that three types of setting be represented: instructional milieus, which include "all places where formal instruction takes place, such as lecture halls, labs, seminar rooms, and classrooms;" academic milieus, which include "typical places outside of the classroom where aspects of academic life are dealt with, such as a study room in a dormitory, the library, an instructor's office, the bookstore, a writing center, or a computer center;" and non-academic milieus, which include places that are not usually associated with academic content but where social and business transactions take place, such as "the business office, international students' office, and the health center, as well as dormitory room and dining areas."

While authentic reading activities can certainly be carried out in each of these environments, it does not seem likely that these different settings would require different types of reading texts, or would influence the difficulty of the reading tasks associated with them. Therefore, we do not recommend that these types of variations in settings be included in the design of the reading test per se. The setting where a text appears could, of course, be addressed by indicating the source from which it was taken (textbook, journal article, etc.) and providing a title or some type of framing information.

Physical settings for reading activities could most authentically be created in combination with the measurement of other language skills. In the case of the reading construct, setting might be relevant to some extent in situations involving reading to integrate across texts. For example, in a lecture hall setting, where an examinee listens to a more extended lecture, some things might be written on blackboards or presented as slides or handouts for reading. Or in a laboratory class, after some oral instruction, a lab manual might present some written text information that needs to be read before answering questions by writing or speaking. In order to situate the reading tasks in a specific location, however, an oral component seems necessary. It is assumed, therefore, that considerations of physical setting for reading tasks will generally be considered in integrative contexts where the situational features of the oral environment might provide greater authenticity.

Content

Any subject area that is typical of academic study could provide appropriate material for the reading test. The current TOEFL test covers a range of very general academic topics broadly classified as topics related to the Arts, Humanities, Social Sciences, Physical Sciences, or Life Sciences. It seems appropriate to continue to include as much topic variety as possible in the new test. As with the current test, however, care should be taken to ensure that specialized knowledge of a particular field is not necessary to understand the information presented in the passages.

Appropriate pragmatic and rhetorical features for TOEFL 2000 reading texts are discussed under the section on text material.

Communicative Purpose

In *TOEFL 2000 Framework: A Working Paper* (Jamieson et al., 1999, p. 15), purpose is defined as “the reason why we engage in tasks.” This definition is then linked to Halliday’s (1973) list of seven categories, six of which are identified as relevant to TOEFL 2000 purposes for which international students would use English in a North American university, namely for heuristic, instrumental, regulatory, personal representational, and interactional purposes. While some or all of these purposes are very relevant to other skill modalities, the heuristic purpose is most relevant to reading in an academic context. Assuming that most of the reading that college students do is heuristic, we believe that the reader’s purpose and goals, more narrowly defined, are critical to reading performance. Van Dijk (1985) has argued that strategies are selected in terms of the reader’s purpose and determined need. Goldman (1997) also notes that “. . . readers’ expectations about their task determine the knowledge and strategies that are brought to bear

during the comprehension process” (p. 366). Such a statement suggests a purpose-driven framework linked more to processing demand than communicative goal. The purpose-driven framework we have proposed for the reading test reflects this emphasis: (a) Reading to Find Discrete Information, (b) Reading for Basic Comprehension, (c) Reading to Learn, and (d) Reading to Integrate Information across Texts.

Register

The term register, as a cover for the uses of texts and the author’s intentions, is reflected in patterns of linguistic features that tend to co-occur as well as arrangements of discourse information and conventional genre forms. For some researchers, register simply refers to occupationally defined texts and specific textual sub-types such as legal register, business register, sports-announcer register, etc. For other researchers, register is equated with genre, and specific functional text types are seen as registers. Examples of these would include sermons, lectures, letters of apology, etc. For yet other researchers, register is a superordinate category for the author’s intent, e.g., to establish relations; provide information; narrate stories, procedures, or ideas; indicate level of support; indicate closeness of relationship; or imply factualness. Biber (1995) gives a good review of these positions.

Viewing register in terms of the varying functional uses that texts serve seems to be most relevant, and most useful, for the TOEFL 2000 project. Identifying occupationally defined language usually does not specify texts that occur all across the space of occupational interactions, nor does it define in sufficient detail the critical marking aspects. A focus only on specific genres is probably not very useful since most testing contexts will allow only a fairly limited range of general texts as input for testing purposes.

A number of authors have proposed a set of underlying parameters that reflect the many uses texts serve that are signaled by, and through, texts. The key to understanding and using these various efforts is to recognize that the parameters are all active and play some role in every text. So, unlike discussion of structure or speech acts or some discussions of style, the many parameters of a useful register theory are all operating at the same time (see Appendix E for a more detailed review of several sets of register variations).

In our view, the best way to proceed would be to set certain register parameters for the TOEFL 2000 reading test and stay within them for the most part in order to develop test forms that are parallel in terms of register. This would mean establishing some system for analyzing the register parameters of reading passages, then incorporating this information into test specifications in order to ensure that test forms of comparable overall register variation are administered to different examinee populations. Later research and experimentation can establish whether or not additional register dimensions should be added. The most common assumption among test score users, and among L2 test takers, is that expository prose types with minimal emotional impact and vagueness constitute the target reading text model (though we will explore argument and historical/autobiographical narrative). We should have strong reasons for changing these assumptions. There are plenty of resources available to continue development of typical TOEFL reading passages of varying length and complexity, and there are many ways to create

new item types from these passages without adding completely new dimensions to the test development matrix from the perspective of register variation.

Text Material. We recommend that tasks in the TOEFL 2000 reading test be based on a variety of text materials. Some could be based on a single text, ranging in length from a short paragraph to a lengthy selection; others could draw on multiple texts, also ranging in length. In some cases, the text might be presented without any visual material, while in other cases a text selection (or selections) could be accompanied by graphics, such as line drawings, schematics, photographs, or maps, or by charts, tables, or graphs presenting data. If technology permits, some of the tasks might incorporate video material.

As defined in *TOEFL 2000 Framework: A Working Paper* (Jamieson et al., 1999) text material for the TOEFL 2000 test will consist of three types of features: (a) grammatical features, which relate to the syntax of the sentences and the vocabulary used in the text; (b) discourse features, which relate to specific discourse features that signal relationships among parts of the text; and (c) pragmatic features, which relate to the intent of the text's creator and, the rhetorical features or organizational patterns of the text. Thus grammatical, discourse, and pragmatic features have been regrouped here into two categories: grammatical/discourse features, and pragmatic/rhetorical features.

Grammatical/Discourse Features

Syntax. The role of syntax is likely to be very important with respect to understanding the construct of reading but much less important as a set of knowledge and related abilities to include in planning task difficulty and developing item types for the TOEFL 2000 test. The contribution of syntactic knowledge to processing efficiencies in reading comprehension was briefly discussed earlier. In particular, syntactic information in reading supports the extraction of appropriate propositional information and also comprises a general set of instructions to readers for building text-model structures that map onto existing structures or generate new sub-structures. Exactly what aspects of syntactic information appear to contribute most to the processes of mapping structures and generating new structures ("shifting" in Gernsbacher's terminology) is not clear, but the current research of Gernsbacher and others is building a better picture (Gernsbacher 1990, 1996, 1997; Givón, 1995; Perfetti, 1997). For example, some research has explored the contributions of articles, tense and aspect marking, locatives, and other systems and structures. It should also be noted at the outset that syntactic features and discourse features may not be as separable as many previous analyses assume; the combining of syntactic markers and discourse features is common in the work cited above, as well as in the work of Freedle and Kostin (1993), Nissan, De Vincenzi, and Tang (1996), and others.

The line of research on syntactic contributions to discourse processing does not typically propose that there are specific syntactic structures that contribute to reading comprehension difficulty in any way that would suggest the assessment of specific isolated structures. Rather, the notion of syntactic support for reading comprehension rests more with the combined sets of signals that structural information provides: it contributes to efficient processing of information in working memory, it establishes and supports semantic relations between arguments and

predicates for proposition formation, and it adds contextual information to help disambiguate lexical meanings. It may also provide important information for other purposes.

With respect to syntactic information as a source of specific reading difficulties, few research studies have isolated significant structural characteristics that deserve to be included as task development variables for testing purposes. Freedle and Kostin (1993) and Freedle (1997) have argued that a small subset of syntactic variables accounts for difficulty in main idea reading comprehension.

Another source of syntactic/structural differences is noted in the work of Just and Carpenter (1992; Carpenter et al., 1994). They argue that syntactic complexity, as a singular concept, leads to lower reading efficiencies. They also note that competing noun-predicate (NP) arguments in the immediate environment reduce processing efficiencies, as does referential distance; the longer the distance between antecedent and co-referent, the lower the efficiency.

Vocabulary. The role of vocabulary in determining task and item difficulty is likely to be a large one. In addition to needing some way to provide frequency data on vocabulary use, vocabulary indices for which definitions are used for a word (knowing a word's different meanings), and for other possible variations (see below) need to be explored. Formality measures (a register feature) related to vocabulary can be assessed by determining the percentage of Latinate words in a text, or word length in a text. This particular measure has generated mixed results, but it should be examined again for the TOEFL 2000 test. Type-token measures or some other measure of the number of new words in a text should be explored. Additionally, a measure of the percentage of uncommon words in a text – an uncommon-type/token ratio – might be more revealing for text readability for second-language (L2) readers. It might also be useful to investigate the difficulty of vocabulary by text type. Other factors relating to vocabulary might include any or all of the following (see Richards, 1976; Nation, 1990):

- Collocability (in fairly tight phrasings)
- Functional limits according to use and situation (register and function constraints)
- Syntactic behavior (parts of speech, sub-categorization, case roles, transitivity)
- Basic forms and derivational possibilities and typical occurrences
- Associational patterns with other words in domains of knowledge and use
- Idiosyncratic features of specific words
- Learning difficulties of certain words (e.g., similar-looking words with different meanings)
- Degree of abstractness-concreteness

Discourse features. Still other consistent contributors to text processing difficulties may be found at the level of discourse organization. The specific role of transition markers is commonly examined, though there does not appear to be a clear synthesis of this research available. Part of the problem is that this topic is addressed through a number of discipline areas and research methods; further, it is not easy to delimit the scope of the notion of “transitional marker.” Most researchers simply note the subset they want to address with a brief nod to some prior rationale,

but this methodological step does not help us establish the potential contribution of transitional discourse markers to differences in reading comprehension abilities.

A number of other features of discourse marking and information structuring should be considered at some point by TOEFL 2000 research studies. These include the roles of theme-rheme structuring, given and new information, definiteness and indefiniteness, noun-predicate density in texts, the positions of main idea and topic sentences in texts, as well as other issues that can be developed from the research literature. A full study of such potential contributing features might be a useful resource document for future specific research studies and for the development of comprehension item types.

Pragmatic/Rhetorical Features

Written discourse has been classified by researchers in a variety of schemes (Brewer, 1980; Britton & Black, 1985; D'Angelo, 1980; Moffett, 1983; Mosenthal, 1985; Vacca & Vacca, 1996; van Dijk & Kintsch, 1983). We recommend that reading passages for the TOEFL 2000 test be classified according to their dominant pragmatic and rhetorical features. The pragmatic features convey the primary intent of the author, while the rhetorical features indicate the higher level organization of the text. The following sections describe these features.

Pragmatic features. According to Brewer (1980), written discourse may serve to inform, entertain, or persuade, or may have a literary-aesthetic intent. The TOEFL 2000 test will focus on academic reading tasks across a range of topics and fields; appropriate types of text materials in these subject areas are texts classified as “expository” or “argumentative persuasive,” the primary intent being to inform or persuade the reader. An additional type of text in these areas may be classified as “historical/biographical narrative,” which would include passages about the contributions of individuals to the disciplines. Fuller descriptions of each of these types of text materials follow.

Exposition – This type of text primarily serves to inform the reader. Often this is the type of reading most prevalent in college classrooms. Lengthy expository text passages may include descriptions, comparisons, contrasts, explanations, and elaborations that provide details about concepts, objects, persons, places, events, and other phenomena. The following passage is characteristic of expository text in that it informs the reader, in this case about the different types of brain wave electrical activity measured by an electroencephalogram.

When you are fully awake and alert, your EEGs contain many beta waves, relatively high-frequency (14 to 30 Hz), low-voltage activity. As you enter a quiet, resting state – for example, just after getting into bed and turning out the light – beta waves are replaced by alpha waves, EEG activity that is somewhat lower in frequency (8 to 13 Hz) but slightly higher in voltage. As you begin to fall asleep, alpha waves are replaced by even slower, higher-voltage delta waves. The appearance of delta waves seems to reflect the fact that increasingly large numbers of neurons are firing together in a synchronized manner.

(Baron, 1992, p. 140)

Typically, the rhetorical features of the text (discussed later in this section) guide the reader in a basic understanding of the passage. Readers who are new to the discipline are often unfamiliar with the vocabulary that is introduced in expository passages; accordingly they must utilize accompanying figures for clarification of the new material, and must have well-developed reading and study strategies for choosing the salient aspects of the passage.

Argumentation/Persuasion/Evaluation – Argumentative/persuasive texts present a point of view about a topic and provide supporting evidence in favor of a position in the analysis of the topic. Good persuasive texts will include carefully crafted positions with reasons and evidence along with an analysis of the opponent’s errors in reasoning.

Argumentative/persuasive texts are characterized by diction that may be personal in tone, by vocabulary that points to an attitude or perspective toward the topic, and by a style that departs from a measured, unbiased stance. For instance, the following example of an argumentative text shows the writer’s opinion about the custom of sending Christmas letters.

I would like to hold a contest for the most fatuous Christmas letter, but I’m afraid I’d be deluged with entries. It is hard to attribute the Christmas letter to a particular type of person or a particular station in life, because almost everyone who has ever had an eventful year, taken an exciting trip, or accomplished a great deal has felt the urge to compose one. I have received them from internationally famous professors who were attempting to describe their world travels, from graduate students describing their Ph.D. research in the field, and from relatives recounting the latest family gossip. Perhaps mimeographed Christmas letters should be used as a vanity indicator, since they expose those among us who yielded to, rather than resisted, the pervasive temptation to blow one’s own horn.

(Johnson, 1971, pp. 44-5)

Argumentative texts are typically found in editorials, essays, political satire, and other types of texts where the intention is to present a point of view. The work of the reader is to analyze the writer’s perspective in relation to the topic and judge the worth of the writer’s presentation, line of reasoning, and evidence.

Historical Biographical/Autobiographical Narrative -- Narrative text tells a story. Consequently, essential elements in this type of text are the setting (which includes the characters and story context) and episodes to reach a goal or solve a problem, which include the initiating event, internal response, attempt, the consequence, and the reaction (Mandler & Johnson, 1977; Stein & Glenn, 1979; Stein & Policastro, 1984). These descriptors of story elements constitute a story grammar and are considered critical components for comprehension of narrative text.

The most common types of college academic reading that contain narrative discourse are historical, biographical/autobiographical, or literary (fiction) text. The TOEFL 2000 test will include passages from a variety of fields, but the use of literary text is not recommended because of fairness issues associated with the cultural background knowledge inherent in such texts.

Types of text selected from historical biographical/autobiographical narratives feature the lives of prominent individuals throughout history. The following examples illustrate this type of passage.

For a few weeks every autumn, when the fields around Chicago were ripe for harvesting, children of Mexican migrant workers joined our classes at the Eugene Field School in Park Ridge. They never stayed long, because their families were always moving on to the next harvest. One year, a boy who was older than I and big for his age took to pushing me and my friends around on the playground. Whatever motivated him, his bullying quickly aroused my fear and my dislike. My reactions to this one boy might well have spilled over to my feelings about the rest of the migrant kids if my mother had not encouraged me to volunteer, along with other girls in my church youth group, to baby-sit for the migrants' younger children on Saturdays so that their older brothers and sisters – my classmates – could join their parents working in the fields.

Just seeing the camp where the families lived made me think for the first time about how my classmates spent their time when they were not in school. I had never before known people who lived in trailers. When we went inside, the mothers seemed nervous about leaving their babies and toddlers in the care of twelve-year-olds who spoke no Spanish. I began to realize the lack of familiarity cut both ways.

The day passed uneventfully, and when the mothers returned, they expressed their pleasure at seeing their children well cared for. It was the return of the fathers, though, that made the greatest impression on me. When the buses dropped them off at the base of the long road to the trailer camp, the children ran as fast as they could to greet them. They were filled with excitement, the same excitement I felt when my own father came home from work at the end of the day. Suddenly those migrant children didn't seem so different from me. This brief encounter helped me begin to appreciate the importance of making judgments about individuals instead of stereotyping whole groups. It also gave me a lot of satisfaction at an early age to be serving families who worked so hard for so little.

(Clinton, 1996, pp. 182-3)

Jung began attending seances and table turnings which were held at the home of relatives every Saturday night. His interest in the occult never diminished, and for his doctoral dissertation he investigated the behavior of a medium, a fifteen year-old girl who performed at the seances of his relatives.

These mysterious phenomena were instrumental in turning Jung's interest to psychology and psychopathology. That fall when he returned to the university he read a textbook on psychiatry by Krafft-Ebing in preparation for his final examinations. The first chapter struck him like a bolt of lightning; he knew immediately that psychiatry was his destined field. In his twenty-fourth year,

then, Jung had finally found the field that was compatible with his interests, speculations, and ambitions. Everything fell into place.

His professors were dismayed by his decision. They were astonished that he would sacrifice a promising medical career for such an absurd field as psychiatry. The medical profession generally was contemptuous of psychiatry; they thought it a lot of nonsense and considered the psychiatrist as peculiar as the patients he treated. Jung characteristically held a firm position on his choice.

(Hall & Nordby, 1973, pp. 21-2)

Although historical biographical/autobiographical narratives vary in style, the primary distinction between this type of text and the other classifications (expository or argumentative/persuasive) is that the writer focuses on a historical sequence of events in the person's life. Further, the author may choose salient events that point towards the reason this particular person is remembered, or, if autobiographical, the authors may choose events that were critical to who he or she has become. Unlike the events related in literary texts, such descriptions of historical and biographical/autobiographical events are recounted as factual episodes. Thus, the intent of this type of text is to inform the reader through a narrative of true events about significant outcomes in the history of the discipline, whether it is psychology, sociology, art, history, biology, or botany, or about a significant impact the events have had on the lives of prominent individuals. Therefore, the task of the reader is to analyze these significant events and to infer the historical and conceptual link to the discipline or individual if not explicitly stated.

Rhetorical features. For the TOEFL 2000 reading test, we recommend that the top-level rhetorical patterns of texts be taken into account in addition to their pragmatic intent. Common rhetorical classifications are taught in many college freshman composition classes (Axelrod & Cooper, 1996; D'Angelo, 1980; Hale et al., 1996) and in content area reading classes (Vacca & Vacca, 1996). Although writers often use a variety of rhetorical types throughout extensive pieces of text, they also employ specific rhetorical patterns, particularly in expository passages, to convey an overall theme or main point in shorter passages. Coupled with the use of cohesive ties to provide transitions between supporting points, writers selectively use rhetorical types to shape ideas and present a cogent point to the reader.

Rhetorical patterns from *TOEFL 2000 Framework: A Working Paper* (Jamieson et al., 1999) are listed in Appendix D of that document. The following is an elaboration of these patterns for the reading section of the TOEFL 2000 test using authentic academic reading texts as examples.

Definition – Writers use definitions to explicate the meanings of terms. When using technical terminology, writers may employ a simple definition, which may be a phrase or sentence; readers may utilize context clues to determine such definitions. However, as a rhetorical pattern, writers use extended definitions to provide full descriptions of concepts. The intent of the entire passage is to describe unfamiliar terminology, elaborate on terms specific to the discipline, and clarify specific uses of the terminology. The following text is an example of an extended definition for the concept of “pollutant” in a college biology lab manual.

A pollutant that reduces a certain population at any step of this food chain will have a similar effect on all other levels of the food chain because all steps of the chain are linked. Thus, water polluted with chemicals such as herbicides that kill algae will decrease populations of zooplankton that eat the algae; this ultimately affects populations of fish and humans that are part of the same food chain. A pollutant is any physical or chemical agent that decreases the aesthetic value, economic productivity, or health of the biosphere. There are many kinds of pollutants, such as noise, chemicals, radiation, and heat.

(Vodopich & Moore, 1996, p. 161)

Illustration – To explicate a concept, a writer may choose to provide examples, a short anecdote, or a familiar description so that vague or abstract concepts are fully described in concrete terms. Cohesive ties that mark this type of rhetorical pattern include “for example” and “to illustrate.” The following example about jet lag illustrates the concept of circadian rhythms.

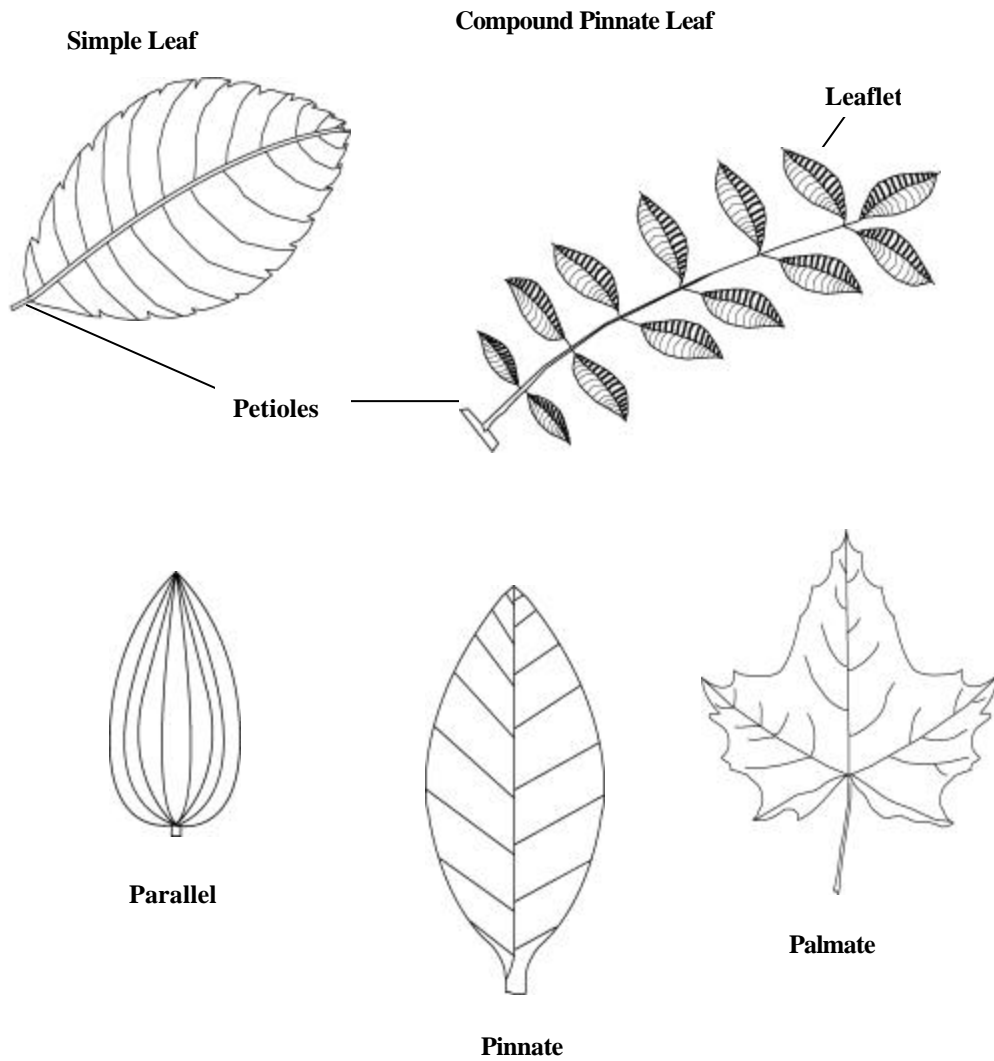
Under normal conditions, circadian rhythms pose no special problems. Most people try to schedule their activities to coincide with their personal highs and lows. Unfortunately, though, there are circumstances under which circadian rhythms may get badly out of phase with our daily activities.

The first of these occurs as a result of modern travel – especially by jet plane. When individuals cross time zones, they have experienced considerable difficulty in adjusting to their new location. The reason for this is clear: Their internal biological clocks are calling for one level or type of activity, while the external world is calling for another one. For example, consider a traveler who flies from New York to London. She departs at 8:00 p.m. and, after a six-hour flight, arrives in London at what her body perceives to be 2:00 a.m. She is very tired and would like to sleep, but in London it is now 7:00 a.m. The day is just shifting into high gear, people are eating their breakfast, and it is broad daylight. The result: Our traveler feels awful. Yes, she can cope – and a few cups of coffee or tea help her get revved up. Yet, if she is like most people, she will soon feel a return of fatigue and will experience many unpleasant sensations as the day progresses. Then, to make matters worse, when she gets into bed at 10:00 p.m. London time, her body has not yet reset its biological clock; to her body it is something like 5:00 p.m., a time when she usually feels especially active. The result: She has trouble falling asleep and experiences even more discomfort.

(Baron, 1992, pp. 132-3)

Classification – This rhetorical pattern groups several items together according to similar features or principles. Writers employ a classification scheme in order to indicate how discrete items belong to a larger group. Readers, in turn, need to note such groupings. Cohesive markers which indicate that the writer is generating such a pattern include transitions marking

membership in a class. Textbooks may include classification charts, hierarchical arrays, or tree diagrams that accompany the text passage to further clarify which items belong to a specific class and which terms provide the general category name. The following passage illustrates this rhetorical text structure.



Leaves typically consist of a **blade** and a **petiole**. The petiole attaches the leaf blade to the stem. Simple leaves have one blade connected to the petiole, whereas **compound leaves** have several leaflets sharing one petiole. **Palmate** leaflets of a compound leaf arise from a central area, as your fingers arise from your palm. **Pinnate** leaflets arise in rows along a central midline.

Leaves are also classified according to their venation (i.e., arrangement of veins). **Parallel veins** extend the entire length of the leaf with little or no crosslinking. **Pinnately veined** leaves have one major vein (i.e., a midrib) from which other veins branch. **Palmately veined** leaves have several main veins each having branches. Veins of vascular tissue in leaves are continuous with vascular bundles in stems.

(Vodopich & Moore, 1996, pp. 270-1)

Comparison/Contrast – Writers employ this rhetorical pattern to designate distinctions among concepts, particularly regarding their similarity or dissimilarity. Cohesive markers such as “similar,” “compared to,” “in contrast to,” or “different from” designate whether concepts are being compared or contrasted. The following passage illustrates this rhetorical pattern.

This automatic processing involves the performance of activities with relatively little conscious awareness. Such processing seems to make little demand on our attentional capacity. Thus, several activities, each under automatic control, can occur at the same time. . . . You engage in automatic processing when you drive your car and listen to the radio at the same time. Automatic processing with respect to a given activity tends to develop with practice, as the components of the activity become well learned and associated with specific stimulus conditions.

In contrast, controlled processing involves more effortful and conscious control of behavior. While it is occurring, you direct careful attention to the task at hand and concentrate on it. Obviously, this type of processing does consume significant attentional capacity. As a result, only one task requiring controlled processing can usually be performed at a time.

Research on the nature of automatic and controlled processing suggests that they differ in several aspects. First, as you might guess, behaviors that have come under the control of automatic processing are performed more quickly and with less effort than ones that require controlled processing. . . .

(Baron, 1992, pp. 134-5)

Cause/Effect – When writers wish to explain why something happened, or the effects of something, they use a cause/effect pattern. Cohesive markers such as “as a result,” “the effect of,” “because,” and “consequently” are often indicators of this pattern. When reading cause/effect texts, readers need to analyze the causes and effects in relation to the overall point.

Another, especially disturbing type of sleep disorder is apnea. Persons suffering from apnea actually stop breathing when they fall asleep. Needless to say, this often causes them to wake up. Since this process can be repeated literally hundreds of times during the night, apnea can seriously affect the health of persons suffering from it.

(Baron, 1992, p. 143)

Problem/Solution – Similar to the pattern of cause/effect, writers utilizing a problem/solution text pattern describe a specific problem or series of problems, then propose a solution, which is a plausible, salutary effect on a course of action.

In eastern forests of Canada and the United States bacterial insecticides may be one important answer to the problems of such forest insects as the budworm and the gypsy moth. In 1960 both countries began field tests with a commercial preparation of *Bacillus thuringiensis*. Some of the early results have been encouraging. In Vermont, for example, the end results of bacterial control were as good as those obtained with DDT. The main technical problem now is to find a carrying solution that will stick the bacterial spores to the needles of the evergreens. On crops this is not a problem – even a dust can be used. Bacterial insecticides have already been tried on a wide variety of vegetables, especially in California.

(Carson, 1962, p. 255)

Analysis – Although many of the previously described text patterns employ an analysis of concepts, this text pattern provides a specific critical review of the facets of a situation, event, idea, or case. Writers utilizing a predominant text pattern of analysis provide an in-depth coverage of specific aspects of a general topic. When writers also provide this analysis in light of a framework of criteria, then the writer employs evidence for justifying a point of view.

In our own American experience, families used to live closer together, making it easier for relatives to pitch in during pregnancy and the first months of a newborn's life. Women worked primarily in the home and were more available to lend a hand to new mothers and to help them get accustomed to motherhood. Families were larger, and older children were expected to aid in caring for younger siblings, a role that prepared them for their future parenting roles.

(Clinton, 1996, p. 70)

So state and nation may often fail to coincide. Yet in ordinary English we use them almost synonymously (along with “country” which probably should really refer only to physical terrain and landscape). And we regard instances where they do not coincide, like those described above, as anomalous. How did a legal concept and a social concept, which certainly do not automatically coincide, come to be so closely associated in our minds?

The reason is that even if a government is generally accepted as appropriate to rule a state, it still faces the difficult task of ensuring people's acquiescence in its laws. It is terribly expensive and counterproductive to enforce all laws, all the time by overt force. Dwight Eisenhower could not simultaneously send troops to all schools that were slow to desegregate! Rather, governments must persuade, cajole, and only to a certain extent force people to obey the laws. Beyond even this, governments generally want the people of a state to be more than just passively obedient to the laws. They want them to contribute positively to the state, as voters, soldiers, volunteer workers in cooperative enterprises. As the modern state began to emerge in the eighteenth and nineteenth centuries, the leaders of states discovered that they could develop a more enthusiastic citizenry if they linked national identity to the state's boundaries. This meant that when the government asked citizens to do something on behalf of the state (obey public health laws, pay taxes, join the army), it was asking not just on its own behalf but on behalf of the nation of which all felt themselves a part.

(Shively, 1997, pp. 120-1)

Evidence about the role that top-level organizing structures play in expository prose is inconsistent. While most results in this line of research show some text-structure pattern as being more difficult for readers, the patterns tend to vary from study to study (e.g., Carrell, 1992; Freedle & Kostin, 1993; Horowitz, 1987; Meyer, 1987). Most studies do agree, however, that a simple description of facts seems to be easiest to comprehend. Any synthesis of this research area would need to account for L1/L2 differences in results.

We have recommended certain pragmatic and rhetorical features, given the test purpose and the organizational framework we propose for the reading test. At this time we do not believe the alternative types of text that we propose are likely to contribute to differences in difficulty of the items associated with them. However, we think it is important for the sake of content representativeness and content comparability to investigate ways in which pragmatic and rhetorical features of parallel tests can be accounted for in test specifications.

Test Rubric. The matrix in Table 1 illustrates the relationships among types of text, reader tasks, and reader purpose. Part A identifies the types of text we recommend for inclusion in the TOEFL 2000 reading test defined in terms of the various pragmatic intentions and rhetorical patterns. Part B identifies various tasks that a reader may perform in engaging an author's writing.

The readers' tasks are related to the four reading purposes that we have chosen to organize the domain. Thus, the task of summarizing may be most compatible with the purpose of reading to learn, the task of comparing/contrasting might most usefully be related to the purpose of reading to integrate information across multiple texts, and so on. The reason for using these combinations of reader tasks/types of texts is that they cover the major domain of reading as it is practiced in formal college and university settings.

Given the various possible combinations between A and B in Table 1, we recommend that these combinations be used to define the domain of reading for the TOEFL 2000 test. For example, for reading to find information and reading for basic comprehension, readers would be required to identify information and interpret texts (and, in some instances, non-prose text information), characterized by various rhetorical patterns: define/describe/elaborate, compare/contrast, present a problem and a solution, explain/justify, persuade, and narrate. For reading to learn, readers might be required to summarize or define/describe/elaborate information based on an author's attempt to define, describe, etc. However, we also recognize that purpose and reader task will require some flexibility; thus, reading purposes (in relation to task) are separated by dashed lines.

We believe that these combinations of types of texts and tasks mesh nicely with the framework we proposed in the previous section, i.e., the four reader purposes applicable in the academic context: (a) reading to find information; (b) reading for basic comprehension; (c) reading to learn; and (d) reading to integrate information across multiple texts.

Table 1
Types of Texts and Tasks for the TOEFL 2000 Reading Test

B. Reader tasks	A. Types of texts defined by pragmatic and rhetorical features					
	EXPOSITION			ARGUMENTATION PERSUASION EVALUATION		HISTORICAL NARRATION
	define/ describe/ elaborate/ illustrate	compare/ contrast/ classify	problem/ solution	explain/ justify	persuade	narrate
identify/ interpret	Reader Purpose: Reading to Find Information and Reading for Basic Comprehension					
summarize	Reader Purpose: Reading to Learn					
define/ describe/ elaborate/ illustrate						
compare/ contrast/ classify						
problem/ solution	Reader Purpose: Reading to Integrate					
explain/ justify	Reader Purpose: Reading to Integrate					
persuade						
narrate						

Tasks to Assess the Four Types of Reading

The following sections describe the types of tasks that could be used in the TOEFL 2000 test to assess the four purposes for reading: reading to find information, reading for basic comprehension, reading to learn, and reading to integrate information across multiple texts.

Reading to find information. At its simplest level, reading to find information involves the process of locating the information that is noted in a question and matching it to identical or closely paraphrased corresponding information in a text. For example, a question might ask an examinee to simply circle a particular word in a text or to circle every occurrence of a particular word or phrase. It is unlikely that the TOEFL 2000 test would measure this type of reading purpose in isolation, however, because the task and level of difficulty would be inappropriate for the test purpose and the task would be unlikely to discriminate among TOEFL examinees unless time is constrained so that automatic processes have to be engaged.

At present, this reading purpose is assessed in combination with reading for basic comprehension. For example, in the question “Why did the Census Bureau revise the definition of ‘urban’ in 1950?” the information from the text that needs to be located in order to answer the question is: “. . . in 1950 the Census Bureau radically changed its definition of ‘urban’ . . .” Locating this information would require the reading to find information purpose, and answering the question would require reading for basic comprehension, which is discussed in more detail below. Some of the current TOEFL reading tasks weigh more heavily on search processes while some involve other comprehension processes to a greater extent.

Another type of reading to find information task might be the question type that asks examinees to locate where in the text a particular type of information can be found. “Where-stated” tasks might be classified as reading to find information tasks or reading for basic comprehension tasks. Ultimately research will determine how particular tasks can best be combined with reading purposes so as to produce the most useful information.

In the TOEFL 2000 test, reading to find information might also require skimming and scanning a prose text or a non-prose document to find a discrete piece of information that can be easily matched to the information requested through a literal match or a close paraphrase. An example of this type of reading to find information might be: “Highlight the information in the bar graph that names the era in which dinosaurs lived,” or a question requiring examinees to locate the particular section of a longer text where a specific type of information would be found or to scan multiple texts for a particular type of information.

Reading for basic comprehension. While reading to find information is an important underlying skill of reading, and might sometimes be tested directly as an independent reading skill, it will certainly be measured indirectly in connection with the second type of reading purpose on the TOEFL 2000 reading test, i.e., reading for basic comprehension. As defined for the purposes of this test, this type of reading requires readers not only to locate information in the text by matching information from the question to the text but also to identify additional new

information in the text that answers the question. For example, in the illustrative question earlier, after examinees have located the place in the text where the Census Bureau is mentioned in conjunction with the year 1950, they still need to understand why the Census Bureau revised the definition of “urban.” Questions testing this type of reading involve both the reader’s knowledge of vocabulary and cohesion/coherence devices and the ability to identify and interpret facts.

The current TOEFL reading test has been extensively analyzed in terms of difficulty variables (Freedle & Kostin, 1993; Sheehan, Ginther, and Schedl, in press). A preliminary analysis of the reading comprehension tasks based on three variables identified by Kirsch and Mosenthal (1990), — type of information, type of match, and plausibility of distractors¹ — indicates that many of the variables that have been useful in predicting difficulty on other reading comprehension tasks will also be useful for describing TOEFL assessment tasks.

In designing new reading to find information tasks and basic comprehension tasks for the TOEFL 2000 test, we recommend that the current paper-and-pencil TOEFL reading texts and tasks be expanded to include not only prose texts but also non-prose texts such as pictures, diagrams, process schematics, procedural schematics, matrix documents, locative documents (including general reference and topographic maps), and quantitative documents (e.g., pie charts, bar charts, line graphs, and timelines). Moreover, in addition to multiple-choice items, new forms of response modes could be used. These might include: (a) open-ended responding with words, phrases, or sentences; (b) point-and-click on the text, and (c) point-click-and-drag information from one part of the screen to another. We would also recommend that a broader range of rhetorical patterns be included. In addition to texts that define, describe, elaborate, illustrate, explain, and justify (as current TOEFL texts do), we recommend including texts that compare and contrast, persuade, and narrate. These recommendations would be in line with making the TOEFL 2000 reading test more similar to the conditions under which college and graduate students read in formal academic settings.

Reading to learn. Reading to learn requires readers to integrate and connect detailed information from the text in a manner that is consistent with the rhetorical pattern of the text. Here we are distinguishing the comprehension of individual main ideas or main points presented in a text, which would be included under the “basic comprehension” purpose and the comprehension of the whole text in the reading to learn purpose, which requires the integration of these ideas into a coherent framework.

From a task perspective, item types that assess reading to learn would require examinees to understand the rhetorical pattern of the text as well as to integrate the content information. For this reading purpose, the issue of text material or text selection becomes especially important because the type of text is more intimately connected to the information presented. It seems likely that the types of texts that are appropriate for this purpose and the types of tasks associated with them will increase the importance of pragmatic and rhetorical features and thereby increase the likelihood that text readability will be associated with item difficulty.

¹ See references in *TOEFL 2000 Framework: A Working Paper* (Jamieson et al., 1999) for a complete list of relevant research.

We see two constructed-response alternatives, or two possible ways, for operationalizing reading to learn tasks. One option is to set a time limit for reading a text, then take the text away and ask examinees to summarize it in order to measure how well they have actually “learned.” This would allow us to include reading rate as a variable related to difficulty. We would want to prototype and research the advantages and disadvantages of examinees being allowed to take notes during the reading period that they could then use in writing the summary. The other possibility would not involve taking the text away but would still require examinees to integrate information across the entire text and evolve a framework for interpretation.

We are not ruling out the possibility that reading to learn could be tested in short-answer, semi-productive, or multiple-choice formats. Research should investigate a number of possible item types. These alternatives need to be explored and research investigating different ways of presenting the prompt needs to be carried out before decisions can be made about the new item types. The research should also look at different language groups to determine whether both alternatives provide equally good information.

Research on types of texts chosen for reading to learn tasks is also needed. We would like to assess whether examinees recognize an author’s rhetorical pattern and can integrate the text information in terms of it. It would be useful to have research that looks at texts varying in terms of how strongly the organizational pattern or classification categories are signaled, i.e., ranging from texts where all of the organizational categories are provided by the author to texts where the organizational pattern is weakly signaled. The same texts could be used and the amount of signaling varied to determine how much this changes the difficulty or the feasibility of the task. We would also like to see research with texts which vary in terms of the types of texts outlined in Table 1.

Finally, we are assuming that passages used for reading to learn tasks that require productive spoken or written responses would also be used for other types of multiple-choice or short-answer tasks so that the test time is used most efficiently. While the reading to learn purpose could be tested in combination with other modalities, it may also be possible to test it alone. We recommend that reading to learn tasks be associated with single passage texts and that multiple texts be reserved for reading to integrate across multiple texts tasks.

Although the following example represents a task that would be scored as an integrated-skills task, we do not mean to suggest that reading to learn tasks have to be connected to other skills. It would be appropriate to test reading to learn as a measure of reading comprehension whether or not it is integrated with other language skills. Research should investigate the feasibility and construct relevance of a number of alternatives.

That said, one way to measure reading to learn might be to require readers to summarize and/or recall text information. An example of a reading to learn task at this level involving a comparison/contrast might be the following text, taken from a history book:

Compare the ideas of Locke and Hobbes. In what ways were Locke's beliefs influenced by Hobbes' thinking? In what ways did their thinking differ?

The ideas of two English philosophers, Thomas Hobbes and John Locke, changed the way Europeans viewed the individual's role in society. During the 1640s, Hobbes witnessed the upheaval of a civil war in England. As a result, he became convinced that if people were left alone without government they would constantly fight among themselves. In 1651, he published his ideas in *Leviathan*. Hobbes described life in a state of nature in which people had no government. Such a life, he claimed, would be "nasty, brutish, and short."

According to Hobbes, to escape the chaos of their natural state, people entered into a contract in which they agreed to give up their freedom to a ruler who guaranteed peace and order. The best government, Hobbes said, was one in which the ruler had absolute power. He insisted that once people entered into such a contract, they could not rebel, even if they thought the ruler was a tyrant. Hobbes' ideas, therefore, supported the rule of absolute monarchs.

In 1690, John Locke published *Two Treatises on Government* in which he agreed with Hobbes that the purpose of government was to create order in society. He also saw government as a contract between the ruler and the ruled. However, Locke's other ideas about government differed greatly from those of Hobbes.

Locke had a more optimistic view of human nature than Hobbes did. He thought people were basically reasonable and would cooperate with one another. Moreover, Locke argued that rulers could stay in power only as long as they had the consent of those they governed. If a ruler were a tyrant, he or she had broken the contract and the people then had the right to rebel.

Locke also presented other ideas that were important in the development of democracy. He believed people had natural rights, including the right to life, liberty, and property. Government was responsible for protecting these rights, he said, but its power should be limited.

(Beers, 1983)

Examinees might be given such a task after having already answered several selected response questions. They might be asked to read the question and take a minute to reread the text before answering the question. Alternately, they might be asked to respond when the text is no longer available. In this case they might be provided with several key points to address in their response to help them remember the text content. For example, in the case of this particular text, they might be asked to compare Hobbes' and Locke's respective views on human nature and human rights and their views on the role of government and the role of rulers.

It is important to note here that in suggesting one possible task and sample text we do not mean to preclude other possibilities. We recommend that extensive research be carried out, particularly in the areas of reading to learn and reading to integrate multiple texts, before decisions are made about operational test design. Reading to learn tasks could be designed as a measure of reading comprehension alone, and reading to learn could also be associated with spoken responses. Research should investigate all possibilities.

Research also needs to be carried out on scoring criteria and equatability of tasks across the various types of texts listed in Table 1.

Reading to integrate information across multiple texts. Reading to integrate requires readers to generate appropriate organizational frameworks which can then be used to relate information from two different passages (or information sources). This integration may require readers to compare/contrast, identify a problem and a solution, explain/justify, persuade, or narrate. The selection of text material will be most critical in establishing the reading to integrate purpose. Careful consideration will need to be given to the pragmatic and rhetorical features of texts used in developing tasks for this purpose. In addition, multiple texts can be combined in a variety of ways: a text passage may be followed by a diagram, a diagram might be followed by a text passage, a text passage might be followed by a map, an outline might be followed by a figure, etc.

Reading to integrate tasks may also make use of language input from other modalities, so that texts may be either spoken or written. From a task perspective, items might ask examinees to read a longer text and summarize it in writing or speaking, or to read a text and complete diagrammatic information (or create a diagram), or listen to one text and read a second text and then compare and contrast them. Multiple texts for reading to integrate tasks might also consist of a prose text and a non-prose document. Research is recommended in order to explore variables related to reading documents.

As for texts used with productive reading to learn tasks, texts used for the reading to integrate purpose could also be used to test other reading purposes with multiple-choice or short-answer responses.

The following is an example of a reading to integrate question that combines a prose text and non-prose text:

Read the two texts below. Text 1 is a prose passage, and text 2 is a bar graph. First, write a brief title for the bar graph. Next, write a summary of the main information in the bar graph and explain how it is related to the information in the passage.

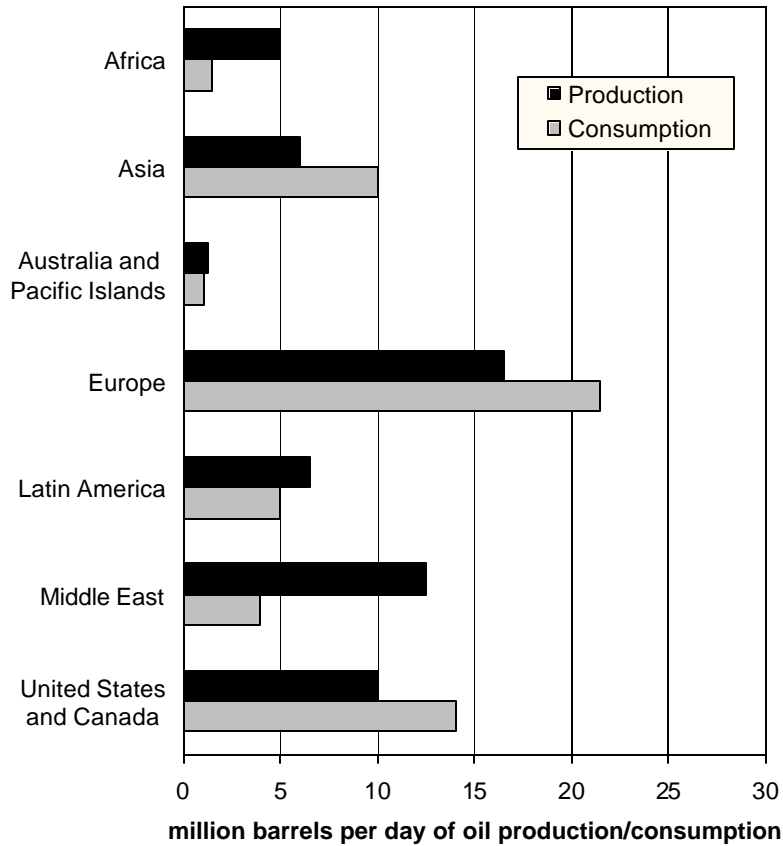
Dimensions of the Energy Problem

Strictly speaking, no energy problem exists. The basic laws of physics dictate that energy is conserved and can only be changed from one form to another or into matter. Fuel, on the other hand, is the accumulation of matter and therefore represents a store of energy. This energy is released in the form of heat when the fuel is burned in chemical or nuclear reactions, which cannot be reversed to regenerate the original fuel mass (at least not without the injection of more energy than was originally released).

As a consequence, a fuel problem does exist. If the supply of fuel is finite, not only will there be no energy supply when the fuel is exhausted, but also all other processes that depend on it will cease. This will affect not only the obvious energy consumers in the United States and the rest of the industrial West, but even the most primitive societies, where the importance of oil-based fertilizer supplies is growing.

Several factors combine to make the problem an urgent one. World population is steadily increasing, which implies that the demand for energy will also increase, although not necessarily in proportion. Social, economic, and political pressure for economic expansion continues in industrialized countries. This implies an increased energy input. The developing countries are becoming aware that their economic position could be improved by increased energy consumption, and they feel entitled to a larger share of the world's energy resources than they now receive. These pressures require that the world energy supply be increased, particularly if the aspirations of some areas are to be met without jeopardizing the living standards of others. Finally, it is now recognized that the supply of the conventional fuels – coal, oil (petroleum), natural gas, uranium, and fuel wood – is limited and insufficient to sustain present rates of development for much longer. Although there may be debate about the exact length of time available before the effects of a worldwide shortage become apparent, there is agreement that such a shortage will occur. It is only a matter of time; in the case of oil, for example, the debate is not about whether, but about when oil production will peak.

(Grolier^a Online Encyclopedia, 1995)



As in the case with reading-to-learn tasks, research will need to be carried out on task comparability and scoring criteria.

Types of Response Formats

In a computer-based test, it is possible to include a much greater variety of response formats than is possible in a paper-and-pencil exam. We recommend that the following response formats be considered for use in the TOEFL 2000 reading test and integrated tasks:

- multiple choice
- open response formats
- click on a word, phrase, or sentence in the text or graphic
- click on and drag a word, phrase, or sentence in the text or graphic
- complete a chart, graph, or table
- create a chart, graph, or table
- extended written and/or spoken response

Tasks using different types of response formats can be used to assess examinees' abilities with respect to the four purposes for reading, and some response formats may be better suited than others to particular measurement objectives.

In redesigning the TOEFL reading test so that it includes measures of reading to find discrete information, reading for basic comprehension, reading to learn, and reading to integrate information across texts, we suggest that readers first be presented with one or more passages and asked a series of reading to find information and reading for basic comprehension questions representing a range of difficulty.

We suggest that these initial tasks include multiple-choice as well as open-ended machine-scorable formats. They might include many of the formats that are used on the current paper-and-pencil TOEFL reading test, as well as new computer-based TOEFL test formats such as the insertion task format.

Once readers have completed the reading to find information and reading for basic comprehension tasks, they could be given one or more reading to learn tasks that might involve additional types of response formats. Reading to learn tasks might ask readers to organize information from the text in a table or a schematic representing a rhetorical pattern.

To measure reading to integrate skills, the system might then present an icon which readers would click on to see a second passage that would address the topic of the first passage but in a different manner. At this point, readers could be asked to relate information in the first passage to information in the second passage.

As with reading to learn tasks, readers' performance on reading to integrate tasks would be evaluated in terms of criteria that would permit exchangeability of scores across different tasks varying in content, type of text, and reader purpose.

Linguistic Variables and Task Difficulty

Tasks for testing these various purposes for reading can be made more or less difficult by varying certain linguistic parameters. We believe the following are an important subset:

- vocabulary,
- syntactic complexity,
- transition markers (cohesion),
- antecedent reference,
- modality (adverbs of attitude),
- amount of text, amount of time allowed,
- distances across text when cycle or integration is involved,
- competing linguistic distractors in the text environment,
- cohesion determiners (e.g., "We bought a camera. The lens was cracked."), grammatical relations as referents (back to subject or back to object), and
- cohesion.

The linguistic variables are relevant to difficulty, however, only insofar as the task makes them relevant, i.e., if a simple locating task is chosen to test a question of fact, the possible impact of most of the linguistic variables is minimized. On the other hand, a task testing reading to integrate information across multiple texts might be influenced by most of these variables.

4. Technological Considerations

The Role of Technology in Reading Comprehension

As a computer-based instrument, the TOEFL 2000 test will allow a greater range of alternatives for measuring reading comprehension, particularly since the construct of reading as defined in this paper encompasses the processing of both text and document materials inclusive of electronic formats. As a construct, communicative competence in the 21st century involves the processing of texts through the medium of computer technology. A broad definition of reading, as well as writing, therefore includes the component of media literacy, which accounts for the many visual and electronic forms of text that are present in our contemporary world. In fact, the International Reading Association and the National Council of Teachers of English (NCTE & IRA, 1996) recently included media literacy – specifically, “viewing” and “representing language” – with the more traditional modalities of reading, writing, listening, and speaking in their definition of the English language arts.

In the description of the construct of reading presented in this paper, each of the four reading purposes aptly applies to electronic environments as well as to hard copy texts. In considering reading in electronic environments, a particularly relevant dimension of the construct of reading comprehension applies to the description of reading to integrate information across texts, since this type of reading is defined through the use of multiple texts. When working online, readers often use multiple screens of information, use texts plus graphics, and work with both electronic and hard copy formats. Therefore, when working with multiple texts, readers need to be able to infer which information generalizes across texts, which information from one text helps elaborate or explain information from another text, and which sections of texts correspond to graphics on the screen. An example of an authentic reading task would be reading several different online texts in order to research a topic for a college paper.

In academic environments, information is disseminated in both paper and electronic formats, and the recent reliance on the use of technology for both teaching and learning is increasing the need for students to become proficient users of computer technology in accessing and reading information. As information becomes more prevalent in electronic environments, users will have greater choice in accessing texts, journals, and documents. In addition, as academic institutions increasingly use computer technology to deliver instruction and for course requirements, students will be expected to learn and be able to use a greater variety of computer environments. Examples already include distance learning, online student registration, specialized Web sites for programs and courses, e-mail for communicating with professors and students, and full libraries of online information.

In many college settings, professors have their own Web sites, and students are expected to be able to access course syllabi, course materials for lectures and activities, and other pertinent course materials. Hansen and Willut (1998) provide an overview of the technologies that students will be expected to use in the year 2000; however, their estimates of the use of technology conservatively project that “courses using computer-based approaches to deliver the core educational experiences will be in the minority in the year 2000” (p. 13). Recent accreditation standards for state, regional, and national agencies now require the infusion of technology in all coursework and require student proficiencies in computer technology.

Access to specialized Web sites, online libraries, bulletin boards, and e-mail provide reading environments quite different from the environments of hard copy text and require a variety of user proficiencies unique to the computer world. For instance, users need to be familiar with how to read and use menus, icons, and navigational aids on the screen and need to have basic proficiency in typing on a keyboard. Future computer designs that include voice recognition may reduce the need for the user to have such computer proficiencies. Touch screens that eliminate the need for using a cursor to make a selection may also reduce the need for users to be able to interact through a keyboard or mouse.

The Reading Comprehension Interface

For the reading comprehension portion of the TOEFL 2000 test, we strongly recommend that a reader-friendly interface be developed. In addition to single passage texts, the reading test is likely to include multiple prose texts and texts integrated with pictures, maps, charts, graphs, diagrams, and video. Because the experience of reading online is generally less appealing than that of reading paper copy, every effort should be made to investigate and create the most user-friendly interface and presentation design possible. The types of online information we would like to draw on for the TOEFL 2000 reading test are summarized as follows:

- Single text – a short paragraph or a longer text;
- Multiple texts – short and/or longer paragraphs;
- Single text plus graphic – a short or longer paragraph accompanied by a graphic; and
- Multiple texts plus graphic – short or longer paragraphs accompanied by a graphic.

The types of graphics that could be utilized are as follows:

- Drawing – line drawing or schematic;
- Picture – picture to illustrate a text passage;
- Map – geographical, population, weather, or other type of map;
- Chart, table, or graph – chart, table, or graph that illustrates data; and
- Video – (if the technology supports such a capability) – video is a recommended source, particularly when the video directly relates to the topics in a hard copy text.

In particular, we recommend the inclusion of navigational aids and icons.

Due to specific factors that influence the reading process, whether the task is in a hard copy or electronic environment, the interface for the reading portion of the TOEFL 2000 test should have the following design features:

- Clarity in resolution – The best state-of-the-art resolution features should be employed.
- Full text viewing – Examinees should be able to choose an icon that allows them to view the entire text.

-
- Scrolling of text – So that test takers are able to access all the text, a scroll feature needs to be included when the text is long.
 - Magnification of the text – Examinees should be able to control a zoom feature to magnify the entire text or a portion of a lengthy text.
 - Speed of access – The TOEFL 2000 test should use state-of-the-art computers that respond quickly when the user clicks on an option.
 - Multiple text viewing – When readers are presented with multiple texts and/or graphics, they should be able to switch from one text to the other very quickly and efficiently so that the multiple texts are displayed without disrupting comprehension.

5. Research Agenda

The goals of the research agenda for the TOEFL 2000 test are to develop a test that is not only reliable but also valid and fair. The design of a new test offers opportunities for integrating test design and research that will contribute greatly to the validity of the test.

Test validation has been conceptualized in many ways, but at its heart it can be seen as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment” (Messick, 1989, p. 13). Validity is a multifaceted concept. Messick (1995) describes six aspects of validity – content, substantive, structural, generalizability, external, and consequential – that can serve as criteria or standards for measurement. Some of these aspects reflect traditional concerns such as whether content is relevant to and representative of the domain, the generalizability of scores over tasks and across time and among raters, and the relationship of scores to external measures and behaviors that are or are not construct relevant. Other aspects, particularly substance and consequence, have received greater emphasis in recent years than in the past.

The substantive aspect of validity concerns construct representation (Embretson, 1983), a theoretical description of the information processes, strategies, and knowledge stores that underlie task performance. With respect to construct representation, assessments can be described from either the task perspective (What are the features of the task?) or the examinee perspective (What processes, skills, strategies and knowledge do people use to solve problems?). Tasks, of course, can be described in many different ways. By introducing a criterion, the relationship between task features and difficulty, a distinction between critical and incidental task features can be supported. The substantive aspect of construct validity is central to the TOEFL 2000 framework because we expect to base test design on an understanding of task difficulty. A basic assumption of this approach, one that needs to be substantiated through research, is that the features of the task and examinee processes are interdependent (Chapelle, Grabe, & Berns, 1997). It should be noted that construct representation has two distinct meanings. One, representativeness, relates to coverage of a domain: Does the assessment represent a reasonable selection of content and processes typical of the domain? The other, representation, refers to a psychological model of the task in terms of features, knowledge, and processes involved in completing the task (Messick, 1995).

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

Because there are so many facets to validation, it is not an activity that should occur at the end of the test design process. It should be an integral part of the process. The decisions that are made in designing a test and the basis for these decisions all become evidence relevant to test validity. Nevertheless, there are sequential constraints that need to be taken into account as to the kinds of validity evidence that can be collected at various stages of test design.

The sequential nature of the test design process offers an organizing frame for laying out the issues that can be addressed at each stage as well as one of a number of criteria for prioritizing research. The stages of test design might be characterized as construct identification, prototyping, pilot testing, and field testing. A brief description of each of these stages follows:

1. Construct identification – The construct to be assessed should be articulated, and the ways that the construct might be operationalized should be proposed.
2. Prototyping – Prototypes of possible tasks and associated scoring rubrics should be developed and evaluated. At the same time, research on tools to support the development of tasks can be initiated, and preliminary consideration should be given to psychometric issues.
3. Pilot testing – Those tasks that are thought to have potential for operational use are pilot tested on small samples to identify potential problems and to clarify their relationship to the constructs that are the objectives of the assessment. The outcome of this phase should be preliminary specifications for task types and the manner in which test sections might be assembled for testing in the next phase.
4. Field testing – Tasks that are deemed to be strong contenders for operational use are assembled into possible operational sections and pretested on samples large enough to support rigorous psychometric evaluation. The function of this stage is to finalize the blueprint for item development and test assembly for operational pretests.

At each of these stages, operational, psychometric, and construct-related issues can be addressed. Operational concerns focus on the feasibility of the proposed assessment. Questions such as what kinds of hardware and software are needed, how much time and effort are required to develop and score tasks, how much time the assessment requires, and whether the cost of the assessment is reasonable, need to be answered. Psychometric issues include evaluating the statistical properties of tasks and the appropriateness of different methods of scaling tasks and generating scores. Finally, construct validation involves documenting that the tasks assess construct-relevant rather than construct-irrelevant processes and skills.

Although the above framework suggests sequential stages, the iterative nature of the test design process should be emphasized. For example, one would expect that the results of pilot testing might dictate revisions to prototyped tasks. At any point in the design process, however, the specific questions that can be answered are constrained by the number of task exemplars that exist and the number of participants we want to involve in the evaluation.

With respect to the research agenda for the TOEFL 2000 reading comprehension test, some of the questions that need to be answered at each stage of the design process are discussed below. While some of the issues described are specific to reading assessment, others are more general and also apply to the other modalities or to integrated tasks.

Construct Identification

The goal of this stage is to produce documents that describe the constructs to be assessed, indicate how they might be operationalized, and clarify the expected uses of the test. Although the research that can be initiated at this stage is constrained because examples of potential tasks do not exist, important topics that can be investigated include the following:

1. Literature reviews of theoretical descriptions of competency and empirical research supporting these views should provide a basis for articulating the constructs to be assessed.
2. Research on how the test is currently used and the characteristics of the population taking the test would be useful for documenting the appropriateness of the test for the kinds of inferences users wish to make, for planning score reports, and for later evaluation of the consequential impact of test revisions. For example, one question that could be addressed is, what strengths and weaknesses do L2 students have, and how do these relate to success or failure in academic settings?
3. Many research issues about the characteristics of texts for reading assessments have been identified. These issues include identifying the range of text features characteristic of academic texts, assessing the contribution of various features to text difficulty, providing empirical support for the test blueprint, and developing tools to support efficient item development. Research on the above topics would be facilitated by a systematic collection of samples of text materials used by students in broad major fields and at various types of institutions or identification of an existing corpus that could be used to answer a number of questions about the range and characteristics of materials that students encounter. The need for the following research activities is anticipated:
 - a) Analyze the rhetorical and linguistic features of authentic texts typical of the domain of academic reading. This will provide information about the range of registers most commonly involved in academic discourse. (See, for example, Biber, 1986.) This information is needed to guide the selection of types of texts that will provide appropriate content representativeness for the TOEFL 2000 test.
 - b) Choose a system for categorizing texts. An empirically supported multidimensional system would be desirable.
 - c) Develop alternative measures of text complexity based on syntactic complexity and/or informational density. While readability features of texts used in reading tests in the past have not been shown to be directly related to item difficulty, it seems to us to be possible that reading text characteristics may be more directly related to task difficulty for the new types of texts and tasks we have proposed for the TOEFL 2000 reading test.

-
- d) Explore whether natural language processing (NLP) tools can be developed to assess text difficulty and to tag text features for test development.
 - e) Determine what other aspects of vocabulary in addition to word frequency contribute to difficulty.
 - f) Analyze the kinds of documents that students must understand and their co-occurrence patterns with different types of texts.
 - g) Determine how various types of signaling influence the comprehension of text structure.
4. Research on the variables controlling the difficulty of current TOEFL tasks should be initiated and research on the variables controlling the difficulty of potential reading-to-learn and reading-to-integrate tasks should be investigated as prototype items are developed.
 5. Investigate the effects of controlling the reading time allowed and/or measuring reading rate.

Prototyping

Exemplars of potential item types and scoring systems will be developed and authored for computer administration. Alternative versions of tasks may need to be developed so that questions about which formats best assess relevant constructs can be answered in subsequent phases. The following research and development activities are needed.

1. Explore how to operationalize the new “Reading to Learn” and “Reading to Integrate” constructs proposed in this reading assessment framework. Develop task exemplars that vary the nature of the prompts, stimulus exposure time, memory demands of the task, and the response formats.
2. Propose scoring systems for the tasks.
3. To ensure optimal screen design, summarize prior research on the impact of presentation features on reading comprehension and consider this research in designing the interface.
4. Develop tutorials instructing examinees how to complete the prototype tasks.
5. As exemplars are developed, conduct preliminary observational studies of small groups of L1 and L2 students to evaluate the tutorial and interface.

Pilot Testing

A small number of exemplars for each prototype task will be developed and pilot tested on small samples of examinees to evaluate which tasks are most suitable for the population and which conditions of administration are appropriate. With respect to the pilot testing, the following activities will be necessary.

1. Develop and evaluate scoring systems for constructed response items.
2. In addition to conducting a preliminary evaluation of the statistical properties of the item types, initiate construct validation research addressing a number of issues. This research is needed to demonstrate that the skills and abilities that account for performance on the test are related to the construct being assessed and not to other sources of individual differences. The outcome of this phase should be tentative or alternative blueprints for the assessment. The types of investigations that can be initiated at this point include studies of:
 - a) User acceptance – Do the constituents who will use the test results believe that the tasks are appropriate?
 - b) Concurrent validity – Is performance on the tasks related to other, concurrent indicators of linguistic competence such as placement in remedial classes, need for tutoring, teacher evaluations, or performance on more complex criterion tasks?
 - c) Construct representation – Is there evidence that the processes and strategies that test takers use to answer questions are construct relevant?
 - d) Impact of test-taker characteristics – Some characteristics, such as how long test takers have studied English or how long they have been in the United States, are construct relevant. Others, such as computer experience, are not. What is the impact of construct relevant and irrelevant factors on performance?
 - e) Native speakers – How do they perform on the tasks?
 - f) Factors that affect task difficulty – In particular, what is the nature of the relationship between reader purpose and task difficulty?

Field Testing

Test sections composed of task types that have been found to have strong potential for inclusion on the new test would be pretested on large samples in operational-like settings to determine statistical properties, to implement and evaluate scoring systems, and to evaluate psychometric models. Topics and questions that must be researched are summarized below.

1. Ways of enhancing scale interpretability should be explored. These might include:

-
- a) scale descriptions linked to proficiency descriptions of tasks; and
 - b) normative information about performance by native speakers applying to different types of programs (two-year colleges, four-year colleges, graduate schools; sciences, social sciences, humanities programs). The latter approach also has implications for test fairness in that it would provide information about the range of performance by native speakers.
2. Additional construct validation research that requires larger samples could be carried out. This includes studies of:
- a) Convergent and discriminant validity – Is performance on the tasks more highly correlated with performance on other tasks that are thought to assess the same construct and less correlated with performance on tasks thought to assess other constructs?
 - b) Construct representation – For example, is there evidence that the difficulty of the tasks can be manipulated in systematic ways that are relevant to the construct?
 - c) Experience – Because it is expected that short-term experience will have an impact on this type of test, it would be useful to document how L2 mastery, as measured by the potential item types, changes over time and what factors facilitate or inhibit development.
 - d) Subpopulation differences – Studies should be carried out to determine if there are sources of construct-irrelevant variance, such as gender, affecting performance on the tasks.
3. The appropriateness of different psychometric models for scoring and scaling items will need to be evaluated.

Further Research

The outcome of the test design process will be a blueprint for test development that will specify in detail the numbers and types of tasks to be developed, the kinds of materials to be used, the task features that should be manipulated to control task difficulty, the composition of test sections, and scoring rubrics.

Many of the issues addressed in previous stages will need further exploration or replication under operational conditions. In addition, there are some questions, such as the consequential effects of the test, that cannot be answered until the test has been operational for a period of time.

6. A Better Reading Test

The existing TOEFL reading test is a good test by current standards. Designing a new test that will constitute a significant improvement over the current one is therefore challenging. The current TOEFL reading test has never been explicitly linked to any particular theory of reading or reading construct, however. The most obvious improvement we can make over the current test, then, is to articulate the construct we want to test and to link the proposed test design to that construct. We believe that the proposed reader-purpose framework will allow for better communication of the principles driving test design and test development, while at the same time make it possible to overlay complementary frameworks driven by processing and task-based views of reading.

Moreover, we believe that the reader-purpose framework expands the construct tested beyond the level of finding discrete information and basic comprehension measured in the current TOEFL reading test. It seems likely that tasks and texts intended to measure reading to learn and reading to integrate information across texts will also tap into the rhetorical and discourse structure of the texts in ways that make text readability more important than it is for reading to locate information and reading for basic comprehension. In addition, the new types of texts and tasks associated with reading to learn and reading to integrate information across texts are likely to raise the ceiling of the test to allow us to discriminate better at higher proficiency levels.

One of the goals of the TOEFL 2000 project is to identify variables contributing to task difficulty in order to improve the interpretability of test scores. In other words, we hope to provide some information about what examinees can and cannot do with the English language rather than merely report what score they received on the TOEFL reading scale. Research can be designed to confirm whether variables thought to influence task difficulty actually do so. This will then allow us to provide useful information to test users and examinees about the strengths and weaknesses of an examinee's performance. It will also provide critical information to test designers and test developers that will allow them to make the most efficient use of examinee time and of items and to create particular types of items to target certain proficiencies and goals. For example, we hope to increase the importance of variables associated with the type of information requested and to decrease the weight of variables associated with the plausibility of the distractors, thus increasing construct validity.

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

Finally, the proposed design is flexible enough that it does not preclude the introduction of additional new item types at some future time. We believe that new item types would be likely to include components that contribute to task difficulty in similar ways to those identified at the time the TOEFL 2000 test is introduced. Identifying and focusing on variables that contribute to difficulty in the design of the new test thus facilitates the process of updating the test from time to time.

References

- Adams, M. J. (1990). *Beginning to read*. Cambridge, MA: The MIT Press.
- Akamatsu, N. (in press). The effects of first language orthographic features on word recognition processing in English as a second language. *Reading and Writing: An Interdisciplinary Journal*.
- Axelrod, R., & Cooper, C. R. (1996). *The St. Martin's guide to writing*. New York: St. Martin's Press.
- Barker, T. A., Torgesen, J. K., & Wagner, R. K. (1992). The role of orthographic processing skills on five different reading tasks. *Reading Research Quarterly*, 27, 346-345.
- Baron, R. A. (1992). *Psychology*. Needham Heights, MA: Simon & Schuster.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney and E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3-73). Cambridge: Cambridge University Press.
- Beers, B. I. (1983). *World history: patterns of civilization*. Englewood Cliffs NJ: Prentice Hall.
- Berman, R. (1986). A cross-linguistic perspective: Morphology and syntax. In P. Fletcher & M. Garman (Eds.), *Language acquisition: Studies in first language development* (2nd ed.). Cambridge: Cambridge University Press.
- Bernhardt, E. B. (1987). Cognitive processes in L2: An examination of reading behaviors. In J. Lantolf & A. Labarca (Ed.), *Delaware Symposium on language studies: Research on second language acquisition in the classroom setting*. Norwood, NJ: Ablex.
- Bernhardt, E. B. (1991). *Reading development in a second language*. Norwood, NJ: Ablex.
- Biber, D. (1988). *Variation across speech and writing*. New York: Cambridge University Press.
- Biber, D. (1992). On the complexity of discourse complexity: A multidimensional analysis. *Discourse Processes*, 15, 133-163.
- Biber, D. (1993). An analytical framework for register studies. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 31-56). New York: Oxford University Press.
- Biber, D. (1995). *Dimensions of register variation*. New York: Cambridge University Press.
- Bowey, J. A., & Francis, J. (1991). Phonological analysis as a function of age and exposure to reading instruction. *Applied Psycholinguistics*, 12, 91-121.

-
- Brewer, W. F. (1980). Literary theory, rhetoric, and stylistics: Implications for psychology. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 221-239). Hillsdale, NJ: Erlbaum.
- Britton, B. K., & Black, J. B. (1985). *Understanding expository text*. Hillsdale, NJ: Erlbaum.
- Brown, T. & Haynes, M. (1985). Literacy background and reading development in a second language. In T. H. Carr (Ed.), *The development of reading skills* (pp. 19-34). San Francisco, CA: Jossey-Bass.
- Carpenter, P., Mikaye, A., & Just, M. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.
- Carrell, P. (1992). Awareness of text structure: Effects on recall. *Language Learning*, 42, 1-20.
- Carson, R. (1962). *Silent spring*. Greenwich, CT: Fawcett Publications.
- Carver, R. P. (1997). Reading for one second, one minute, or one year from the perspective of reading theory. *Scientific Studies of Reading*, 1 (1), 3-43.
- Chapelle, C., Grabe, W., & Berns, M. (1997). *Communicative language proficiency: Definition and implications for TOEFL 2000*. (TOEFL Monograph Series Report No. 10). Princeton, NJ: Educational Testing Service
- Chen, R. S., & Vellutino, F. (1997). Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research*, 29, 1-24.
- Clarke, M. A. (1979). The short circuit hypothesis of ESL reading or when language competence interferes with reading performance. *Modern Language Journal*, 64, 203-209.
- Clinton, H. R. (1996). *It takes a village*. New York: Touchstone.
- Clyne (1994). *Intercultural communication at work*. New York: Cambridge University Press.
- Cohen, A. D., Olshtain, E., & Rosenstein, D. S. (1986). Advanced EFL apologies: What remains to be learned? *International Journal of the Sociology of Language*, 62, 51-74.
- Crystal, D. (1991). *A dictionary of linguistics and phonetics*. Cambridge, MA: Basil Blackwell.
- Crystal, D. (1992). *An encyclopedic dictionary of language and languages*. Cambridge, MA: Basil Blackwell.
- Cummins, J. (1979). Linguistic interdependence and educational development of bilingual children. *Review of Educational Research*, 49, 222-251.

-
- Cummins, J. (1991). Interdependency of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70-89). Cambridge, England: Cambridge University Press.
- Cummins, J., Swain, M., Nakajima, K., Handscombe, J., & Green, D. (1981). *Linguistic interdependence in Japanese and Vietnamese students*. Report prepared for the Inter-America Research Associates, June. Toronto: Ontario Institute for Studies in Education.
- D'Angelo, F. (1980). *Process and thought in composition* (2nd ed.). Cambridge, MA: Winthrop.
- Devine, J. (1987). General language competence and adult second language reading. In J. Devine, P. L. Carrell, & D. E. Eskey (Eds.), *Research on reading English as a second language* (pp. 73-86). Washington, DC: Teaching English to Speakers of Other Languages.
- Devine, J. (1988). A case study of two readers: Models of reading and reading performance. In J. Devine, P. L. Carrell, & D. E. Eskey (Eds.), *Interactive approach to second language reading* (pp. 127-130). Cambridge: Cambridge University Press.
- Dillon, G. (1983). Interpersonal functions of textual analysis. Lecture presented at the University of Southern California, Los Angeles, March 1983.
- Drum, P. A., Calfee, R. C., & Cook, L. K. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly*, 16, 486-514.
- Durgunoglu, A. Y. (1997). Bilingual reading: Its components, developments, and other issues. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism, psycholinguistic perspectives*. Mahwah, NJ: Erlbaum.
- Durgunoglu, A. Y., Nagy, W. E., & Hancin, B. J. (1993). Cross-language transfer of phonemic awareness. *Journal of Educational Psychology*, 85, 453-465.
- Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Favreau, M., & Segalowitz, N. (1982). Second language reading in fluent bilinguals. *Applied Psycholinguistics*, 3, 329-341.
- Ferreiro, E. (1986). The interplay between information and assimilation in beginning literacy. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 15-49). Norwood, NJ: Ablex.
- Fowler, A. E., & Liberman, I. Y. (1995). The role of phonology and orthography in morphological awareness. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 157-188). Hillsdale, NJ: Lawrence Erlbaum.

-
- Freedle, R. (1997). The relevance of multiple-choice reading test data in studying expository passage comprehension: The saga of a 15 year effort towards an experimental/correlational merger. *Discourse Processes*, 23, 399-440.
- Freedle, R., & Kostin, I. (1993). *The prediction of TOEFL reading comprehension item difficulty for expository prose passages for three item types: Main idea, inference, and supporting idea items* (TOEFL Research Report No. 44). Princeton, NJ: Educational Testing Service.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum
- Gernsbacher, M. A. (Ed.). (1994). *Handbook of psycholinguistics*. San Diego, CA: Academic Press.
- Gernsbacher, M. A. (1996). The structure-building framework: What it is, what it might also be, and why. In B. K. Britton & A. C. Graesser (Eds.), *Models of text understanding* (pp. 289-311). Hillsdale, NJ: Erlbaum.
- Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, 23, 265-304.
- Geva, E., & Siegle, L. S. (1999). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing: An Interdisciplinary Journal*.
- Geva, E., Wade-Wooley, L., & Shany, M. (1997). Development of reading efficiency in first and second language. *Scientific Studies of Reading*, 1(2), 119-144.
- Gholamain, M., & Geva, E. (1999). The concurrent development of word recognition skills in English and Farsi. *Language Learning*.
- Givón, T. (1995). Coherence in text vs. coherence in mind. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text* (pp. 59-115). Philadelphia: John Benjamins.
- Goldman, S. R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, 23, 357-398.
- Goodman, K. S. (1973). Psycholinguistic universals of the reading process. In F. Smith (Ed.), *Psycholinguistics and reading* (pp. 21-29). New York: Holt, Rinehart and Winston.
- Goswami, U., & Bryant, P. (1992). Rhyme, analogy, and children's reading. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition*. Hillsdale, NJ: Erlbaum.

-
- Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1-13). Mahwah, NJ: Erlbaum.
- Green, D. W., & Meara, P. (1987). The effects of script on visual search. *Second Language Research, 3*, 102-117.
- Grolier Online Encyclopedia (1995). Danbury, CT: Grolier Electronic Publications, Inc.
- Gundel, J. K., & Tarone, E. E. (1983). Language transfer and the acquisition of pronominal anaphora. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 281-296). Rowley, MA: Newbury House.
- Guthrie, J. T. (1988). Locating information in documents: Examination of a cognitive model. *Reading Research Quarterly, 23*, 178-199.
- Guthrie, J. T., & Kirsch, I. S. (1987). Distinctions between reading comprehension and locating information in text. *Journal of Educational Psychology, 79*, 220-227.
- Guthrie, J. T., & Mosenthal, P. (1987). Literacy as multidimensional: Locating information and reading comprehension. *Educational Psychologist, 22*, 279-297.
- Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning, 26*, 321-351.
- Hakuta, K. (1982). Interaction between particles and word order in the comprehension and production of simple sentences in Japanese children. *Developmental Psychology, 18*, 62-76.
- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1996). *A study of writing tasks assigned in academic degree programs* (TOEFL Research Report No. 54). Princeton, NJ: Educational Testing Service.
- Hall, C. S., & Nordby, V. J. (1973). *A primer of Jungian psychology*. New York: New American Library.
- Hansen, E. G., & Willut, C. K. (1998). *Computer and communications technologies in colleges and universities in the Year 2000* (TOEFL Monograph Series Report No. 12). Princeton, NJ: Educational Testing Service.
- Hasuike, R., Tzeng, O. J. L., & Hung, D. L. (1986). Script effects and cerebral lateralization. In J. Vaid (Ed.), *Language processing in bilinguals: Psycholinguistic and neurological perspectives* (pp. 275-288). Hillsdale, NJ: Erlbaum.
- Hatch, E., Polin, P., & Part, S. (1974). Acoustic scanning and syntactic processing: Three reading experiments - first and second language learners. *Journal of Reading Behavior, 6*, 275- 285.

-
- Haynes, M., & Carr, T. H. (1990). Writing system background and second language reading: A component skills analysis of English reading by native-speaking readers of Chinese. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 375-421). San Diego, CA: Academic Press.
- Horiba, Y. (1993). Second language readers' memory for narrative texts: Evidence for structure-preserving top-down processing. *Language Learning, 43*, 345-72.
- Horowitz, R. (1987). Rhetorical structure in discourse processing. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending spoken and written language* (pp. 117-160). San Diego, CA: Academic Press.
- Irujo, S. (1986). Don't put your leg in your mouth: Transfer in the acquisition of idioms in a second language. *TESOL Quarterly, 20*, 287-304.
- Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (1999). *TOEFL 2000 framework: A working paper*. (TOEFL Monograph Series Report No. 16). Princeton, NJ: Educational Testing Service.
- Johnson, S. K. (1971). Sociology of Christmas cards. *Society, 8*, 40-45.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.
- Kintsch, W. (1995). How readers construct situation models for stories: The role of syntactic cues and casual inferences. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text* (pp. 139-160). Philadelphia: John Benjamins.
- Kirsch, I. S., & Mosenthal, P. B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*, 5-30.
- Koda, K. (1989). Effects of L1 orthographic representation on L2 phonological coding strategies. *Journal of Psycholinguistic Research, 18*, 201-222.
- Koda, K. (1990). The use of L1 reading strategies in L2 reading. *Studies in Second Language Acquisition, 12*, 393-410.
- Koda, K. (1993). Transferred L1 strategies and L2 syntactic structure during L2 sentence comprehension. *Modern Language Journal, 77*, 490-500.
- Koda, K. (1996). L2 word recognition research: A critical review. *The Modern Language Journal, 80*, 450-460.

-
- Koda, K. (1997). Orthographic knowledge in L2 lexical processing: A cross-linguistic perspective. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: Rationale for pedagogy* (pp. 35-52). New York: Cambridge University Press.
- Koda, K. (in press). Development of Kanji knowledge among adult L2 learners of Japanese. In I. H. Nara (Ed.), *Advances in Japanese pedagogy*. Columbus, OH: The Ohio State University Foreign Language Publications Office.
- Koda, K. (in press). Development of L2 intraword structural sensitivity and decoding skills. *Modern Language Journal*.
- Koda, K., Takahashi, E., & Fender, M. (in press). Effects of L1 processing experience on L2 intra-word awareness. *Ilha do Desterro*, Cognitive perspectives on second/foreign language acquisition/learning.
- Lautamatti, L. (1987). Observations on the development of topic in simplified discourse. In U. Connor & R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 tests* (pp. 92-126). Reading, MA: Addison-Wesley.
- Legareta, D. (1979). The effects of program models on language acquisition by Spanish speaking children. *TESOL Quarterly*, 13, 521-534.
- MacWhinney, B., & Bates, E. (Eds.). (1989). *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press.
- Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111-151.
- Mason, J. (1980). When do children begin to read? *Reading Research Quarterly*, 15, 203-227.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Meyer, B. (1987). Following the author's top-level organization: An important skill for reading comprehension. In R. Tierney, P. Anders, & J. Mitchell (Eds.), *Understanding readers' understanding: Theory and practice* (pp. 59-76). Hillsdale, NJ: Erlbaum.
- Moffett, J. (1983). *Teaching the universe of discourse*. Boston: Houghton Mifflin.
- Mosenthal, P. B. (1985). Defining the expository discourse continuum: Towards a taxonomy of expository text types. *Poetics*, 14, 387-414.

-
- Muljani, M., Koda, K., & Moates, D. (1998). Development of L2 word recognition: A Connectionist approach. *Applied Psycholinguistics*, *19*, 99-114.
- Nagy, W. E., & Anderson, R. C. (1997). Metalinguistic awareness and literacy acquisition in different languages. In D. Wagner, R. Venezky, & B. Street (Eds.), *Literacy: An international handbook*. New York: Garland.
- Navon, D., & Shimron, J. (1984). Reading Hebrew: How necessary is the graphemic representation of vowels? In L. Henderson (Ed.), *Orthographies and reading: Perspectives from cognitive psychology, neuropsychology, and linguistics* (pp. 91-102). Hillsdale, NJ: Erlbaum.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- National Council of Teachers of English/International Reading Association. (1996). *Standards for the English language arts*. Urbana, IL: NCTE.
- Nissan, S., DeVincenzi, F., & Tang, K. L. (1996). *An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension* (TOEFL Research Report No. 51). Princeton, NJ: Educational Testing Service.
- Oller, J. W., Jr. (1972). Assessing competence in ESL reading. *TESOL Quarterly*, *4*, 107-116.
- Oller, J. W., Jr., & Tullius, J. (1973). Reading skills of non-native speakers of English. *IRAL*, *11*, 69-80.
- Olshtain, E. (1983). Sociocultural competence and language transfer: The case of apology. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 232-249). Rowley, MA: Newbury House.
- Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension fostering and monitoring activities. *Cognition and Instruction*, *1*, 75-117.
- Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145-174). Hillsdale, NJ: Erlbaum.
- Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849-894). San Diego, CA: Academic Press.
- Perfetti, C. A. (1997). Sentences, individual differences, and multiple texts: Three issues in text comprehension. *Discourse Processes*, *23*, 337-355.
- Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly*, *33*, 283-319.

-
- Perkins, K. (1992). The effect of passage topical structure types on ESL reading comprehension difficulty. *Language Testing*, 9, 163-172.
- Purves, A., & Hawisher, G. (1990). Writers, judges, and text models. In R. Beach & S. Hynd (Eds.), *Developing discourse practices in adolescence and adulthood* (pp. 183-199). Norwood, NJ: Ablex.
- Raphael, T. E., & Pearson, P. D. (1982). *The effects of meta-cognitive strategy awareness training on students' question answering behavior* (Technical Report No. 238). Urbana, IL: University of Illinois, Center for the Study of Reading.
- Richards, J. (1976). The role of vocabulary teaching, *TESOL Quarterly*, 10, 77-89.
- Rutherford, W. E. (1983). Language typology and language transfer. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 358-370). Rowley, MA: Newbury House.
- Ryan, A., & Meara, P. (1991). The case of invisible vowels: Arabic speakers reading English words. *Reading in a Foreign Language*, 7, 531-540.
- Saito, Y. (1989, December). *The effects of article deletion in English on the cognitive processes reflected in the eye movements and metacognitive awareness of native readers and Japanese readers of English: An eye-tracking study*. Paper presented at the annual National Reading Conference, Austin, TX.
- Sasanuma, S. (1984). Can surface dyslexia occur in Japanese? In Henderson, L. (Ed.), *Orthographies and reading: Perspectives from cognitive psychology, neuropsychology, and linguistics* (pp. 43-56). Hillsdale, NJ: Erlbaum.
- Scarcella, R. C. (1983). Discourse accent in second language performance. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 306-326). Rowley, MA: Newbury House.
- Scheuneman, J., Gerritz, K., & Embretson, S. (1991). *Effects of prose complexity on achievement test item difficulty* (ETS Research Report No. 43). Princeton, NJ: Educational Testing Service.
- Segalowitz, N. S. (1986). Skilled reading in the second language. In J. Vaid (Ed.), *Language processing in bilinguals: Psycholinguistic and neurological perspectives* (pp. 3-19). Hillsdale, NJ: Erlbaum.
- Segalowitz, N. S., Poulsen, C., & Komoda, M. (1991). Lower level components of reading skill in higher level bilinguals: Implications for reading instruction. *AILA Review*, 8, 15-30.

-
- Sheehan, K., Ginther, A., & Schedl, M. (in press). *Development and validation of a proficiency scale for the TOEFL reading comprehension section* (TOEFL Research Study). Princeton, NJ: Educational Testing Service.
- Shively, W. P. (1997). The setting of power: The state. In C. Justice (Ed.), *Social science*. New York: McGraw Hill.
- Singer, M. (1990). *Psychology of language: An introduction to sentence and discourse processing*. Hillsdale, NJ: Erlbaum.
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17, 38-62.
- Skehan, P. (1997). *A cognitive approach to language learning*. London: Oxford University Press.
- Skehan, P. (1998). Task-based instruction. In W. Grabe, et al. (Eds.), *Annual Review of Applied Linguistics*, 18, 268-286.
- Skutnabb-Kangass, T., & Toukomaa, P. (1976). *Teaching migrant children's mother tongue and learning the language of the host country in the context of socio-cultural situation of the migrant family*. Helsinki: Finnish National Commission for UNESCO.
- Slobin, D. I. (1985). (Ed.), *The crosslinguistic study of language acquisition* (Vol. 2). Hillsdale, NJ: Erlbaum.
- Slobin, D. I., & Bever, T. G. (1982). Children's use of canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, 12, 229-265.
- Stanovich, K. E. (1991a). Changing models of reading and acquisition. In L. Rieben & C. A. Perfetti (Eds.), *Learning to read* (pp.19-32). Hillsdale, NJ: Erlbaum.
- Stanovich, K. E. (1991b). Word recognition: Changing perspectives. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research II* (pp. 418-452). New York: Longman.
- Stanovich, K. E., & Stanovich, P. (1995). How research might inform the debate about early reading acquisition. *Journal of Research in Reading*, 18(2), 87-195.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24, 402-433.
- Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing*. Norwood, NJ: Ablex

-
- Stein, N. L., & Policastro, M. (1984). The concept of story: A comparison between children's and teacher's viewpoints. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 113-155). Hillsdale, NJ: Erlbaum.
- Sticht, T. G., & James, J. H. (1984). Listening and reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 293-317). New York: Longman.
- Sulzby, E. (1986). Writing and reading: Signs of oral and written language organization in the young child. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 50-89). Norwood, NJ: Ablex.
- Troike, R. C. (1978). Research evidence for the effectiveness of bilingual education. *NABE Journal*, 3, 13-24.
- Turvey, M. T., Feldman, L. B., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and reading: Perspectives from cognitive psychology, neuropsychology, and linguistics* (pp. 81-90). Hillsdale, NJ: Erlbaum.
- Tyma, D. (1981). Anaphoric functions of some demonstrative noun phrases in EST. In *English for academic and technical purposes – Studies in honor of Louis Trimble* (pp. 65-75). In L. Selinker, E. Tarone, & V. Hanzeli, (Eds.) Rowley, MA: Newbury House.
- Urquhart, S., & Weir, C. (1998). *Reading in a second language: Process, product and practice*. London and New York: Longman.
- Vacca, R. T., & Vacca, J. L. (1996). *Content area reading*. New York: HarperCollins.
- Vaid, J. (1995). Script directionality affects nonlinguistic performance: Evidence from Hindi and Urdu. In I. Taylor & D. R. Olson (Eds.), *Scripts and literacy: Reading and learning to read the world's scripts* (pp. 295-310). Dordrecht: Kluwer Academic.
- Van Dijk, T. (1985). Strategic discourse comprehension. In T. Ballmer (Ed.), *Linguistic dynamics: Discourses, procedures, and evolution* (pp. 30-61). Berlin: Walter de Gruyter & Co.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly*, 33, 321-363.
- Vodopich, D. S., & Moore, R. (1996). *Biology laboratory manual*. Boston: McGraw-Hill.

-
- Wagner, R., & Stanovich, K. (1996). Expertise in reading. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 189-225). Mahwah, NJ: Erlbaum.
- Whitman, R. L., & Jackson, K. L. (1972). The unpredictability of contrastive analysis. *Language Learning*, 22, 29-41.
- Witte, S. P. (1983). Topical structure and revision: An exploratory study. *College Composition and Communication*, 34(3), 313-341.
- Yanco, J. J. (1985). Language contact and grammatical interference: Hausa and Zarma in Niamey, Niger. *Studies in African Linguistics*, 9, 318-322.
- Zehler, A. M. (1982). Reflection of first-language-derived processes in second-language acquisition. *DAI*, 43, 3054A.
- Zobl, H. (1983). Markedness and the projection problem. *Language Learning*, 33(3), 293-313.
- Zobl, H. (1984). The wave model of linguistic change and the naturalness of interlanguage. *Studies in Second Language Acquisition*, 6(2), 160-185.
- Zwaan, R., & Taylor, C. (1996). The influence of language proficiency and comprehension skill on situation-model construction. *Discourse Processes*, 21, 289-327.

Appendix A. Linguistic and Processing Variables

Language Knowledge Variables

Reading researchers interested in cognitive factors in reading, and individual differences among readers, have found that a number of basic linguistic variables contribute consistently to reading comprehension abilities across a wide range of reader groups. These variables include:

- Automatic orthographic-form recognition (Stanovich, 1991; Wagner & Stanovich, 1996)
- Phonemic analysis (Stanovich, 1991, Stanovich & Stanovich, 1995)
- Word recognition (Stanovich, 1991; Perfetti, 1994, 1997)
- Thorough word representation (Perfetti, 1992, 1994)
- Syntactic parsing (Gernsbacher, 1990, 1997; Perfetti, 1997; Kintsch, 1995)
- Proposition encoding and integration (Singer, 1990; various in Gernsbacher, 1994)

These factors are reviewed briefly in Grabe (1997).

Numerous variables thought to influence the difficulty of reading tests and tasks have been identified in previous research. Based on their analyses of seven different reading tests administered to students in elementary and secondary school, Drum, Calfee, and Cook (1981) concluded that skills necessary for successful performance included:

- 1) Accurate and fluent word recognition
- 2) Knowledge of specific word meanings
- 3) Knowledge of syntactic/semantic clause and sentence relationships
- 4) Recognition of the superordinate/subordinate idea structure of passages
- 5) Identification of the specific information requested in questions
- 6) Evaluation of the alternate choices in order to select the one that best fits
 - (a) the syntactic/semantic requirements of the question and
 - (b) the idea structure of the passage (pp. 488-89)

This list suggests that vocabulary and syntax, as well as cohesion and coherence relations, are important variables in the comprehension of reading texts. Scheuneman, Gerritz, and Embretson (1991) found many of the same skills to be necessary for successful performance on verbal reasoning items. They also found that the number and type of propositions embedded within items and passages accounted for additional variation in item difficulty for NTE[®] and GRE[®] items designed for adults.

Tyma (1981) analyzed difficulty contributed by demonstrative noun phrases in science texts. Lautamatti (1987) investigated five types of topical structures to look at relationships between and among sentences, and she postulated that texts that have more “easy” sentence types and limited topical development would be easier for second-language readers. Witte (1983) and Perkins (1992) found some evidence that this is the case. Perkins studied the effects of passage topical structure types on second-language reading difficulty using TOEFL reading comprehension passage sets and found that questions based on sentences in which topical

information occurred in the subject position tended to be easier than those in which this information was elsewhere.

Kirsch and Mosenthal (1990) identified three variables in document literacy tasks: (a) degree of correspondence between the phrasing used in the task directive (the question) and the phrasing used in the document; (b) the type of information requested (single or multiple points of information, concrete or abstract pieces of information, etc.), and (c) plausibility of the distractors.

Freedle and Kostin (1993) were able to account for 31 to 61 percent of the variance in difficulty on items in the TOEFL reading test with 32 variables, 14 of which were related to the text and 14 of which were related to lexical overlap between the text and the item.

Carpenter, Miyake, and Just (1994) identified linguistic ambiguities, syntactic complexities, and distance across text references or concepts over which relevant information must be retained as features of the texts themselves that strain working memory, thereby making comprehension more demanding.

Processing Variables

In the process of comprehending a text, the reader's language knowledge interacts with cognitive processing abilities to arrive at comprehension of a text. Processing efficiencies in working memory seem to be the basic mechanism for discussing these issues. Using a framework developed by Just and Carpenter, the following outline notes cognitive processing factors involved in reading.

A. *Basic Processing Efficiencies In Working Memory.* Efficiencies are due to the multiple simultaneous use of sets of processes and information, such as noted above, that are active in memory, as well as the rapid use of these processes and information resources. Processing limitations that may cause individual variation in reading include the following. (Many of these factors can also be seen as task variables.)

Language proficiency knowledge limits:

- Referent ambiguity tolerance and resolution
- Syntactic complexities encountered
- Distance across referent or concepts
- Time constraints
- Other task interference
- Speed of lexical access
- Speed of syntactic processing
- Speed of suppression mechanism

Processing limitations may also be due to:

- Contextual interference

-
- Emotional distress
 - Fluency and rate control
 - Text model building and bridging inferences

B. *Non-language Processing Efficiencies.*

- Visual processing
- Dual-coding theory

C. *Comprehension Variables.*

- Text model building
- Text structuring (Carrell, 1992)
- Discourse organization
 - Style
 - Register
 - Purpose for text, genre
- Signaling of discourse structure
- Information structuring
 - Given-new relations
 - Foregrounding main information, backgrounding secondary information
 - Important information in first-mention position
 - Marking thematic information with repetition, pronouns, unusual structures
- Situation model building (Perfetti, 1997; Zwaan & Taylor, 1996)
 - Inferencing
 - Strategic planning and goal setting
 - Strategic monitoring
 - Strategic repair and compensation
 - Situation model construction

References on listening-reading correlations

- Bell, L., & Perfetti, C. (1994). Reading skill: Some adult comparisons. *Journal of Educational Psychology, 86*, 244-55.
- Carver, R. (1997). One second of reading, one minute of reading, one year of reading. *Scientific Studies of Reading, 1*, 1-43.
- Chen, R., & Vellutino, F. (1997). Predictors of reading ability. *Journal of Literacy Research, 29*, 1-24.
- Cunningham, A., Stanovich, K., & Wilson, M. (1990). Cognitive variation in adult college students differing in reading ability. In T. Carr & B. Levy (Eds.), *Reading and its development* (pp. 129-59). San Diego: Academic Press.

-
- Curtis, M. (1980). Development of components of reading skill. *Journal of Educational Psychology, 72*, 656-69.
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Verbal Learning and Verbal Behavior, 19*, 450-66.
- Daneman, M. (1991). Individual differences in reading. In R. Barr et al. (Eds.), *Handbook of reading research, II*. (pp. 512-38). New York: Longman.
- Gough, P. B., & Walsh, M. A. (1991). Chinese, Phoenicians, and the orthographic cipher of English. In S. Brady & P. Shankweiler (Eds.) *Phonological processes in literacy* (pp. 199-210). Hillsdale, NJ: Erlbaum.
- Gough, P. B., Hoover, W., & Petersen, C. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 1-13). Mahwah, NJ: Erlbaum.
- Hoover, M., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*, 127-160.
- Jackson, M., & McClelland, J. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology General, 108*, 151-181.
- Palmer, J., MacLeod, J., Colin, M., Hunt, E., Davidson, J. E. (1985). Information processing correlates of reading. *Journal of Memory and Language, 24*, 59-88.
- Sticht, T., & James, J. (1984). Listening and reading. In P. D. Pearson, et al. (Eds.), *Handbook of reading research* (pp. 293-317) New York: Longman.
- Townsend, D., Carrithers, C., & Bever, T. (1987). Listening and reading processes in college and middle school age readers. In R. Horowitz & S. Samuels (Eds.), *Comprehending oral and written language* (pp. 217-242). New York: Academic Press.

Appendix B. Text and Task Variables

Text Variables

Few strong factors consistently stand out as basic to the establishment of the reading construct, but the distinction between expository and narrative prose types may be one such source of variation. Students perform differently on tasks that involve expository texts than they do on tasks that involve narrative texts. At the same time, there is no consistent evidence that any particular type of expository prose pattern of organization is more difficult than others (though a descriptive collection of statements commonly appears as easier to understand in experimental research). With respect to genre variation (or broadly interpreted register variation), it is probably fair to say that professional prose texts, written for professional colleagues, are more difficult to comprehend than instructional textbook excerpts, and both are probably harder to understand than popular texts in trade publications. Note also that some researchers would treat all of the variation possibilities within this section as sources of register variation, seeing this as a cover concept for ways that texts vary in response to intended uses and situations of use (Biber, 1993, 1995). The following may be helpful in considering possible texts for use in research tasks and prototype design.

Expository texts: Types and attributes

- Logical patterns of arrangement (top-level structuring)
- Non-episodic texts
- Argumentative
- Persuasive
- Real world
- Informational, informational density
- Spatial

Narrative texts: Types and attributes

- Causal
- Episodic
- Argumentative
- Generate involvement
- Tension
- Imaginative, temporal distance
- Constructive

Genre variation (perhaps register variation)

- Professional research texts
- Textbooks
- Popular writing
- Many specific genres

Register variation (further textual variation that changes the way a reader approaches a text).

These sources of textual variation are strongly reflected in specific lexical and structural choices made by authors of texts, both in terms of selecting one option as opposed to another and in terms of frequency of use. Terms for general parameters of register variation are likely to be overlapping:

- Formal vs. informal
- Personal vs. impersonal
- Concrete vs. abstract
- Situation independent vs. situation dependent (contextualized vs. decontextualized)
- Interpersonal vs. informational
- Politeness vs. solidarity
- Higher status vs. equal status vs. lower status
- No graphics vs. graphics/no support vs. graphics support vs. graphics interactive

Task Variables

A. Mosenthal and Kirsch, in various publications, argue that a set of task factors can account for individual variation in both document and prose literacy assessments. These task factors include:

Type of information requested (five levels of difficulty)

- Objects
- Times, attributes, amounts
- Manner, goals, purposes
- Cause, effect, evidence
- Equivalence, difference, theme

Type of match

- Locate
- Cycle
- Integrate
- Generate
- Complicating processing factors

Plausibility of distractors

- No matching distractors
- Matching distractor by lexical overlap with question or with true answer in key paragraph
- Matching distractor by lexical overlap with both question wording and true answer, one of which appears in key paragraph
- Matching distractors of both types appear multiple times and once in key paragraph

-
- Matching distractors of both types appear in key paragraph

Document structure (document complexity)

- Fry readability formula

B. Skehan (1996, 1997, 1998) proposes another view of task variability as a general frame for language learning and assessment. It would have to be assessed differently for its possible contribution to “tasks” as part of understanding the reading construct.

1. Task characteristics

- Familiarity
- Computation demand
- Code complexity
- Cognitive complexity
- Communicative stress

Time

Modality

Scale

Stakes

Control

- Support for task

2. Task planning differences

- Planning time available/amount of planning time
- No planning time

3. Task implementation conditions

- Pre-task activities
- During-task activities
- Post-task activities

4. Task performance-outcome differences

- Complexity
- Accuracy
- Fluency

Further sources on reading difficulty and complexity

1. See Freedle and Kostin (1993) for variables that contribute to reading difficulty: They began with 93 possible variables which were narrowed down to about 10 useful variables, depending on the test-taking pool used. After regressions, the useful variables contributing to difficulty included use of negative forms, location of main idea statements, and percentage of word overlap with key testing section (as a distractor variable).

-
2. See Nissan et al. (1996) for variables that contribute to comprehension difficulty in listening comprehension dialogue items. These included the use of infrequent vocabulary, the type of utterance pattern, negative forms in the stimulus, implicit vs. explicit information tested, and the role of speaker as institutional voice vs. personal voice.
 3. Gernsbacher's (1990, 1996, 1997) Structure Building Framework provides additional ways to consider the roles of syntax and lexical forms for building text comprehension structures. Basically, a reader, with each syntactic clause unit, either lays a foundation for text model building, elaborates the text model (mapping), or builds a new fragment (shifting). Poor readers tend to build too many fragments (for a number of possible reasons). Gernsbacher also adds processing factors: an information enhancement mechanism (speeding activation) and a suppression mechanism (inhibition process) for removing excess or unimportant information. Poor readers seem to have a much weaker information suppression mechanism, creating working memory inefficiencies. These are worth considering for the processing aspect of the reading construct and for sources of individual differences among readers. It is not clear how such ideas will translate into assessment environments.
 4. Biber (1992, 1995) explores the ways that texts vary as part of his theory of register variation. His approach suggests both a number of ways in which texts contribute to variation in reading comprehension and a different way to examine text complexity. In his general approach, he establishes communicative functions for text types based on lexical-grammatical variation within the texts. His work indicates ways that specific linguistic elements work together to determine discourse purposes. His work also forces a rethinking of simple assumptions about complexity measures that are commonly used in research.

Appendix C. Other Factors that May Contribute to Test Variation

The following factors may contribute to test variation, either as aspects of the reading construct or as additional sources of test variation.

1. Affective Variables
 - Motivation
 - Involvement
 - Interest
 - Anxiety
 - Self-esteem
 - Test wiseness and experience
2. Topic Variables
3. Background Knowledge Factors
4. L1 Factors
 - Reading abilities in first language
 - Social context variables
 - Training
 - Practice
5. General Cognitive Problem-solving Abilities

Appendix D. L1 and L2 Differences

The following list of specific L2 factors which may influence interpretation of the reading construct should be kept in mind in carrying out TOEFL 2000 research and test design:

1. L1 and L2 readers typically have very different sizes of vocabulary knowledge. These differences are reflected in different numbers of words, different ranges of meanings and nuances for words, and different thoroughness of word representations.
2. L1 and L2 readers are likely to have differing approaches to the range of authentic texts that they encounter. Many L2 readers regularly consider whether or not they are likely to understand a text or text genre (perhaps like L1 readers with difficult science texts). L2 students usually have many more encounters with frustration-level reading texts. This should have some long-term impact on L2 reading practice and on motivation.
3. The L2 language threshold is a uniquely L2 issue. L1 students have enough language from the time that they start reading so that basic language knowledge will not be a severe impediment to comprehension.
4. L2 readers typically have a much greater awareness of language as part of their reading resources. They should respond better to language awareness instruction, patterns of discourse organization, and discourse signaling forms. Metalinguistic awareness may be a more important resource (and need) than it is for L1 readers.
5. There are real sources of linguistic differences across any two languages that may have an impact on L2 reading comprehension. These linguistic differences may be influential at the orthographic recognition level, lexical meaning level, morphological complexity level, syntactic constraints level, or discourse organization level, as well as in expected frequency of occurrence of syntactic and discourse uses. There is also an issue of perceived differences by a reader between the L1 and the L2.
6. There is a question of the extent to which L1 linguistic knowledge transfers to the L2 reading, and whether such transfer may be a support for L2 reading or an interference. There seems to be good evidence for transfer effects at early levels of L2 reading ability. It also seems that transfer is stronger at word recognition and discourse levels and less strong at morphological and syntactic levels except for beginning L2 readers.
7. The fact that L2 reading must involve processing with two languages is, in itself, a difference with likely implications for L2 reading. A few studies have reported that L2 readers simply have slower processing efficiencies, as indicated by word recognition and reading rates.
8. The use of various supporting resources for L2 reading marks it as different from L1 reading. Typically the use of glosses, bilingual dictionaries, cognates, and translations are unique features of L2 reading (though students reading science texts may encounter similar issues).

-
9. The extent of reading practice in the L2 will mark the typical L2 reader as different from the L1 reader. Stanovich's exposure to print research would argue that amount of reading in the L2 makes a strong contribution to individual differences. It is reasonable to assume that the large majority of L2 readers cannot have the same extent of exposure to print in the L2 as does the L1 reader. L2 reading exposure will also be reflected in slower L2 reading rates and reading fluency in general.
 10. The level of cultural knowledge of L2 situations and language uses that an L2 reader has cannot compare with the cultural resources used by an L1 reader. This issue represents the broadest level of the influence of background knowledge on reading comprehension.
 11. The L2 reader typically has a more developed metacognitive awareness of the learning activities involved while reading in the L2. If for no other reason, most L2 readers are older when they encounter conceptual information in L2 reading, and they are not limited by conceptual knowledge to the extent that an L1 reader might be at the same level of language knowledge. The L2 student typically also has had some success with L1 reading and can draw on that success. Being older when engaging in L2 reading, the student may also know more about how to learn in general and use this knowledge to help bootstrap L2 reading.
 12. The L2 reader is influenced by the level of L1 reading abilities that has developed. Research typically shows a strong correlation between L1 and L2 reading abilities. So L1 reading abilities will play a strong role in L2 reading performance.
 13. The differing social context experiences for learning to read in the L1, and the differing opportunities to use the L2, will influence performance in L2 reading. Students from cultures who do not read extensively, who read texts as though they were sacred, who do not question texts, who perform texts on a regular basis, who are not exposed to a wide range of genres and texts, who do not see others read for enjoyment, who do not use reading for many purposes, or who cannot afford the time for reading in the L1 are all likely to have differing profiles of L2 reading.
 14. A further difference for L2 reading is the likelihood of different motivations for reading, as well as differing senses of self-esteem, interest, involvement with reading, and emotional response to reading. Since most L2 readers are more extrinsically driven to read the L2, there may be variations in reading development for L2 readers that differ from L1 reading development.

Appendix E. Register Variations

Dillon (1983), in early exploration of this idea, proposed the following five parameters of variation along which all texts can be placed: personal-impersonal, distance-solidarity, superior-equal, oblique-confronted, formal-informal. The point of these dimensions is that they should offer some independent contribution to understanding what the text is signaling on multiple levels of interpretation, serving multiple purposes.

A second set of register parameters, used for cross-cultural text comparison, was proposed by Purves and Hawisher (1990):

- Personal-impersonal (extent of reference to writer's thoughts and feelings)
- Ornamented-plain (figurative-literal)
- Abstract-concrete (extent of specific information and detailed references)
- Single-multiple (focus on singular aspect or multiple aspects of subject)
- Propositional-appositional (multiple propositional relations and embedding vs. simple additive/conjunction signaling)

A third set of register parameters is proposed by Clyne (1994):

- Form vs. content emphasis
- Verbal vs. literate emphasis
- Rhythm of discourse
- Directionality
- Abstractness vs. concreteness

A fourth system of parameter variation, proposed by Biber (1988, 1993, 1995), is well supported empirically and has gone through a number of refinements. Briefly, his parameters of register variation include the following:

- Involved vs. informational production
- Narrative vs. non-narrative discourse
- Situation dependent vs. elaborated reference
- Overt expression of argument
- Abstract vs. non-abstract style
- On-line information elaboration marking stance
- Academic hedging



Test of English as a Foreign Language
P.O. Box 6155
Princeton, NJ 08541-6155
USA

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

Phone: 609-771-7100

E-mail: toefl@ets.org

Web site: <http://www.toefl.org>