TOEFL®

Monograph Series

MS - 23
AUGUST 2002

The Effects of Notetaking, Lecture Length and Topic on the Listening Component of TOEFL 2000

Patricia L. Carrell
Patricia A. Dunkel
Pamela Mollaun

ETS® Educational Testing Service

THE EFFECTS OF NOTETAKING, LECTURE LENGTH AND TOPIC ON THE LISTENING COMPONENT OF TOEFL 2000

Patricia L. Carrell
Patricia A. Dunkel
Pamela Mollaun

Educational Testing Service
Princeton, New Jersey
RM-02-04

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

Copyright © 2002 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 logos, TOEFL, the TOEFL logo, and TSE are registered trademarks of Educational Testing Service.

The Test of English as a Foreign Language and Test of Spoken English are trademarks of Educational Testing Service.

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

Email: toefl@ets.org

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

Foreward

The TOEFL Monograph Series features commissioned papers and reports for TOEFL 2000 and other Test of English as a Foreign Language™ (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 new forms of assessment 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 were the working papers that laid out the TOEFL 2000 conceptual frameworks with their accompanying research agendas. The initial framework document, Monograph 16, described the process by which the project was to move from identifying the test domain to building an empirically based interpretation of test scores. The subsequent framework documents, Monographs 17-20, extended the conceptual frameworks to the domains of reading, writing, listening, and speaking (both as independent and interdependent domains). These conceptual frameworks guided the research and prototyping studies described in subsequent monographs that resulted in the final test model.

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 present study examined the effects of notetaking, lecture length, and topic, as well as two aptitude variables on listening comprehension with ESL students representative of the TOEFL population. A total of 234 ESL students at five participating universities in the United States took a computer-based listening comprehension test, a short-term memory test, the listening comprehension section of a disclosed Institutional (paper-and-pencil) TOEFL, a debriefing questionnaire, and a biodata questionnaire. Results revealed positive effects for notetaking and lecture length, as well as significant interactions between notetaking and topic, and between notetaking and lecture length. No differences in the pattern of results occurred when listening comprehension proficiency and short-term memory were taken into consideration with the three main factors.

Key words/phrases: Listening comprehension, notetaking, lecture length, topic, short-term memory

Acknowledgments

The authors wish to thank the consultants at the five cooperating universities:

Samuel T. Lee, University of Southern California (USC), Sylvia D. Smythe, University of Southern California (USC), Alfred D. Stover, University of Arizona (UAZ), Beverly Ruiz, Indiana University (IU), J. Becky Pharis, Southern Illinois University at Carbondale (SIUC), and Ray Graham, Brigham Young University (BYU).

Further, the authors wish to acknowledge the assistance of the following individuals: Frances McCarty, statistical consultant, and Joanne Crump, research assistant, both at Georgia State University (GSU).

At ETS, we wish to thank Carol Taylor, Lou Mang, and Vanessa Hubbard.

Finally, we wish to acknowledge the support of Craig Carrell, who provided financial accounting and reporting support to the project.

Table of Contents

Page

List of Tables

Page

List of Figures

Statement of Problem or Rationale

Taking notes while listening to a lecture is widely accepted as a useful strategy for augmenting student attention and retention of academic discourse. Yet previous research with both native and nonnative speakers of English has yielded mixed results regarding the facilitating effects of notetaking.

Because listeners in a college classroom are usually given the opportunity to take notes while listening to a lecture, allowing them to do so on the TOEFL test would simulate more closely the college classroom experience. Allowing students to take notes might, thereby, increase the validity of the TOEFL testing task.

Therefore, the study reported herein investigated the effects of allowing notetaking along with the effects of lecture length and topic with a sample of students representative of TOEFL examinees.

Literature Review

Notetaking is intuitively appealing to the lecture listener and is generally viewed as one class of mathemagenic activity that facilitates the process of learning and remembering lecture material (Clerehan, 1995; Crawford, 1925a, b, c; Dunkel, 1988; Dunkel & Davy, 1989; Kiewra, 1987; McKenna, 1987; Palmatier & Bennet, 1974). Van Meter, Yokoi, and Pressley (1994) note that most college students take notes in most classes because notetaking is the primary means of creating "a record of information" that is presented in lectures. In a study of college students' theory of notetaking, Van Meter et al. asked students what they hoped to achieve by taking notes. Although the major goal of notetaking is doing well in courses, the students designated a number of primary goals related to attention (it increases attention to the lecture); understanding (it increases student comprehension and memory of material presented in the lecture); organization (it provides an opportunity to connect ideas, provide structure, or generate holistic representation of lecture content); study aid (it informs about the content of exams); homework aid (it informs about solutions to practice problems and provides information relevant to written assignment). When asked about the **content** placed in their notes, the students reported that they placed into their notes the following: content redundant with the text; material the professor stressed; content on the board or overheads; content cited in the syllabus; definitions, main points, important concepts and ideas; and information not well understood or not familiar. They noted that guest lecturer and film content is not noted down, nor is content that is common knowledge. In response to a query about the **structure** and preferred methods of notetaking, while preferred methods varied from student to student, commonalities noted included inclusion of key terms, an outline of some sort (e.g., flagging relationships between more and less important content), and personal shorthand.

A number of experimental studies have been carried out to examine the effect of notetaking on lecture information retention and recall. In one of the early experimental studies of the effectiveness of notetaking versus nonnotetaking for native English speakers,[1] Crawford (1925a) concluded that "taking notes on a point does not guarantee its being recalled at the time of the quiz, but failing to take note of it very greatly decreases its chances of being recalled"

[1]It should be noted that the vast majority of experimental studies have focused on the study of native-English-speaker notetaking during and after listening to a lecture given in English. Few empirical studies have been carried out on the effectiveness of notetaking for nonnative English speakers listening to lectures in English (see Dunkel, 1985).

(p. 289). Half a century later, Aiken, Thomas, and Shennum (1975) noted that lecture material was twice as likely to be recalled if it was configured in note form than if it was not. However, the case for the efficacy of notetaking during lecture learning is, by no means, as clear-cut as it seems. In 1978, Hartley and Davies summarized the experimental or quasiexperimental research concerning the effect of the process of notetaking on lecture information recall. They determined that only 17 studies of the 35 examined supported enhancement of recall as a function of the process of notetaking. The research of the 1980s and 1990s provided more conflicting evidence concerning the utility of notetaking (see Dunkel, 1985). In fact, researchers are still seeking to determine whether notetaking per se during a lecture presentation is facilitative, debilitative, or of no particular use to the listening comprehension and recall of lecture material by native-speaker and nonnative-speaker listeners. The research includes studies of an experimental nature (Chaudron, Loschky, & Cook, 1994; DiVesta & Grey, 1972; Dunkel, 1985; Kiewra, DuBois, Christian, McShane, Meyerhoffer, & Roskelley, 1991), a qualitative-interpretive type (Adamson, 1993; Benson, 1989; Flowerdew, 1994; King, 1994; Van Meter, et al., 1994); and a phenomenologic variety (Bilbow, 1989; Fahmy & Bilton, 1990). All seek to assess the effect of notetaking during lecture learning.

The facilitative effect of notetaking is thought to derive from one or both of its two postulated functions: (a) the encoding function and (b) the external storage function. Encoding, the process (or act) of notetaking, supposedly aids lecture learning by activating attentional mechanisms and engaging the learner's cognitive processes of coding, integrating, synthesizing, and transforming aurally received input into a personally meaningful form. The external storage function of notetaking is seen as important because the notes taken serve as an external repository of information that permits later revision and review to stimulate recall of the information heard. As Cohn, Cohn, and Bradley (1995) argue, the encoding function involves a process of transforming and reorganizing material heard as a way of learning from the act of notetaking itself, whereas the external storage function involves the use of notes for the purpose of review as a means of learning. Kiewra and his colleagues (1989, 1991) contend that the traditional measure of the external storage function actually represents a combined function of both encoding and external storage. In their 1989 paper, Kiewra, Benton, Christian, Kim, and Lindberg suggest that "students who take and review their own notes have actually experienced both the encoding function plus the external storage function of notetaking" (p. 217), and that researchers should reclassify the traditional external storage function as an "encoding-plus-storage function"[2] if the same person who took the notes uses them in recognition/recall of information.

The storage (or product) function of notetaking is thought to derive when listeners have their notes for later review and study for tests on the lecture information. This function has been probed experimentally, according to Kiewra (1987), by comparing the performance of notetakers who have access to (and review) their notes before taking a recognition or recall test of lecture information, and notetakers who do not review their notes. In 24 studies reviewed by Hartley (1983) and/or Kiewra, listeners who reviewed their notes performed higher on postlecture achievements tests than did those who did not review their notes, thereby demonstrating support for the product function. In 8 other studies, no significant differences were found between reviewers and nonreviewers. Interestingly, no study indicated that reviewing notes was debilitative for lecture listeners. Additionally, Kiewra points out that nonexperimental (correlational) studies similarly support the product function of notetaking. Studies conducted by Crawford (1925a, b, c), Fisher and Harris (1973), and Kiewra (1984, 1987) all found significant correlations between amount of notetaking and achievement on information recall/recognition tests when notes served their external storage function and were reviewed.

A number of studies have found notetaking (i.e., the <u>very act</u> of notetaking) to have a positive encoding effect only <u>under certain conditions</u> (Chaudron, Cook, & Loschky, 1988; Einstein, Morris, & Smith, 1985; Fischer & Harris, 1973; Hult, Cohn, & Potter, 1984; Kiewra, Benton, Risch, & Christensen, 1995; Kiewra & Fletcher, 1984; Rickards, Fajen, Sullivan, & Gillespie, 1997). Rickards and colleagues (1997), for example, found that, overall, recall of text information was maximized for notetakers who heard and took notes on "signaled" texts (i.e., texts in which the speaker provided cues concerning the important content and its organizational structure); on nonsignaled texts, recall of notetakers was minimized. However, it was also found that notetaking in the presence of signals enhanced information recall of field-dependent, but not

[2]Kiewra et al. (1989) proposed a new means for examining the external storage function independently. This new function was represented by absent students who had no opportunity to view (or encode) the lecture presentation, but who were each provided with a set of the "borrowed" notes of attending students for review purposes. The variation produced three notetaking conditions: the original encoding function (take notes/no review), the newly classified encoding-plus-storage function (take notes/review), and the new, independent external storage function (borrow notes/review). Findings indicated that notetaking served a minimal encoding function and that its primary value was in providing external storage; the combination of encoding plus storage (take notes/review) was most facilitative of postlecture test performance.

field-independent learners.[3] In a study of the effect of field independence-dependence and study technique on lecture learning, Frank (1984) found that field-independent students performed better than field-dependent students under the student's notes condition. It was further noted that the notes of the field-independent students were more efficient and tended to be in an outline format more often than notes of field-dependent students.[4] Examining the effect of different notetaking formats (i.e., convention, outline, or matrix notetaking) on the recall of information contained in a 19-minute lecture, Kiewra et al. (1995) found that outline notetaking with subsequent review of the notes produced higher recall than did conventional or matrix notetaking. In a study of computer-based notetaking, Armel and Shrock (1996) found that those required to take notes on the computer screen scored significantly higher on the information-retention posttest than did those for whom notetaking was optional or not allowed.

Other studies have failed to detect the facilitative effect of the act of notetaking (Chaudron, et al., 1994; Dunkel, 1985; Hale & Courtney, 1994; Kiewra, 1985). Hale and Courtney (1994), for example, found that allowing participants taking the Test of English as a Foreign Language (TOEFL) to take notes on the lectures or minitalks heard had little effect on test performance. Reflecting on the findings, the researchers speculated that because the talks were relatively short (less than 2 minutes),[5] memory for the information contained in the minitalks might not have been heavily taxed, especially since the comprehension test questions did not query about facts and details that might easily be forgotten without the aid of notes— details, however, that might have been remembered if notetaking had been allowed (i.e., if the information had been noted down in the listener's notes and available for the listener's review).

[3]Rickards et al. (1997) posited that notetakers find it difficult to search for the structure in nonsignaled text under the constraints of listening. Interviewing college students about their notetaking strategies and practices, Van Meter, Yokoi, and Pressley (1994) found that the students surmised that the lecturer's signals explicitly pointed out the important content and indicated relationships between/within the information heard; the signals were used to form the basis for the students' notetaking.

[4]Field-independent and field-dependent students listened to a taped lecture under one of four study technique conditions: (a) no notes, (b) student's notes, (c) outline framework plus student's notes, and (d) complete outline plus student's notes. A 10-minute review period followed the lecture.

[5]Hale and Courtney (1994) note that the kinds of comprehension questions asked in their study might also have been a factor in the lack of notetaking effect. "Typically, TOEFL minitalks are followed by questions that tap general understanding of the passage. Students are not asked to remember very specific details, such as names and dates. And, although they are occasionally asked about the main point of the passage, this type of question is usually among the easiest and often can be answered correctly without need for high-level discourse processing. Most of the questions deal with information contained in the talk that is specifically stated and does not involve a level of detail that is so minor as to be difficult to retain over a short period of time" (p. 9). Our study examined not only gist questions but also detail questions, and those stated explicitly and implicitly in the minitalk.

Like Hale and Courtney (1994), Dunkel (1985) failed to find a positive effect of notetaking on ESL listeners' comprehension/recognition of information presented in a 22-minute English minilecture. Dunkel did, however, detect a sizable "memory effect" on performance; listeners with high short-term memory ability accurately recalled significantly more lecture concepts and details than did listeners with low short-term memory ability. (Memory ability was measured by performance on the digit span subtest of the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974).) Level of English proficiency was also found to have a positive effect on comprehension test performance. Native speakers of English (higher proficiency listeners) outperformed nonnative speakers of English (lower proficiency listeners) in recognizing concepts and details presented in the lecture, whether or not they took notes.

Although the results of Dunkel's (1995) study failed to support the notion that the mere "act of notetaking" augments cognitive encoding of aurally received information, they highlight the role of participant aptitudes in the processing of lecture material. According to Waldberg (cited in Ganske, 1981), aptitude variables account for a large proportion of variance (approximately 40% to 60%) in experimental studies of learning and aptitude (the dependent variable in much educational research). The results of Dunkel's study suggested that participant aptitudes (e.g., English language proficiency and short-term memory) had a decided effect on lecture learning.[6] They may, in fact, have been of sufficient importance that they vitiated the effect of notetaking for the ESL listeners. Examining the strong positive correlation found between memory and recall performance in Berliner's 1971 notetaking study, Weener (1974) went so far as to propose that all studies on the effects of notetaking should include an analysis of the interactive effect of memory on notetaking and lecture comprehension/recall. Although the present study focused primarily on analyzing the effect of notetaking in relation to the lengths and topics of four computer-based lectures or minitalks, the effects of these three main variables

[6]Hughes and Suritsky (1994) studied the notetaking of 30 learning disabled (LD) and nondisabled university students. They found that the LD university students performed significantly lower recalling the cued and noncued information. Participant aptitudes (in this case, learning ability) played a notable role in the effect of notetaking and the recall of information.

in conjunction with two additional variables, listening comprehension proficiency, and short-term memory, were also analyzed.[7]

The effect of notetaking was investigated in the context of two lecture lengths (2.5 minutes and 5 minutes) and two topic types (physical sciences and arts/humanities) and within the context of an aggregate score representing comprehension and recall of the following item types: (1) main idea, (2) supporting idea, (3) detail, and (4) minor detail. The response types included (1) multiple choice, (2) order/match, (3) multiple selection multiple choice, and (4) constructed response. The information types included (1) explicit information and (2) implicit, inferred information.[8] (See Table 3 for a listing of the item types, response types, and information types associated with the longer and shorter versions of the lectures used in this study.) The current TOEFL computer-based listening format served as the testing environment.

Relevance of the Problem to the TOEFL 2000 Project

The Importance of Investigating the Notetaking Variable

The intuitive belief held by college students and lecturers alike that notetaking promotes lecture learning causes listeners (e.g., TOEFL examinees) to place great value on their ability to take notes during lecture (or minitalk) presentations.[9] Not being allowed to take notes during the TOEFL minitalks seems to concern many TOEFL test takers (personal communication with G. Hale of Educational Testing Service, March 1993), regardless of the fact that researchers have not been able unequivocally to document that notetaking per se has a facilitative effect on ESL lecture processing in general, and TOEFL minitalk processing in particular. Many examinees (and test users) believe that being allowed to take notes on the minitalks could enhance performance and would give a better snapshot of the examinees' listening comprehension ability.

---

[7]Participants who took notes were allowed to review their notes to answer postlecture comprehension/recall questions. Thus, the storage, as well as the encoding, function of notetaking were probed in the current study. In Dunkel's (1985) study, only the encoding function of notetaking was examined because participants were prevented from reviewing or using the notes they had taken to answer the postlecture test questions. Dunkel strongly suggested that further research into the impact of notetaking should incorporate a review condition even when the intent is to probe the act of notetaking, because review of notes may enhance the effect of notetaking and is more similar to the real-world purpose of taking notes on aurally received information.

[8]In the present study, an aggregate percentage score was used in the analysis; however, in future analyses of the data, it will be possible to scrutinize the data in terms of the various kinds of item types, response types and information types created for the dependent measure.

[9]In a survey of 164 American and international students enrolled in an American research university, when Dunkel and Davey (1989) asked participants whether notes taken were useful for organizing information heard in a lecture. 96.3% of the Americans and 89.4% of the internationals responded in the affirmative.

Although researchers (e.g., Dunkel, 1985; Dunkel, Mishra, & Berliner, 1989; Hale & Courtney, 1994) have not been able to find support for this notion, it seemed time to reexamine this issue within the context of (1) the longer lectures or minitalks being considered for use in the TOEFL 2000 and (2) the computer-based testing (CBT) environment. The issue of notetaking was revisited mainly to provide data that could be used to help determine whether TOEFL 2000 examinees should be allowed to take notes during the lecture/minitalk section of the computer-based TOEFL.

The Importance of Investigating the Length and Topic Variables

The researchers hoped to provide the TOEFL 2000 program with data about what effect doubling the present length of the minitalks has on listener test performance. In their framework for testing academic listening comprehension as part of the TOEFL 2000 initiative, Bejar, Douglas, Jamieson, Nissan, and Turner (1998) point out that this framework will be used to "define listening as it will be measured in TOEFL 2000" (p. 4), and will be used to identify those variables perceived most relevant to the task of assessing the skill of academic listening. Among the variables identified by Bejar et al. is one investigated in this study: the length of the listening text or passage[10] and its effect on both the difficulty of the listening task and specific item difficulty.

The length of the text that serves as the input for the listener-examinee is of major relevance to the TOEFL 2000 project, since it affects the face, content, and construct validity of the TOEFL in the eyes of both examinees and score users. Clearly, in a university setting, academic listeners are required to listen to, extract information from, and remember information presented in lectures longer than 2.5 minutes, the amount of time currently relegated to the TOEFL minitalks. If the TOEFL 2000 is to aim for greater face (as well as construct) validity, it is thought the minilectures need to be longer than 2.5 minutes. How much longer has yet to be determined, so, in the present study, we examined the effect of doubling the length of the minitalk

[10]The question of whether length of text affects task or item difficulty was investigated by Nissan and her colleagues (cited in Bejar et al., 1998, p.16) who found that the number of words in the text did not contribute to performance difficulty. There was, however, "little variation in the length of the texts" used, so the issue remains open to further investigation. In Dunkel (1985), short-term memory ability in English, as measured by an English-digit-span test, was directly related to a participant's ability to understand and recall lecture information. However, Henning (cited in Bejar et al.) found no evidence of a relationship between memory load and item performance. Bejar et al. note that Henning's "findings must be cautiously interpreted due to his operationalization of passage length" (p. 16). Yepes-Baraya, Yepes, and Gorham (cited in Bejar et al.) are examining the relationship between text length and memory, using relatively short texts of up to 2.5 minutes. The question of memory load and listening ability remains open to investigation, especially for texts longer than 2.5 minutes.

input and determining the effect of that doubling on listeners' performance on the types of items and response types and information types identified in the framework (see Bejar et al., 1998; Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 1997).

We also examined the effect that topic of the minitalk has on the information recall of examinees, to determine whether topic interacts with the effects of notetaking and lecture length. Researchers have analyzed the effect of a number of text variables, such as rhetorical organizers (Chaudron & Richards, 1986; Dunkel & Davis, 1995; Meyer & Freedle, 1984); amount of lexical overlap (Freedle & Felbaum, cited in Freedle & Kostin, 1999); and redundancy (Chiang & Dunkel, 1992) in the listening text; and topical differences in the minitalks (Freedle & Kostin, 1999). Most recently, Freedle and Kostin (1999) examined the relationship between the ease/difficulty of TOEFL listening items and a number of text variables, one of which included the topical effects of subject matter. In an ex post facto correlational analysis of 337 listening comprehension items associated with 69 minitalks passages, the researchers found that the non-academic subject matters were associated with easier listening items, but most of the academic subject matters were associated with more difficult listening items representing, specifically, physical sciences, biological science, humanities, and arts. Freedle and Kostin suggest that instead of relying on examination of accrued TOEFL data, additional empirical work should be conducted to clarify how text variables (e.g., topic) affect listener performance on the TOEFL minitalk passages. The present study included an analysis of minitalk topic (physical sciences versus arts and humanities) to see what, if any, influence topic had on the other major variables under investigation: notetaking and lecture length.

The Importance of Investigating the Short-term Memory and Overall Level of Listening Proficiency Variables

The importance of investigating short-term memory in the context of notetaking, lecture length and topic is discussed above in the literature review (see especially Dunkel, 1985). It was also decided to examine the other major variable involved in the comprehension and retention of lecture information, namely, the listener's overall level of listening proficiency. Although it might be assumed that overall level of listening proficiency (as measured, for example, by the listening comprehension section of the paper-and-pencil TOEFL) would be highly correlated with performance on the more specialized type of lecture-listening test represented by the computer-based test in the present study, we deemed it important to test that assumption, especially because some research has found interaction between overall level of listening proficiency and lecture topic (Hansen & Jensen, 1994).

Research Questions

Four primary research questions were addressed in this study:

1. Is listening comprehension, as reflected by the percent correct score on a computer-based test, affected by the opportunity to take and use notes?
2. Is listening comprehension, as reflected by the percent correct score on a computer-based test, affected by the length (2.5 versus 5 minutes) of the minitalk used to present the content?
3. Is listening comprehension, as reflected by the percent correct score on a computer-based test, affected by the topic (arts/humanities versus physical sciences) of the minitalk?
4. Are there interactions among notetaking, length, and topic that affect listening comprehension, as reflected by percent correct score on a computer-based test?

Two secondary research questions were also addressed:

1. What effect does adding the variable overall English listening proficiency (as measured by the paper-and-pencil TOEFL) have on the answers to the primary research questions above?
2. What effect does adding the variable short-term memory (STM) (as measured by the digit span test) have on the answers to the primary research questions above?

Methodology

Participants

Data from 234 participants are included in the study.[11] Of the 234, 139 reported themselves as male, 88 as female. Participants were international students studying English as a second language at five participating institutions: Brigham Young University in Provo, Utah (BYU); Indiana University in Bloomington (IU); Southern Illinois University in Carbondale (SIUC); University of Arizona in Tucson (UAZ); and University of Southern California in

[11] Due to missing values for some data for some participants, totals may not equal 234.

Los Angeles (USC). Participants were representative of the usual TOEFL test-taking population in terms of both their biographical data and their general levels of English proficiency as measured by the Institutional TOEFL listening comprehension section (minimum = 31, maximum = 66, M = 48.61, SD = 6.21). Participants' minimum age was 18, maximum age was 58, average age was 24.67 (SD = 5.24). They had spent an average of 56 months studying English (minimum 0 months, maximum 267 months, SD = 48.8). They had spent on average 7 months in the United States (minimum 1 month, maximum 69 months, SD = 9.00). They represented various regions of the world, various native language backgrounds, and various fields of study. (See Table 1.) As incentive, participants were given gift certificates at the incampus bookstores (value between $15 and $20, depending upon the number of participants at the institution).

Table 1

Biographical Data on Participants

Institution		Region of the World	
BYU	N = 50	Asia	N = 115
IU	N = 63	South/Latin America	N = 53
SIUC	N = 50	Middle East/North Africa	N = 35
UAZ	N = 44	Europe	N = 17
USC	N = 27[12]	Sub-Saharan Africa	N = 6

Greatest Representation in the Study:

Countries		Native Languages	
Korea	N = 37	Spanish	N = 43
Japan	N = 29	Korean	N = 37
Taiwan	N = 24	Arabic	N = 33
United Arab Emirates	N = 16	Chinese	N = 34
Mexico	N = 12	Japanese	N = 29
PRC	N = 10	Portuguese	N = 10
All other countries	N < 10	All other native languages	N < 10

Academic Level		Fields of Study	
Preuniversity	N = 56	Engineering/Mathematics Computer Science/Physical Sciences	N = 74
Undergraduate	N = 81	Business	N = 45
Graduate	N = 84	Humanities	N = 32
Other	N = 6	Social Sciences	N = 15
		All other fields of study	N < 15

Materials and Procedures

Participants were tested in two sessions: Session 1 and Session 2, with a maximum time of two weeks between the sessions, and a minimum time of no less than one hour between the sessions. Each session lasted between 1 hour and 1 hour and 15 minutes.

Materials administered in the first session consisted of an informed consent form, a biodata questionnaire (see Appendix A), a short-term memory test, and the listening comprehension section of a disclosed Institutional, paper-and-pencil TOEFL.

[12]USC was the first institution to administer the two separate testing sessions. Because of some logistical glitches in this first administration, which were remedied before the other institutions administered the instruments, a number of USC participants had to be dropped from the study.

Materials administered in the second session consisted of the computer-based test designed for this study and a debriefing questionnaire (see Appendix B).

Short-term Memory Test

The digit span subtest of the WAIS-III was used as the test of short-term memory. The test has a reliability of .78, according to Sattler (1981). This test, which is the most widely used test of short-term memory, also has the advantage of minimizing the influence of language and culture on nonnative speakers' measure of memory span in English. In the digit span test, the digits forward task involves primarily rote learning and short-term memory, whereas the digits backward task requires considerably greater transformation of the stimulus input prior to recall. Mishra, Ferguson, and King (1985) note that the digit span is a measure of memory, attention, sequencing ability, mnemonic strategies, and speed of item identification.[13] Ordinarily administered individually, the digit span test was administered in this study to small groups of students orally via audiotape and paper and pencil. Participants in the test are given a list of series of single digits that the participants listen to; they are then asked to write down in the order (digits forward) or in the reverse order (digits backward) of presentation when all the digits in the series have been presented. At first, the series of digits is rather short (3 digits in a series digit forward; 2 digits in a series digits backward), but gradually the list increases until the series reaches 9 digits in a series (digits forward) or 8 digits in a series (digits backward). A total of 28 of the 30 series were used in the present study.[14] The number of correct series is tabulated to provide a digit span or STM measure. For the purposes of the study, only the digits forward results are included.[15]

Listening Comprehension Test, Paper-and-Pencil Institutional TOEFL

To establish initial proficiency levels, participants were administered the TOEFL listening comprehension component. The listening section of the paper-and-pencil test has three parts.

---

[13] As Klatzky (1980) notes, not all people have the same memory span; one person's may be seven, someone else's may be five. Memory span varies with age as well as among individuals of the same age. These differences among individuals led to the early use of memory span as a measure of mental abilities, as well as of short-term or working memory. Although the span of immediate memory can be said to be about seven words (Miller, 1956), according to Klatzky, it is also seven letters (if the letters do not form words) or seven nonsense syllables. "That is, the memory span is not defined in terms of any particular unit—word, letter, or syllable—but instead seems to be about seven of whatever units are presented. Thus, participants can remember seven letters if they do not form any particular pattern (X, P, A, F, M, K, I); but they can remember many more letters if they form seven words. That is because they are able to recode multiple-letter sequences into single units when the sequences form meaningful words" (Klatzky, p. 91).

[14] The first two series in the digits forward test, with only two digits in each series, were omitted. The test began with the series having three digits.

[15] Test administrators expressed concern that participants may not all have followed the directions as given when doing the digits backward part of the test.

In Part A, examinees listen to 30 brief conversations, each followed by a multiple-choice question. The questions test understanding of the focus and/or purpose of the exchange. Parts B and C consist of five longer conversations or monologues followed by three to five questions. There are 20 items in Parts B and C combined.

Computer-based Test of Listening Comprehension

Eight listening comprehension lectures, with six to eight related items per text, were designed for the study. The sets were modeled on the minitalks currently used in the computer-based TOEFL.[16] In order to investigate the effect of notetaking in the context of lecture length and topic, the following specifications were included in the test design: lecture length: four lectures were “short,” similar in length to the current TOEFL minitalk. These short talks averaged 2 ½ minutes (ranging from 2'19" to 2'45"). Four lectures were “long,” twice the length of the short talks. These long talks averaged 5 ¼ minutes (ranging from 5'07" to 5'29").

Topic: Two of the major content categories of the current computer-based TOEFL were included, the two deemed to be the most different or distinct from each other.[17] Four lectures (two long and two short) were based on topics in the arts and humanities category, and four (two long and two short) in the physical sciences category. Topics in the arts and humanities included the following: a comparison of features of the Renaissance, Baroque, and Neo-Classical styles of art (hereafter Baroque, long); a discussion of the Dada movement in 20th century abstract art (hereafter Dada, long); analyzing works of art in terms of visual elements (hereafter Form, short); a description of techniques used in printmaking, the basic process, creating multiples, the difference between relief and intaglio (hereafter Prints, short). Topics in the physical sciences category included factors that lead to irregular land surface formations, land subsidence, karst topography (hereafter Karst, long); causes of erosion to desert land forms (hereafter Deserts, long); an explanation for recent changes in Louisiana wetlands (hereafter Wetlands, short); and a

[16]In the current computer-based TOEFL, the minitalk is a monologue representing a short segment of an academic lecture in one of four major content areas: arts and humanities, social science, life science, or physical sciences. Each talk averages 2.5 minutes and is followed by 6 selected-response items. Selected-response item formats include multiple choice, multiple selection multiple choice (requires selecting more than one answer choice), order/match (requires moving statements or phrases into a specified order or category), and visual (requires selecting a response based on a visual in the question or in the options). Items in the current test are designed to test general comprehension of the talk: the main idea or gist of the talk and supporting details. Items testing minor or incidental details are avoided.

[17]The study was limited to two content categories to keep the study manageable logistically, including both the number of participants and the length of the test for any individual participant.

discussion of hydroelectricity as an alternative energy source, how it works, its limitations and advantages (hereafter Hydro, short).

Table 2 depicts the actual length of each of these lectures, as well as the number of words.

Table 2

Length of Each Lecture – Long and Short

	Arts and Humanities			Physical Sciences		
	Topic	Minutes	Words	Topic	Minutes	Words
Long	Baroque	5' 29"	748	Karst	5' 18"	792
	Dada	5' 22"	848	Deserts	5' 07"	831
Short	Form	2' 31"	365	Hydro	2' 36"	377
	Prints	2' 19"	403	Wetlands	2' 45"	422

Item Type: Each of the eight sets included a main idea (MI) item type (as in the current computer-based TOEFL). Because the current TOEFL does not test facts or details that might be easily forgotten without notetaking, the specifications for testing details were broadened to include three types of detail questions that could now be investigated with notetaking as a variable in the study. The three types added were supporting information (SI), details (D), and minor details (MD). SI items tested broader concepts related to the main idea, generally requiring integration of information presented in the talk, either explicitly or implicitly. D items tested key points in the talk, presented with some redundancy. MD items tested specific details, such as names and dates, presented with limited redundancy.

Information Type: As in the current computer-based TOEFL, items tested information explicitly mentioned in the talk (EX) and information not explicitly mentioned (NEX), but implied and intended to be inferred.

Response Type: The selected-response item types included in the study were similar to those in the current computer-based TOEFL: multiple-choice (MC) items, order/match (O/M) items, and multiple selection multiple choice (MSMC) items. Visual-response items were not included in the study because it had been decided that no content visuals were to be used in the talks. In addition to the MC, O/M, and MSMC selected-response items, the study included a constructed-response (CR) item type. These responses were limited to one word or a short phrase, which participants typed into a box on the screen.

Number of Items: Each short lecture was followed by six items (five selected-response and one constructed-response). Each long lecture was followed by eight items (six selected-response and two constructed-response).

Table 3 shows the item distribution in the eight sets for each topic.

Table 3

Computer-based Test Format

Short sets - six items:

Item Type	Information Type	Response Type
Main Idea	Not Explicit, but Implied	Multiple Choice
Supporting Information	Explicit or Not Explicit, but Implied	Multiple Choice
Detail	Explicit	Multiple Choice
Detail	Explicit	Order/Match or Multiple Selection Multiple Choice
Detail	Explicit	Constructed Response
Minor Detail	Explicit	Multiple Choice
Long sets - eight items (six as above and two as below)		
Supporting Information	Not Explicit, but Implied	Multiple Choice
Minor Detail	Explicit	Constructed Response

The computer-based test served as the testing environment. During oral presentation of the lecture, a context visual appeared on the screen. To establish setting, the context visual depicted a professor and several students in a classroom setting. The items were presented both orally and printed on the screen. To respond, participants were required to click on the correct answer choice (multiple-choice items), click and move options to the appropriately marked space (order/matching items), or to type in a short answer (constructed-response items). Participants were informed that they would not be penalized for spelling or grammar errors on the constructed-response items. That is, a spelling close enough to make a word recognizable was accepted (e.g., renasance, runessance, renissance, and rennascence were all accepted as spellings for Renaissance). However, a gross misspelling that rendered the word unrecognizable or in a different form was not accepted (e.g., renacentism, renacissm, renainess, and renasentist were unacceptable as spellings for Renaissance).

Before beginning the test, participants were required to work through a brief set of tutorials providing information on how to use the computer. These included general information about the test format, how to adjust the volume, how to answer the questions, and several practice items.

During the test, each participant listened to four talks (two short and two long) and answered the related 28 items. Participants were instructed to take notes during two of the talks (one long and one short) and were not permitted to take notes during the other two talks (one long and one short). Specific instructions were given before each pair of talks regarding notetaking. That is, immediately before the pair of lectures (one long and one short) on which notetaking was permitted, participants received instructions that they could take notes; immediately before the pair of lectures (one long and one short) on which notetaking was not permitted, participants received instructions that they could not take notes. Test administrators distributed and collected the paper used for notetaking at appropriate times for each participant.

The listening material was presented only once. Participants were given 30 minutes to answer all the questions (excluding time spent listening to the lecture).

In total, there were 16 different forms of the computer test. Forms 1-8 were on the content category of arts and humanities; forms 9-16 were on the content category of physical sciences. On forms 1-4 and 9-12, notetaking was permitted on the first two lectures, but not on the last two. On forms 5-8 and 13-16 notetaking was permitted on the last two lectures, but not on the first two. Forms 1, 5, 9, and 13 had lectures in the order: short, long, short, long; forms 2, 6, 10, and 14 had the lectures in the order: long, short, short, long; forms 3, 7, 11 and 15 had the lectures in the order: short, long, long, short; forms 4, 8, 12 and 16 had the lectures in the order: long, short, long, short. Participants were randomly assigned to forms. (See section on results, distribution of participants across test forms.) Computer instructions informed participants to raise their hands to receive notetaking paper when they were beginning to work on a section on which notetaking was allowed. Computer instructions also informed participants to raise their hands when they finished those lectures so the notetaking paper could be collected before the participants continued on sections on which no notetaking was permitted.

The computer test was administered in computer laboratories of the five participating universities. Aggregate percent correct scores (across all item types, information types, and response types) on each of the computer-based subtests were used in the statistical analyses.

Debriefing Questionnaire

At the end of the computer test, a computer screen thanked the participants for their participation and told them that the administrator would give them a brief questionnaire about their experiences with taking notes during the test. The debriefing questionnaire was modeled closely on Hale and Courtney's (1994) survey questionnaire so direct comparisons could be made with Hale and Courtney's results. The instrument consisted of a total of 22 items using a 5-point Likert scale for responses, with 5 = agree strongly, and 1 = disagree strongly. A copy of the instrument is found in Appendix C.

Statistical Analyses

Chi-square analyses were conducted to confirm that participants in the study were appropriately distributed across topics by field of study, gender, institution, and region of the world. Chi-square analyses were also conducted to analyze the debriefing questionnaires.

In terms of addressing the primary research questions, the data were analyzed by a 2 x 2 x 2 analysis of variance with repeated measures (ANOVA-R) for two factors. The between subjects factor was the topic of the minitalk. The topics were classified into two main categories, arts/humanities and physical sciences. The within-subjects factors were length of the minitalk and notetaking status. The minitalks were classified as either being short (approximately 2.5 minutes) or long (approximately 5 minutes). The two levels of notetaking included listening with notetaking allowed and listening with notetaking disallowed. When notetaking was allowed the participants were subsequently allowed to use their notes when completing the short test following the minitalk. The results of both the main and interaction effects will be reported. Tests of simple effects were performed to interpret statistically significant interaction effects. This is the most frequently used method of interpreting interaction effects. In addition, separate error terms were used in the simple effects analyses (Maxwell & Delaney, 1990). An alpha level of $p < .05$ was used to determine statistical significance.

In addressing the secondary research questions, the data were analyzed by incorporating two additional between-subjects factors into the model described above. To facilitate interpretation and to consider the effects of each factor separately, a separate analysis was conducted for each factor. The two additional factors were overall English listening proficiency, as measured by the listening section of a disclosed form of the paper-and-pencil (Institutional) TOEFL, and short-term memory, as measured by the digit span test. These two factors were included in the analysis as between-subjects factors by using a median split to form two groups for each variable. The median was computed utilizing all of the 234 participants who had data available on each variable (paper-and-pencil TOEFL and digit span memory test), N = 227. The two groups were defined as those students with scores at or above the median and those with scores below the median for each of the variables.

Results

Reliability of the Computer-based Test

Table 4a

Reliability and Item Characteristics by Test Topic for Arts and Humanities Topics

Topic	Item	Item Difficulty	Item Discrimination	Cronbach's Alpha
Baroque	1	.73	.27	
	2	.41	.16	
	3	.13	.14	
	4	.21	.34	
	5	.40	.17	
	6	.15	.26	
	7	.61	.07	
	8	.30	.28	.46
Dada	1	.35	.18	
	2	.43	.13	
	3	.42	.01	
	4	.24	.25	
	5	.45	.11	
	6	.43	.26	
	7	.18	.05	
	8	.26	.11	.33
Form	1	.40	.27	
	2	.72	.22	
	3	.25	.34	
	4	.54	.33	
	5	.22	.24	
	6	.53	.24	.52
Prints	1	.72	-.02	
	2	.36	.20	
	3	.16	.39	
	4	.13	.27	
	5	.23	.15	
	6	.35	.20	.39

Table 4b

Reliability and Item Characteristics by Test Topic for Physical Sciences Topics

Topic	Item	Item Difficulty	Item Discrimination	Cronbach's Alpha
Deserts	1	.57	.22	
	2	.65	.15	
	3	.21	.15	
	4	.17	.04	
	5	.39	.24	
	6	.17	.28	
	7	.26	.26	
	8	.32	.32	.44
Karst	1	.64	.39	
	2	.40	.04	
	3	.61	.39	
	4	.21	.05	
	5	.25	.03	
	6	.62	.13	
	7	.39	.03	
	8	.13	.03	.32
Wetlands	1	.49	.06	
	2	.67	.31	
	3	.32	-.08	
	4	.46	.15	
	5	.45	.12	
	6	.47	.16	.27
Hydro	1	.30	.31	
	2	.33	.23	
	3	.47	.30	
	4	.47	.33	
	5	.47	.24	
	6	.32	-.02	.46

Tables 4a and 4b contain the reliability coefficients and item characteristics for each test used for the arts and humanities and physical sciences topics, respectively. The reliability coefficients are relatively low for all tests, ranging from .27 to .52. However, given the brevity of the tests (six or eight items) these reliability coefficients should not be considered surprising.

The item difficulties represent the proportion of examinees responding correctly to the items. For this sample, the items would be considered in the moderate to difficult range, with the majority of the item difficulties at or below .50. The item discriminations represent the corrected item-total correlations. The corrected item-total correlation is the point-biserial correlation coefficient between the individual item score and the sum of the scores on the remaining items. The item discriminations are quite low. However, given the small number of items and the difficulty of some of the items, these values might be expected.

Reliability of the Short-term Memory Test

The reliability of the short-term memory test, the digits forward test of the WAIS-III, consisting of 14 items, was examined by computing a split-half reliability coefficient and Kuder Richardson formula 20. The split-half reliability coefficient with the Spearman-Brown correction ($r = .81$) was based on an odd-even split. The items are designed to get progressively more difficult with consecutive pairs being similar with respect to difficulty. The measure of internal consistency was computed using the Kuder-Richardson formula 20 (KR-20) and was found to be .75. Given the brevity of the test, the moderate to high reliability coefficients would seem to suggest that the reliability of this test for this particular group was adequate.

Distribution of Participants Across Topics

Chi-square analysis of the participant data revealed that the participants were equally dispersed across the arts and humanities and physical sciences topics by field of study (Table 5), by gender (Table 6), by institution (Table 7), and by region of the world (Table 8).

Table 5 shows that the fields of study of the participants were about equally distributed across the arts/humanities and physical sciences topics, although there were more engineering, physical sciences, mathematics, and computer science majors than from any other fields of study represented. Also, relatively few arts/humanities and social sciences majors were represented.

Table 5

Chi-Square Analysis for Field of Study by Topic

Field of Study	Arts/Humanities	Physical Sciences	Total
Eng/Math/CS/PS	34 (18.3%)	31 (16.7%)	65 (35.1%)
Business	32 (17.3%)	19 (10.3%)	51 (27.6%)
Humanities	14 (7.6%)	15 (8.1%)	29 (15.7%)
Social Sciences	6 (3.2%)	8 (4.3%)	14 (7.6%)
Other	13 (7.0%)	13 (7.0%)	26 (14%)
Total	99 (53.5%)	86 (46.5%)	185 (100%)

$\chi^2 = 7.14$ ($df = 4$), $p = .623$ (ns)

Table 6 shows that the genders of the participants were about equally distributed across the arts/humanities and physical sciences topics, although there were fewer females overall.

Table 6

Chi-Square Analysis for Gender by Topic

Gender	Arts/Humanities	Physical Sciences	Total
Female	46 (22.1%)	38 (18.2%)	84 (40.3%)
Male	64 (30.8%)	60 (28.9%)	124 (59.7%)
Total	110 (52.9%)	98 (47.1%)	208 (100%)

$\chi^2 = .219$ ($df = 1$), $p = .89$ (ns)

Table 7 shows that the institutions of the participants were about equally distributed across the arts/humanities and physical sciences topics.

Table 7

Chi-Square Analysis for Institution by Topic

Institution	Arts/Humanities	Physical Sciences	Total
BYU	27 (12.6%)	22 (10.3%)	49 (22.9%)
IU	25 (11.7%)	24 (11.2%)	49 (22.9%)
SIUC	22 (10.3%)	23 (10.7%)	45 (21.0%)
UAZ.	22 (10.3%)	22 (10.3%)	44 (20.6%)
USC	17 (7.9%)	10 (4.7%)	27 (12.6%)
Total	113 (52.8%)	101 (47.2%)	214 (100%)

$\chi^2 = 1.7$ ($\underline{df} = 4$), $\underline{p} = .79$ (ns)

Table 8 shows that participants from different regions of the world were about equally distributed across the arts/humanities and physical sciences topics.

Table 8

Chi-Square Analysis for Region by Topic

Region	Arts/Humanities	Physical Sciences	Total
Asia	61 (29.6%)	45 (21.8%)	106 (51.4%)
Africa/SubS	3 (1.4%)	2 (1.0%)	6 (2.4%)
S/Latin Am.	22 (10.6%)	30 (14.5%)	52 (25.1%)
Middle East	11 (5.3%)	15 (7.2%)	26 (12.5%)
Europe	12 (5.8%)	5 (2.4%)	17 (8.2%)
Total	109 (52.7%)	97 (47.3%)	206 (100%)

$\chi^2 = 6.57$ ($\underline{df} = 4$), $\underline{p}= .16$ (ns)

Key: Asia (e.g., Korea, China [PRC, Taiwan]), Japan, Indonesia, Burma, Mongolia, Thailand, Vietnam, Laos, Hong Kong; Africa/SubSaharan (e.g., Senegal, Mali, Burkina Faso, Mozambique, Ivory Coast, Kenya, Zimbabwe, South Africa, Guinea); South/Latin America (e.g., Nicaragua, Mexico, Panama, Puerto Rico, Venezuela, Chile, Argentina, Brazil, Colombia, Honduras, Peru); Middle East/North Africa (e.g., Saudi Arabia, United Arab Emirates, Oman, Jordan, Syria, Qatar, Kuwait, Egypt, Libya, Morocco, Ethiopia, Iran, Iraq, Bahrain, Mauritania); Europe (e.g., Spain, Italy, Portugal, Albania, Russia, Turkey, Bulgaria, Romania).

Analysis of Notetaking, Length, and Topic (Primary Analysis)

The analysis used to address the primary research questions revealed significant main effects for the following factors: notetaking, $F(1,211) = 9.91$, $p < .01$; and length, $F(1,211) = 15.02$, $p<.001$. In addition, the following interaction effects were found: notetaking and topic, $F(1,211) = 5.52$, $p<.05$; and notetaking and length of minitalk, $F(1,211) = 36.63$, $p<.001$. The ANOVA-R results are reported in Table 9. (Two hundred and thirteen participants had complete item responses to the computer-based test.) Using eta squared as a measure of effect size, it should be noted that the effects are in the small to moderate range. However, the effect for the notetaking x length interaction could be considered large.

Table 9

Repeated Measures Analysis of Variance of the Effects of Topic, Notetaking, and Lecture Length on Percent Correct Scores

Source	MS	df	F	Eta Squared
Between subjects				
Topic	2778.33	1	2.60	0.012
Error	1066.96	211		
Within subjects				
Notetaking	3291.57	1	9.91**	0.045
Notetaking x Topic	1832.26	1	5.52*	0.025
Error	332.02	211		
Length	3905.82	1	15.02***	0.066
Length x Topic	505.32	1	1.94	0.009
Error	260.06	211		
Notetaking x Length	7993.69	1	36.63***	0.148
Notetaking x Length x Topic	393.92	1	1.81	0.008
Error	218.23	211		

* p < .05, ** p < .01, *** p < .001

The cell means and standard deviations are reported in Tables 10 and 11.

Table 10

Means and Standard Deviations (Percent Correct Scores) for Notetaking Status and Lecture Length by Topic

Topic		NS	NL	NNS	NNL
Arts/Humanities	Mean	44.25	36.73	32.60	34.62
(N = 113)	SD	25.37	20.20	21.29	19.62
Physical Sciences	Mean	48.83	34.50	39.33	41.00
(N = 100)	SD	22.30	19.87	25.02	18.56
Total	Mean	45.93	35.68	35.76	37.62
(N = 213)	SD	23.99	20.03	23.30	19.35

Key: NS = notetaking allowed, short lecture; NL = notetaking allowed, long lecture; NNS = no notetaking allowed, short lecture; NNL = no notetaking allowed, long lecture.

Table 11

Means and Standard Deviations for the Main Effects of Topic, Notetaking Status, and Lecture Length

Effect	Mean	SD	N
Notetaking Status (w)*			
Notetaking allowed	40.81	18.88	213
Notetaking disallowed	36.69	18.72	
Length of minitalk (w)			
Short	40.85	20.13	213
Long	36.65	16.22	
Topic (b)			
Arts and humanities	37.05	16.73	113
Physical science	40.67	15.87	100

*(w) – within-subject factor, (b) – between-subjects factor

Interaction of Notetaking and Topic

The results of the post hoc analysis on the interaction between notetaking and topic are reported in Table 12.

Table 12

Means, Standard Deviations, and ANOVA-R Results for the Simple Main Effects Analysis for the Notetaking and Topic Interaction

Topic	Source	MS	df	F	Eta Squared	Notetaking Status	Mean	SD
Arts/ Humanities	Notetaking	2671.94	1	15.81***	.124	Notetaking allowed	40.49	20.28
	Error	168.96	112			Notetaking disallowed	33.61	17.81
Physical Sciences	Notetaking	50.00	1	.31	.003	Notetaking allowed	41.17	17.25
	Error	162.67	99			Notetaking disallowed	40.17	19.2

***p<.001

Thus, students taking the arts and humanities topics scored significantly higher when notes were allowed than when they were not allowed. However, students taking the physical sciences topics performed no differently when they were allowed to take notes and when they were not allowed to take notes. Based on the eta squared value, the effect of notetaking for the arts and humanities group could be considered moderate to large.

Figure 1 depicts the interaction between notetaking and topic.

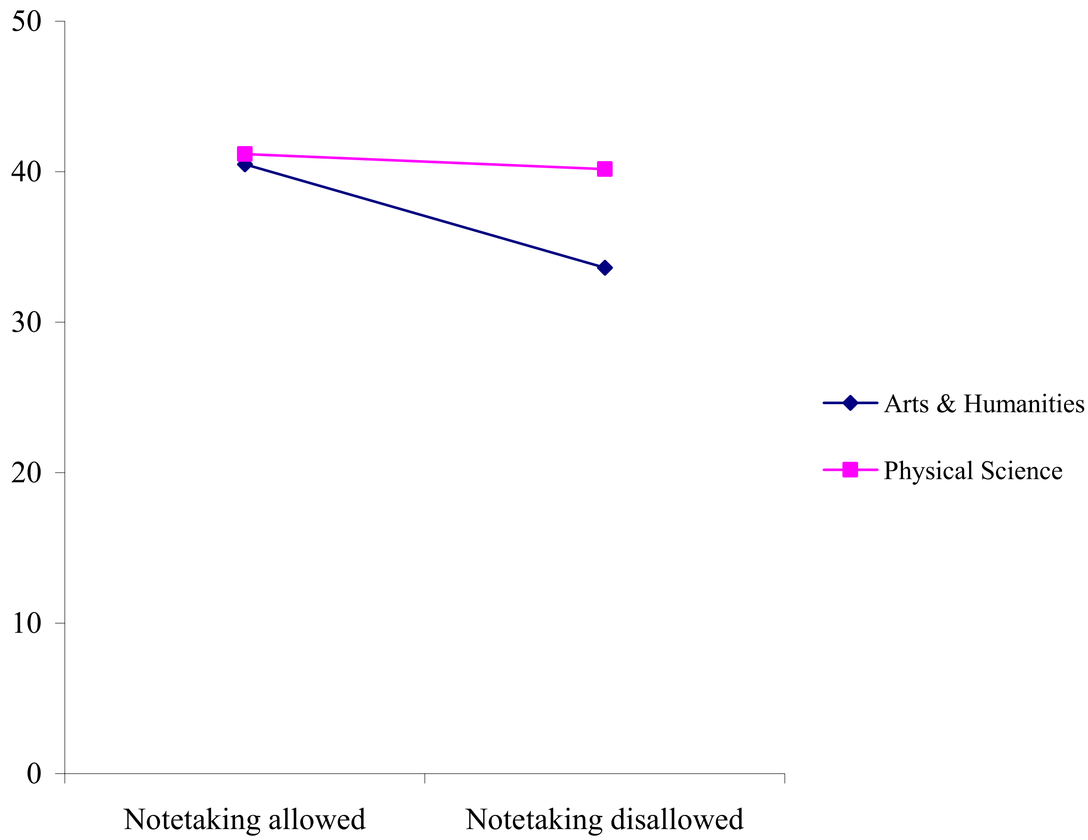

Figure 1. **Means of the Percent Correct Scores for Interaction Between Notetaking and Topic.**

Interaction of Notetaking and Length

The results for the post hoc analysis of the interaction between notetaking and length of minitalk are reported in Table 13.

Table 13

Means, Standard Deviations, and ANOVA-R Results for the Simple Main Effects Analysis for the Notetaking and Length Interaction

Minitalk Length	Source	MS	df	F	Eta Squared	Notetaking Status	Mean	SD
Short	Notetaking	11019.82	1	35.74***	.144	Notetaking allowed	45.93	23.99
	Error	308.35	212			Notetaking disallowed	35.76	23.30
Long	Notetaking	399.43	1	1.60	.007	Notetaking allowed	35.68	20.03
	Error	249.81	212			Notetaking disallowed	37.62	19.35

***p<.001

Thus, on the short minitalks, students did significantly better when they were allowed to take notes than when they were not allowed to take notes. The effect of notetaking in this case could be considered large. On the long minitalks, there was no significant difference when notetaking was allowed and when it was disallowed.

Figure 2 depicts the interaction between notetaking and length of minitalk.

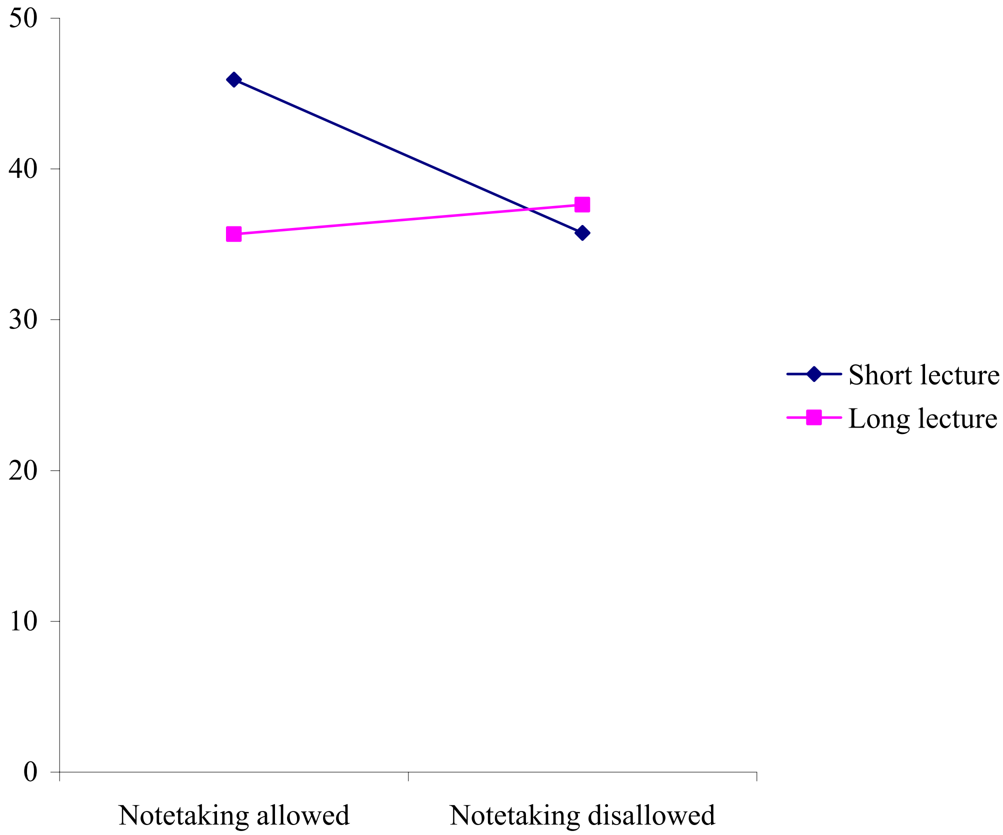

Figure 2. **Means of the Percent Correct Scores for Interaction Between Notetaking and Lecture Length.**

Analysis of Listening Comprehension Proficiency and Short-term Memory Span When Added to Basic Model (Secondary Analysis)

The secondary analyses were conducted in order to examine the additional effects of overall English listening comprehension proficiency and short-term memory on test performance. Each additional factor, based upon a median split, was entered into the 2 x 2 x 2 basic model (notetaking, length, topic) used for the primary analysis and considered separately. The pattern of results was examined for similarities with the results for primary analysis.

Adding Listening Comprehension Proficiency Into the Basic Model

On the listening comprehension section of the paper-and-pencil Institutional TOEFL, students scored a minimum of 31 and a maximum of 66, with a median of 49. Splitting the participants according to the median (the low group ranging from 31 to 48, the high group ranging from 49 to 66) resulted in the following high and low groups according to their listening comprehension proficiency; see Table 14. Two hundred and eight participants had complete data for these secondary analyses.

Table 14

Means and Standard Deviations (Percent Correct Score) by TOEFL Group (Median Split), Topic, Notetaking Status, and Lecture Length

Topic	TOEFL Group		Mean	SD	N
Arts/Humanities	High	NS	56.32	22.69	58
		NL	46.12	19.34	
		NNS	40.23	23.16	
		NNL	41.81	20.08	
	Low	NS	30.45	20.80	52
		NL	25.72	14.94	
		NNS	24.04	15.63	
		NNL	26.44	15.78	
Physical Sciences	High	NS	55.95	20.94	56
		NL	40.85	19.14	
		NNS	49.11	23.45	
		NNL	46.43	17.95	
	Low	NS	37.30	19.41	42
		NL	26.79	18.22	
		NNS	26.59	20.51	
		NNL	33.04	16.29	

Note: NS - notetaking allowed, short lecture; NL - notetaking allowed, long lecture; NNS - no notetaking allowed, short lecture; NNL - no notetaking allowed, long lecture.

When listening comprehension proficiency was added to the basic model, resulting in a 2 x 2 x 2 x 2 model (notetaking, length, topic, listening comprehension proficiency) the results were those reported in Table 15.

Table 15

Repeated Measures Analysis of Variance of the Effects of Topic, Notetaking, Lecture Length, and TOEFL Group on Percent Correct Scores

Source	MS	df	F	Eta Squared
Between Subjects				
Topic	1985.79	1	2.73	.013
TOEFL group	68627.79	1	94.44***	.316
Topic x TOEFL group	271.16	1	.37	.002
Error	726.67	204		
Within Subjects				
Notetaking	3239.58	1	9.93**	.046
Notetaking x Topic	1326.55	1	4.07*	.020
Notetaking x TOEFL group	424.14	1	1.30	.006
Notetaking x Topic x TOEFL group	1026.58	1	3.15	.015
Error	326.37	204		
Length	3440.95	1	13.44***	.062
Length x Topic	380.53	1	1.49	.007
Length x TOEFL group	1281.34	1	5.00*	.024
Length x Topic x TOEFL group	176.07	1	.69	.003
Error	256.09	204		
Notetaking x Length	7464.83	1	33.69***	.142
Notetaking x Length x Topic	351.19	1	1.59	.008
Notetaking x Length x TOEFL group	.00	1	.00	.000
Notetaking x Length x Topic x TOEFL group	270.15	1	1.22	.006
Error	221.58	204		

***p<.001, **p<.01, *p<.05

Thus, in terms of English listening comprehension proficiency as measured by the listening comprehension section of the paper-and-pencil Institutional TOEFL, the results were similar to those reported for the primary analysis with significant main effects for notetaking and length of minitalk, as well as significant interaction effects for notetaking by topic and notetaking by length of minitalk. In addition, the interaction between length of minitalk and TOEFL median group was found to be significant. This effect was examined by analyzing simple main effects. For students with TOEFL listening comprehension scores below the median score there was no difference between the mean percent scores for the long minitalk ($\underline{M}$ = 27.79, $\underline{SD}$ = 12.72) and the short minitalk ($\underline{M}$ = 29.34, $\underline{SD}$ = 14.60). However, for students with TOEFL listening comprehension scores at or above the median, there was a difference for the mean percent correct scores for the long and short minitalks, $\underline{F}(1,113) = 17.70$, $\underline{p} < .001$. The mean for the short minitalks was 50.37, with a standard deviation of 18.96; the mean for the long minitalks was 43.80, with a standard deviation of 15.21.

Adding Short-term Memory Span Into the Basic Model

On the digits forward part of the test of short-term memory, participants scored a minimum of 2 and a maximum of 14, with a median score of 7. Splitting the participants according to the median (the low group ranging from 2 to 6, the high group ranging from 7 to 14) resulted in high and low groups according to listening comprehension proficiency; see Table 16.

Table 16

Means and Standard Deviations by Short-term Memory (STM) Group, Topic, Notetaking Status, and Lecture Length

Topic	STM Group		Mean	SD	N
Arts/Humanities	High	NS	46.92	26.82	65
		NL	40.00	21.10	
		NNS	34.62	22.69	
		NNL	37.31	20.55	
	Low	NS	40.00	22.59	45
		NL	31.39	17.60	
		NNS	29.63	19.43	
		NNL	30.56	17.79	
Physical Sciences	High	NS	47.00	20.40	50
		NL	35.25	19.35	
		NNS	38.00	24.75	
		NNL	42.00	18.34	
	Low	NS	48.96	24.17	48
		NL	34.38	20.17	
		NNS	40.97	25.02	
		NNL	39.32	18.59	

Note: NS - notetaking allowed, short lecture; NL - notetaking allowed, long lecture; NNS - no notetaking allowed, short lecture; NNL - no notetaking allowed, long lecture.

When short-term memory was added to the basic model, resulting in a 2 x 2 x 2 x 2 model (notetaking, length, topic, short-term memory) the results were those reported in Table 17.

Table 17

Repeated Measures Analysis of Variance of the Effects of Topic, Notetaking, Lecture Length, and Short-term Memory (STM) Group on Percent Correct Scores

Source	MS	df	F	Eta Squared
Between Subjects				
Topic	4007.15	1	3.84	.018
STM group	2136.91	1	2.05	.010
Topic x STM group	2616.18	1	2.51	.012
Error	1044.44	204		
Within Subjects				
Notetaking	3160.82	1	9.48**	.044
Notetaking x Topic	1394.24	1	4.18*	.020
Notetaking x STM group	28.84	1	.09	.000
Notetaking x Topic x STM group	66.99	1	.20	.001
Error	333.46	204		
Length	4107.30	1	15.77***	.072
Length x Topic	464.06	1	1.78	.009
Length x STM group	454.15	1	1.74	.008
Length x Topic x STM group	80.58	1	.31	.002
Error	260.40	204		
Notetaking x Length	7293.22	1	32.75***	.138
Notetaking x Length x Topic	289.56	1	1.30	.008
Notetaking x Length x STM group	26.70	1	.12	.001
Notetaking x Length x Topic x STM group	23.87	1	.11	.001
Error	222.68	204		

***p<.001, **p<.01, *p<.05

In terms of short-term memory, the same pattern of results obtained in the primary analysis was obtained when the short-term memory factor was added to the model. There were no significant effects related to the short-term memory grouping.

Results of the Debriefing Questionnaire

For analysis purposes, the 5-point Likert scale used in the debriefing questionnaire, where 5= strongly agree, 4 = agree, 3 = neither agree nor disagree, 2 = disagree, and 1 = strongly disagree, the “strongly agree” and “agree” categories were collapsed into “agree” and the “strongly disagree” and “disagree” categories were collapsed into “disagree.” The proportions of participants agreeing, disagreeing, and neither agreeing nor disagreeing are presented in Table 18.[18] See Appendix C for the entire debriefing questionnaire.

[18]Proportions/percentages may not total exactly 100 because of rounding.

Table 18

Frequency and Percentage Distributions of Participants Agreeing/Disagreeing/Neither Agreeing nor Disagreeing With Debriefing Questionnaire Statements

Statement Number	N	Agree # (%)	Disagree # (%)	Neither Agree nor Disagree # (%)
S1	205	138 (67)	26 (13)	41 (20)
S2	206	155 (75)	23 (11)	28 (14)
S3	205	129 (63)	27 (13)	49 (24)
S4	206	32 (16)	124 (60)	50 (24)
S5	205	92 (45)	71 (35)	42 (20)
S6	204	97(48)	41 (20)	66 (32)
S7	206	94 (46)	46 (22)	66 (32)
S8	206	85 (41)	50 (24)	71 (34)
S9	206	96 (47)	47 (23)	63 (31)
S10	204	128 (63)	31 (15)	45 (22)
S11	205	61 (30)	72 (35)	72 (35)
S12	206	129 (63)	30 (15)	47 (23)
S13	206	42 (20)	117 (57)	47 (23)
S14	202	60 (30)	77 (38)	65 (32)
S15	205	25 (12)	130 (63)	50 (24)
S16	205	31 (15)	127 (62)	47 (23)
S17	203	13 (6)	156 (77)	34 (17)
S18	206	117 (57)	28 (14)	61 (30)
S19	204	85 (42)	47 (23)	72 (35)
S20	206	102 (50)	42 (20)	62 (30)
S21	206	73 (35)	77 (37)	56 (27)
S22	206	75 (36)	74 (36)	57 (28)

Table 19

Chi-Square and Cramer's V Results on Distributions of Participants Agreeing/Disagreeing/Neither Agreeing nor Disagreeing With Debriefing Questionnaire Statements

Statement Number	Chi-Square	p-value	V
S1	108.2	.001	.51
S2	163.0	.001	.63
S3	84.3	.001	.45
S4	69.2	.001	.41
S5	18.4	.001	.21
S6	23.1	.001	.24
S7	16.9	.001	.20
S8	9.0	.011	.15
S9	18.2	.001	.21
S10	80.9	.001	.64
S11	1.2	ns	----
S12	81.6	.001	.45
S13	51.2	.001	.35
S14	2.3	ns	---
S15	88.0	.001	.46
S16	77.4	.001	.43
S17	176.2	.001	.66
S18	58.9	.001	.38
S19	10.9	.004	.16
S20	27.2	.001	.26
S21	3.6	ns	---
S22	2.9	ns	---

Analysis of the Cramer's V tests for the strength of association of the significant chi-square statistics indicates that a moderate to strong association exists between the variables of interest (i.e., examinees' perceptions of the value of notetaking and examinees' agree/disagree/neither agree nor disagree responses). Hatch and Lazaraton (1991) claim that a Cramer's V of .41 indicates that "a strong relationship" (p. 416) exists between the variables of interest.

Analysis of the chi-square data revealed a number of significant differences in participants' responses to the questions on the debriefing questionnaire about notetaking in the testing situation.[19] On only 4 of the 22 questions did participants fail to register differing opinions about whether they (1) agreed, (2) disagreed, or (3) neither disagreed nor disagreed concerning the statements in the survey. Concerning Statement 11, **"It was difficult to locate the information in my notes and then to mark the answers on the computer,"** 30% agreed, 35% disagreed, and 35% neither agreed nor disagreed ($\chi^2 = 1.2$, $\underline{df} = 2$, $\underline{p} = .554$). In response to Statement 14, **"The questions were about things I had written down in my notes,"** 30% agreed, 38% disagreed, and 32% neither agreed nor disagreed ($\chi^2 = 2.3$, $\underline{df} = 2$, $\underline{p} = .322$). With reference to Statement 21, **"I relied on my memory more than my notes to answer the text questions,"** 35% agreed, 37% disagreed, and 27% neither agreed nor disagreed ($\chi^2 = 3.6$, $\underline{df} = 2$, $\underline{p} = .164$). In response to Statement 22, **"I have had training in developing notetaking skills in English,"** 36% agreed, 36% disagreed, and 28% neither agreed nor disagreed ($\chi^2 = 2.9$, $\underline{df} = 2$, $\underline{p} = .225$).

The remainder of the 18 statements elicited significantly different levels of agreement/ disagreement or neutral reaction about the value of notetaking (see Table 18). Agreement was voiced by 67% of participants to Statement 1: **"Taking notes helped me to answer the questions better than if I had not been able to take notes,"** whereas only 13% disagreed, and 20% were neutral ($\chi^2 = 108.2$, $\underline{df} = 2$, $\underline{p} = .001$). Agreement was indicated by 75% of the participants to Statement 2 that **"taking notes made it easier to remember the lecture information,"** whereas only 11% disagreed, and 14% neither agreed nor disagreed ($\chi^2 = 163.0$, $\underline{df} = 2$, $\underline{p} = .001$). In response to Statement 3 that **"I felt more at ease when I could take notes than when I could not,"** 63% agreed, only 13% disagreed, and 24% expressed neither agreement nor disagreement ($\chi^2 = 84.3$, $\underline{df} = 2$, $\underline{p} = .001$). In response to Statement 4 that **"taking notes made answering the test questions more difficult,"** only 16% of the participants agreed, while 60% disagreed and

[19]Participants in the study took notes on two lectures and were prohibited from taking notes on two lectures. It was anticipated that, as a result, participants would think about their notetaking experiences during lecture processing when responding to the questions posed.

24% neither agreed nor disagreed (χ^2 = 69.2, df = 2, p = .001). When asked to register their reactions to Statement 5, **"Taking notes helped me listen carefully to the lectures,"** 45% of the participants agreed, 35% expressed disagreement, and 20% registered neither agreement nor disagreement (χ^2 = 18.4, df = 2, p = .001). Forty eight percent of the participants registered their agreement with Statement 6, **"Taking notes helped me understand the lectures."** Twenty percent disagreed that taking notes helped their understanding, and 32% could neither agree nor disagree that the heuristic of notetaking helped their understanding (χ^2 = 23.1, df = 2, p = .001). Interestingly, when asked if **"taking notes distracted [them] from paying close attention to the information in the lectures"** (Statement 7), 46% agreed, 22% disagreed, and 32% were neutral (χ^2 = 16.9, df = 2, p = .001).

Statements 8 through 17 related to specific aspects of notetaking experience (see Hale & Courtney, 1994), and the following reactions were registered. When asked to indicate their agreement, disagreement, or neutral opinion that they "**wanted more time to review [their] notes before answering the test questions"** (Statement 8), 41% agreed that they did, 24% did not agree that they wanted more time, and 34% indicated neither agreement nor disagreement (χ^2 = 9.0, df = 2, p = .011). When asked to respond to Statement 9, **"I found it difficult to listen to the test questions and at the same time to look for the specific information in my notes,"** 47% of the participants agreed, 23% disagreed, and 31% were neutral (χ^2 = 18.2, df =2, p = .001). A relatively larger percentage of the participants concurred that they **"used [their] notes when answering the test questions"** (Statement 10): 63% agreed, 15% disagreed, and 22% neither agreed nor disagreed that they did so (χ^2 = 80.9, df = 2, p =.001). When asked to agree or disagree or to indicate neither their agreement nor their disagreement with Statement 12, **"Taking notes would have helped me more if I had to answer an essay question,"** 63% agreed, 15% disagreed, and 23% neither agreed nor disagreed (χ^2 = 81.6, df =2, p = .001). **"I had enough time to take as many notes as I wanted"** (Statement 13) elicited agreement from only 20% of the participants; 57% disagreed, and 23% registered neither agreement nor disagreement (χ^2 = 51.2, df = 2, p = .001). When asked whether **"the lectures were too short for notetaking to help [listeners] very much"** (Statement 15), only 12% agreed that they were. However, 63% of the participants disagreed, and 24% were uncertain (χ^2 = 88.0, df = 2, p = .001). When asked whether they remembered enough of the lecture to answer the test question without taking notes (i.e., **"I remembered enough of the lecture to**

answer the test questions without taking notes” [Statement 16]), only 15% agreed that they did. Sixty-two percent disagreed that they remembered enough of the lecture to answer the questions without taking notes, and 23% registered neither agreement nor disagreement with the statement (χ^2 = 77.4, df = 2, p = .001). A low percentage of respondents (06%) agreed with Statement 17, **“The lectures were too easy for notetaking to help me very much,”** whereas a large percentage (77%) indicated their disagreement; 17% were unable to agree or disagree (χ^2 = 176.2, df =2, p = .001).

Five items additional to those constructed by Hale and Courtney (1994) were included in the survey questionnaire. When asked their reactions to Statement 18, **“Taking notes was important for me to get a better score on the text,”** 57% of the participants agreed, 14% disagreed, and 30% were uncertain (χ^2 = 58.9, df = 2, p = .001). The item that related most specifically to the computer-based testing situation asked whether **“taking notes on the sheet of paper supplied to participants interfered with [their] concentration during the computerized lectures”** (Statement 19). Forty-two percent of the respondents agreed that it did, 23% disagreed, and 35% indicated neither agreement nor disagreement (χ^2 = 10.9, df = 2, p = .004).[20] Statement 20 asked whether **“the talks were too long for notetaking to help [them] very much”;** 50% agreed, 20% disagreed, and 30% neither agreed nor disagreed (χ^2 =27.2, df = 2, p = .001).

Interpretation of Findings

Interpretation of Main Findings From Primary and Secondary Analyses

Interpretation of Main Findings from Notetaking, Length, and Topic (Primary Analysis)

Results from the primary analysis show significant interactions between notetaking and length, and between notetaking and topic. These two significant interactions mitigate the significant main effects both for notetaking and for length. The significant interaction between notetaking and length revealed that when the lecture length was shorter (between 2'19" and 2'45") and participants were allowed to take notes, the participants performed far better than they

[20]One wonders if taking notes on the computer screen rather than on a sheet of paper would have elicited similar responses from the participants. Future studies should seek to address this issue by designing studies that allow test takers to take notes on the computer screen via the computer keyboard rather than on a separate sheet of paper using a pen or pencil, as was done in the present study.

did when lecture length was shorter and they were not allowed to take notes or when lecture length was longer (between 5'07" and 5'29") regardless of whether they were allowed or not allowed to take notes. The combination of both shorter lecture length and the ability to take notes worked together to enhance performance. These results run counter to those of Dunkel (1985) as well as of Hale and Courtney (1994), who found no effect for notetaking, although those researchers did not consider the factor of length per se. Given that these results are different from what might have been expected, namely, that notetaking on a longer lecture would have a more positive effect on lecture comprehension and recall than notetaking on a shorter lecture, another study should be undertaken to determine whether this finding can be replicated.

The significant interaction between notetaking and topic revealed that for the physical sciences topics there was no difference between notetaking being allowed and not being allowed. However, for the arts and humanities topics, there was a difference between being allowed to take notes or not, with participants performing far worse on the arts and humanities topics when they were not allowed to take notes. When participants were allowed to take notes on the arts and humanities topics, they performed about the same as participants on physical sciences topics, with or without notetaking being allowed. Evidently, the arts and humanities topics were more difficult than the physical sciences topics for these participants. This may be due to the fact that a larger proportion (35%) of all participants in our study, regardless of what topics they were randomly assigned to, were scientifically oriented (i.e., were either already majoring in or intended to major in engineering, physical sciences, mathematics, or computer science). A smaller proportion (16%) of the sample were participants either already majoring in or intending to major in humanities fields (cf. Table 1). As a result, it seems that participants were advantaged when they were allowed to take notes on the arts and humanities topics, possibly because a large proportion of the participants were not as familiar with those topics as with the physical sciences topics. It may be that participants' greater familiarity with topics in the physical sciences allowed them to recall more of the information without having taken it down in note form.

A reviewer of this report has noted that the lower representation of humanities majors compared to engineering and physical science majors could have resulted in sampling bias. Thus, the interaction of notetaking and topic might be due to this potential bias. Therefore, with a different sampling distribution of major fields of study this result could differ. Further investigation of this should be undertaken.

Interpretation of Main Findings from Adding Listening Comprehension and Short-term Memory to Notetaking, Length, and Topic (Secondary Analyses)

When the participants were divided into high and low groups based upon their scores on the listening comprehension section of the Institutional TOEFL, utilizing a median split, and when this variable was added into the model for analysis, none of the basic results was affected. That is, the same main effects and same interaction effects were significant with and without the addition of the listening proficiency variable. In addition to a significant effect for the paper-and-pencil Institutional TOEFL, notetaking and length continued to be significant main effects, and notetaking and length and notetaking and topic continued to be significant interactions, as discussed above. There was also a significant interaction between length and the Institutional TOEFL. Participants in the low Institutional TOEFL group performed virtually the same whether the lectures were shorter ($\underline{M}$ = 29.59) or longer ($\underline{M}$ = 28.00); however, participants in the higher listening group performed better on the shorter lectures ($\underline{M}$ = 50.40) than on the longer lectures ($\underline{M}$ = 43.80).

Those participants with higher levels of listening ability in English as a foreign/second language were more advantaged when the lectures were shorter, whereas participants with lower levels of listening ability in EFL/ESL were less affected by lecture length. Their ability in English listening comprehension may have been too low to be affected by length one way or another. In other words, a floor effect may have prevailed for the low proficiency listeners. Examinees may need to have a certain level of listening proficiency for the effects of notetaking, length, and topic to significantly affect test performance.

Although individual aptitude traits, such as short-term memory, have been claimed to play a role in cognitive, information processing tasks in both first (Klatzky 1980) and second language (Skehan, 1989), and although some empirical studies have borne this out (Dunkel, 1985), our results found an insignificant effect for short-term memory (as measured by the digit span forward test) on performance on the computer-based test.

Research on information processing since the early work of Shannon and Weaver (1949) has focused on capacity limitation. Many models of information processing attribute this to limitation on short-term memory and attention (Broadbent, 1958). Presumably, the limited capacity of short-term memory affects performance in cognitive tasks like reading or listening comprehension and general problem solving (Newell & Simon, 1972). Gernsbacher (1990), wondering whether and why poorer access to recently comprehended information marks less skilled L1 comprehenders, concluded that one possible explanation is that less skilled comprehenders are plagued by smaller memory capacities. Yet, he found that within the "normal range of adults' comprehension skill" (p. 180) more skilled and less skilled comprehenders are indistinguishable according to traditional short-term memory measures, such as the digit span test of the WAIS. He notes, for example, that no difference was found in the memory span, as measured by the traditional digit span test of the WAIS, of more skilled ($\underline{M}$ = 7.13, $\underline{SD}$ = 1.93) comprehenders and less skilled ($\underline{M}$ = 7.45, $\underline{SD}$ = 2.19) comprehenders. He reasons, therefore, that the less skilled comprehenders were not plagued by smaller short-term memory capacities.

The digit span has been assumed to reflect the storage capacity of short-term memory, and it is a ubiquitous component of intelligence tests (Wechsler, 1944). However, it has not been found to correlate well with performance on such higher level tasks as reading comprehension (Perfetti & Lesgold, 1977) or even the amount of information estimated to be represented in primary or secondary memory (Martin, 1978).

Although early theories viewed short-term memory as a fixed number of slots or bins (Waugh & Norman, 1965), Baddeley and Hitch (1974) have argued that this focus is too much on the storage function of short-term memory and not enough on the processing functions. Hence, they prefer the name "working memory" and have argued for the importance of both the storage and processing functions of the working memory system (Turner & Engle, 1989).

A number of different working memory span tests have been devised to tap both the processing and storage functions of working memory; one that has been widely used with good results is the reading span test (Daneman & Carpenter, 1980). Others include the listening span test (Daneman & Carpenter) and the speaking span test (Daneman & Green, 1986). However, much of the recent research on working memory tests, all of which are designed to tax the processing and storage functions, suggests that the functional capacity of working memory may vary with the processing characteristics of the task being performed (Daneman & Green), and consequently questions the usefulness of a central working memory construct for measuring and understanding individual differences in language performance (Daneman & Tardif, 1987). By contrast, Turner and Engle (1989) argue that working memory is task independent. Carpenter, Miyake, and Just (1994) provide an overview of a number of issues surrounding working memory capacity and capacity constraints.

Thus, short-term or working memory is a complex phenomenon. While the digit span measurement of short-term memory has been traditionally used to assess this capacity in connection with listening comprehension, it is probably not an appropriate measure because it is limited to measuring only the "storage" and not the "processing" function of short-term or working memory. We recommend that further testing of listening comprehension include other measures of working memory.

Some Evidence for the Concurrent Validity of the Computer-Based Test

Results from the secondary analysis showed that the Institutional TOEFL listening comprehension test (the paper-and-pencil test) correlated moderately well with the computer-based test constructed for this study (r = .71, N = 208 participants who had both a computer-based listening comprehension test score and a paper-and-pencil listening comprehension test score).[21] In a sense this significant and substantial correlation validates the computer-based test of listening comprehension, even though the computer-based test focused specifically on listening to short academic lectures and was delivered in a different medium, and even though it contained questions somewhat different from those in the paper-and-pencil test (e.g., different types of multiple-choice items—such as order/match, and open-ended questions—and constructed-response items). Obviously, if one were to administer the listening comprehension section of the Institutional TOEFL test, one would not want to administer what amounts to essentially the same test via computer.

Interpretation of Debriefing Questionnaire Results

Due to the repeated-measures design of the study, all students participated in both the notetaking and nonnotetaking conditions. Moreover, instructions at the end of the computer-based test informed students that they would next be asked to complete a debriefing questionnaire about

[21] TOEFL reports a correlation of .82 between the listening components of the paper-and-pencil traditional test and the computer-based test scores on a concordance sample (TOEFL, 1998, p. 31).

their experiences during the notetaking condition. Still, it is conceivable that respondents may not have been able to focus their responses specifically on their experiences in the notetaking condition as opposed to their experiences in the nonnotetaking condition. However, we do not have evidence to believe that was the case, primarily because participants engaged in both notetaking and nonnotetaking conditions successively, within a relatively short period of time and immediately prior to answering the debriefing questionnaire. Thus, we trust respondents were able to focus on answering the items in terms of their recent notetaking experience on the computer-based test.

Students' responses suggest that they perceived a level of comfort and ease from being allowed to take notes while listening to the lectures, that they believed notetaking aided performance in answering questions about the lectures, and that their recall of information was positively influenced by being allowed to take notes. For example, 67% agreed that notetaking helped them answer the questions better than if they were not able to take notes; 75% agreed that notetaking made it easier to remember the information from the lecture; and 63% felt more at ease when they were allowed to take notes during lecture learning. Participants perceived the positive benefits of notetaking even for the relatively short lectures (2'19" – 2'45"). Sixty-three percent disagreed that the lectures were too short for notetaking to help their performance; only 12% found the lectures too short for notetaking to be of use. It is interesting to note, however, that when asked if the lectures were too long[22] for notetaking to help, 50% agreed they were, whereas only 20% disagreed, and 30% neither agreed nor disagreed. It appears that the notetaking on relatively short lectures was considered more effective than merely listening and trying to recall information heard. In addition, being allowed to take notes seems to have aided the students' processing of the lecture information. Sixty-two percent disagreed that they could remember enough information contained in the minilectures without taking notes. Seventy-seven percent disagreed that the lectures were too easy for notetaking to have helped, and 60% disagreed that taking notes made answering the questions more difficult.

Fewer than half the students perceived that being able to take notes helped them listen more carefully (46% agreement) and also helped them understand the lectures (47% agreement),

[22]One presumes the participants were responding in terms of the longer minilectures that were approximately 5 minutes (5'07" – 5' 29") in length.

and they expressed a sense that taking notes was even somewhat of a distraction from or interfered with listening. Forty-six percent agreed that taking notes distracted them from paying close attention to the information in the lectures, and 42% agreed that taking notes interfered with concentration during the lectures.

Although the majority of participants agreed that they **used** their notes when answering the test questions (63% agreed, 15% disagreed), participants generally also admitted that they had **difficulty** using their notes: 41% agreed that they wanted more time to review their notes before answering the test questions, and 47% found it difficult to both listen to the test questions and look for information in their notes. Moreover, 63% felt that being able to take notes would have been more helpful if the questions had been essay questions rather than the multiple-choice questions and short-answer open-ended questions on the computer-based test.

In sum, participants generally perceived a number of different types of benefits from being able to take notes during lecture listening, but, at the same time, they reported (1) some interference and distraction effects as a result of notetaking and (2) some difficulty in effectively utilizing their paper-and-pencil notes in a computer-based testing situation. It is interesting to note that although most participants perceived a benefit from being allowed to take notes, this was not necessarily reflected in their performance on the computer-based test, especially on the longer minitalks. There may have been a mismatch between students' feelings about the benefit of notetaking and the actual effect of notetaking. Further research is needed to examine this issue.

Relationship of Our Debriefing Questionnaire Results to Those of Hale and Courtney (1994)

Interestingly, the survey findings of the present study are largely in harmony with those found by Hale and Courtney in their 1994 investigation. Hale and Courtney reported percentages of participants "agreeing" with statements similar to those asked in the present study (i.e., statements 1, 2, 3, 4, 5, 6, 8, 10, 13, 14, 15, 17). It should be noted that Hale and Courtney collapsed participants' responses into a 2-point scale (agree/disagree) rather than the 3-point (agree/disagree/neutral) scale used in the present study. Still, participants in the present study registered remarkably similar proportions of agreement with those found by Hale and Courtney. To illustrate, in the present study, 67% of the participants voiced the feeling that taking notes helped them to answer the test questions better than if they had not been able to take notes.

Hale and Courtney reported that 56% of those allowed and 57% of those urged to take notes voiced similar feelings. We found that 75% of the participants agreed that taking notes made it easier to remember the lecture information; Hale and Courtney reported that 77% and 72%, respectively, of those allowed and urged to take notes agreed with the idea. Sixty-three percent of our participants, as compared to 64% and 61% of those allowed and urged to take notes by Hale and Courtney, agreed that they felt more at ease when they could take notes on the minilectures. When asked whether taking notes helped them listen more carefully to the lectures, the percentage of students agreeing (45%) with the idea in the present study was similar to those in the Hale and Courtney study (45% for those allowed to take notes). However, those urged but not required to take notes by Hale and Courtney responded with only a 37% agreement rate to this notion. When asked whether taking notes helped listeners to understand the lectures, similar percentages of agreement were again reported by Hale and Courtney (48% of those allowed to take notes) and the researchers of the present study (48%); only 39% of those urged to take notes by Hale and Courtney expressed agreement with the sentiment. Finally, when asked whether they had enough time to take as many notes as they wanted, 20% of the participants in the present study agreed, compared to 18% and 15%, respectively, of those allowed and urged to take notes in the Hale and Courtney study.

There were, however, some interesting differences between our results and those of Hale and Courtney (1994). For example, 21% of both those allowed and urged to take notes in the Hale and Courtney study found the talks too easy for notetaking to help very much, compared with only 6% in the present study. Similarly, 46% of both those allowed and urged to take notes in the Hale and Courtney study found the talks too short for notetaking to help very much, compared with only 12% in the present study. We attribute these differences to the fact that the talks in the Hale and Courtney study were less than 2 minutes in duration, whereas they were longer in the present study, even in the short condition (2'19" – 2'45"), much less in the long condition, where they were even longer (5'07" – 5'29").

It is interesting to note that participants in the present study were evenly distributed, percentage-wise, when asked whether the questions were about things they had written in their notes, whereas Hale and Courtney (1994) found stronger agreement (i.e., 51% of those allowed and 56% of those urged to take notes). Approximately one third of Hale and Courtney's participants agreed (33% of those allowed and 37% of those urged to take notes) that taking notes made the test more difficult, whereas only 16% of the participants in the present study agreed that taking notes made it more difficult to answer the questions. Finally, while only about half (49% of those allowed and 50% of those urged to take notes) of the participants in the Hale and Courtney study agreed that they actually used their notes when answering the questions, a higher proportion, almost two thirds (63%), said they did in the present study. Moreover, Hale and Courtney's study posed primarily main idea questions, and this could explain why their participants perceived notetaking as less critical or useful.

Although participants said that they used their notes when answering the questions, we do not know whether those who took "higher quality" notes were better able to answer questions than those whose notes were of "lower quality." As Clerehan (1995) notes, "rarely have the notes taken by (L1) students been systematically investigated as objects of interest" (p. 137). The situation is similar for L2 participants in research investigations. In one of the few studies that has examined the content of lecture attendees' notes Clerehan observes that L1 students recorded significantly fewer of the hierarchical structures, particularly the top-level elements, of the lecture.[23] As a result, it seems advisable to examine the quality of the notes taken by the participants in this study to see exactly what was recorded, if anything, and to analyze the relationship between the content of the notes and examinees' performance on the postlecture questions. We cannot simply assume that a notetaker in the study was a "good" notetaker. Neither can we assume that the examinees took notes when they were allowed to. In a study of the effects of required and optional computer-based notetaking on the achievement of mandatory and optional notetaking during a computer-based instruction (CBE) lesson on The Human Heart and Its Functions, Armel and Shrock (1996) found that those required to take notes performed better on the postinstructional quiz than did those permitted to or precluded from taking notes. The required notetaking[24] group scored significantly higher on the posttest than did the optional (notetaking permitted) or the control (notetaking denied) group. Forced notetaking improved achievement on an immediate posttest of

[23]Einstein, Morris, and Smith (1985) similarly found that L1 notetakers recalled many more "high-importance" propositions than "low-importance" propositions; nonnotetakers recalled an equal number of high- and low-importance propositions" (p. 522). The results suggest that "notetaking enhances organizational processing of lecture information" (p. 522). Once again, investigation of the content of examinees' notes would shed light on the kind of notetaking subjects engaged in.

[24]The notetakers used the keyboard to type notes into a designated "Notes" area on their screens. If any participants in the required notetaking group tried to continue the lesson without taking notes, a phrase reminding the participants that they had to take notes appeared on the screen. The instruction would not advance until notes containing at least five words were entered.

information recall. Even the notetaking-as-optional group scored significantly higher than the group that was not permitted to take notes. The researchers also found that those required to take notes took significantly longer than either the optional notetaking group or the control group to complete the instructional program.

In the present study, notetaking was optional; examinees were not required to take notes. Moreover, because it was not part of the design of this study, no one, including the investigators, has examined the quality of the notes taken by the participants. It may, therefore, be necessary to examine the content of the notes under the notetaking-allowed condition to determine what kinds of information examinees included in their notes and what relationship exists, if any, between the content of the notes and subsequent performance on the postlecture test. Study of the content of the notes would further illuminate the findings of the present study. It may also be interesting to investigate whether students take more (and more useful) notes when allowed to take notes on their computer screens. In the present study, examinees listened to a minilecture and took paper-and-pencil notes. Could they take faster and fuller notes if they could use the keyboard to enter their notes? A study of the effectiveness of computer-based notes versus paper-and-pencil notes might shed light on the phenomenon of notetaking in the twenty-first century when students listen to a lecture on screen and take notes simultaneously on screen.

Summary of Main Findings

The following summarize our main findings:

1. A positive effect for allowing notetaking was found, which was not the case in previous experimental work by Dunkel (1985), and Hale and Courtney (1994). The interactive effect of notetaking and length, as well as of topic, has been noted. However, the fact remains that this study provides rare support for the value of notetaking for L2 learners.

2. A positive effect for lecture length was found; shorter lectures produced higher percent correct scores than did longer lectures.

3. However, these two main effects are mitigated by the interaction effects found for the following:

3a. An interaction between notetaking and topic was found; students performed least well on arts and humanities topics when no notetaking was allowed, performed best on arts and humanities topics when notetaking was allowed, and performed virtually the same on physical sciences topics regardless of whether notetaking was allowed or disallowed.

3b. An interaction between notetaking and lecture length was found; students performed best when notetaking was allowed on short lectures, and performed less well when not allowed to take notes on short lectures or when lectures were long, regardless of whether the students could or could not take notes.

4. No differences in the pattern of results were found when listening comprehension proficiency (as measured by the listening comprehension section of the Institutional TOEFL) and short-term memory were added to the equation along with the three main factors of notetaking, lecture length, and topic. However, there was a significant interaction between lecture length and listening comprehension, with participants with lower listening comprehension performing virtually the same whether lectures were short or long, but participants with higher listening comprehension performing far better on shorter lectures than on longer lectures.

Implications of These Findings for TOEFL 2000

1. Allow examinees the opportunity to take paper-and-pencil notes while listening to the computer-based minilectures. Clearly, the finding of a significant interaction for notetaking and length, and for notetaking and topic, as well as the questionnaire data, suggest that examinees feel better when they are allowed to take notes. Allowing notetaking seems to enhance the learners' level of comfort during the minitalk section of the computer test. This perceived comfort of being able to jot down notes while listening to the minitalks may also allow examinees to demonstrate higher levels of performance, because they will not have to rely so heavily on their memories to store all the information heard in minitalks. They can reference their notes to check information asked in the test questions. Furthermore, the face validity of the test should improve somewhat if notetaking is allowed, because university lecturers encourage (and even expect) students to listen and take notes on their lecture presentations. In allowing notetaking, TOEFL 2000 would be reiterating one of the traditional approaches to learning from lectures—that is, listening to the discourse heard and noting down information perceived as important or relevant for recall in an examination.

2. We recommend keeping the current short minitalk format unless further study should determine that test scores on longer minitalks more accurately reflect the construct being measured. Whether TOEFL 2000 should increase the length of the minitalks from 2.5 minutes to 5 minutes or more remains a question for further study. In this study, notetaking helped listeners on the shorter (traditional 2.5 minute) lectures, but not the longer lectures (5 minutes). It should be noted that the lectures that were increased in length also had a concomitant increase in information density of the lecture. It may be that if the lectures had been increased in length without an increase in information load, with the addition of iteration and expansion of information, a slower pace, and addition of backtracking, fillers, and so forth, the students might have done as well on the longer talks as on the shorter. TOEFL might wish to conduct another study that investigates the increase in minitalk length in two different ways: (a) longer minitalks that carry a concomitant increase in information density as a result of the increase in length (which was the case in the present study) and (b) longer minitalks that, while longer, do not increase the information load, by virtue of the addition of iteration, elaboration, etc. Our finding in the current study reinforces the principle that notetaking on shorter minitalks is advantageous, when compared with the first type of longer minitalk. However, it may be that on the second type of longer minitalk students would have done as well or better on the longer lectures as on the shorter lectures, with or without notetaking.

3. Notetaking effects may interact with topic to affect performance; all topics may not behave the same with respect to notetaking effects. Possibly due in part to the backgrounds of examinees, this is worthy of further investigation.

4. One conclusion that could be drawn from the correlation between the computer-based test with its novel item and response types and the traditional paper-and-pencil audiotape Institutional TOEFL test is that TOEFL 2000 might wish to include some of these novel item types (detail and minor detail) and novel response types (constructed response). However, before that possibility is implemented, further study needs to be undertaken examining our results by different item types, because the current study used only an aggregate score encompassing all of the different item, information, and response types.

A follow-up study on the various item types, response types, and information types utilized in this study is warranted, because, as previously stated, the current study used only an aggregate percent correct score encompassing all of these different types. Further research on the various item types and response types included in this computer-based test might end up suggesting that such novel item and response types could eventually be added to an operational version of the TOEFL 2000. We included novel item and response types in this study, and noted that reliability (concurrent validity) did not decrease appreciably compared to that of the traditional paper-and-pencil test.

Further Research

We recommend that the following research be conducted to further pursue the findings of this study.

Additional analyses should be conducted on the data already collected in this study. First, the notes students actually took should be examined to identify the kinds of notes they took. All we can state at this point is that approximately 90% of the students, when allowed to take notes, made some type of written notation. Second, the quality of the notes taken by the students should be analyzed by content analysis. And finally, the relationship between the notes taken and their quality, on the one hand, and test performance, on the other, should be explored. For example, did students who took good notes when they were allowed to perform better than students who either did not take notes when they were allowed to or whose notes were not of high quality?

A follow-up study on the various item, information, and response types utilized in this study is warranted, because, as previously stated, the current study used only an aggregate score encompassing all of these different types.

Further analyses should be conducted on the relationships between the debriefing questionnaire responses and student performance on the computer-based test. For example, did students who perceived that notetaking helped them actually perform better when they were allowed to take notes?

Finally, we recommend that further studies be conducted to examine the question of notetaking directly on the computer compared with the paper-and-pencil notetaking allowed in this study, and to investigate the different ways in which the minitalks might be lengthened and thus have a greater degree of authenticity (i.e., with or without increase in information density).

References

Adamson, H. D. (1993). *Academic competence*. New York: Longman.

Aiken, E. G., Thomas, G. S., & Shennum, W. A. (1975). Memory for a lecture: Effects of notes, lecture rate, and information density. *Journal of Educational Psychology, 67*, 439-444.

Armel, D., & Shrock, S. (1996). The effects of required and optional computer-based notetaking on achievement and instructional completion time. *Journal of Computing Research, 14*, 329-344.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (vol. 8, pp. 47-89). New York: Academic Press.

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (1998). *A TOEFL 2000 framework for testing listening comprehension: A report*. Princeton, NJ: Educational Testing Service.

Benson, M. (1989). The academic listening task: A case study. *TESOL Quarterly, 23*, 421-445.

Berliner, D. (1971). *Aptitude-treatment interactions in two studies of learning from lecture instruction.* Berkeley, CA: Far West Lab for Educational Research & Development. ERIC Document Reproduction Service (No. ED 046 249).

Bilbow, G. T. (1989). Towards an understanding of overseas students' difficulties in lectures: A phenomenographic approach. *Journal of Further and Higher Education, 3*, 85-99.

Broadbent, D. E. (1958). *Perception and communication.* Oxford, England: Oxford University Press.

Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia and aging. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1075-1122). San Diego: Academic Press.

Chaudron, C., Cook, J., & Loschky, L. (1988). *Quality of lecture notes and second language listening comprehension* (Tech. Rep. No 7). Honolulu: University of Hawaii at Manoa, Center for Second Language Classroom Research.

Chaudron, C., Loschky, L., & Cook, J. (1994). Second language listening comprehension and lecture notetaking. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 75-92). New York: Cambridge University Press.

Chaudron, C., & Richards, J. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics, 7*, 113-127.

Chiang, C., & Dunkel, P. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL Quarterly, 26*, 345-74.

Clerehan, R. (1995). Taking it down: Notetaking practices of L1 and L2 students. *English for Specific Purposes, 14*, 137-155.

Cohn, E., Cohn, S., & Bradley Jr., J. (1995). Notetaking, working memory, and learning in principles of economics. *Research in Economic Education, 26*, 291-307.

Crawford, C. C. (1925a). The correlation between college lecture notes and quiz papers. *Journal of Educational Research, 12*, 282-291.

Crawford, C. C. (1925b). The correlation between lecture notes and quiz papers. *Journal of Educational Psychology, 12*, 379-386.

Crawford, C. C. (1925c). Some experimental studies of the results of college notetaking. *Journal of Educational Research, 12,* 379-386.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25*, 1-18.

Daneman, M., & Tardif, T. (1987). Working memory and reading skill re-examined. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 491-508). London: Erlbaum.

DiVesta, F., & Grey, G. S. (1972). Listening and notetaking. *Journal of Educational Psychology, 63*, 8-14.

Dunkel, P. (1985). *The immediate recall of English lecture information by native and non-native speakers of English as a function of notetaking*. Unpublished doctoral dissertation, University of Arizona.

Dunkel, P. (1988). The content of L1 and L2 students' lecture notes and its relation to test performance. *TESOL Quarterly, 22*, 259-281.

Dunkel, P., & Davey, S. (1989). The heuristic of lecture notetaking: Perceptions of American and international students regarding the value and practice of notetaking. *English for Specific Purposes, 8,* 33-50.

Dunkel, P., & Davis J. (1995). The effects of rhetorical signaling cues on the recall of English lecture information by native and nonnative speakers of English as a second language. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 55-74) New York: Cambridge University Press.

Dunkel, P., Mishra, S., & Berliner, D. (1989). Effects of notetaking, memory, and language proficiency on lecture learning for native and nonnative speakers of English. *TESOL Quarterly, 23,* 543-549.

Einstein, G. O., Morris, J., & Smith, S. (1985). Notetaking, individual differences, and memory for lecture information. *Journal of Educational Psychology, 77,* 522-532.

Fahmy, J. J., & Bilton, L. (1990). Listening and notetaking in higher education. In S. S. Anivan (Ed.), *Language teaching methodology for the nineties, Anthology Series, 24*, (pp. 106-126). Singapore: SEAMEO Regional Language Centre.

Fischer, J. L., & Harris, M. B. (1973). Effects of notetaking and review on recall. *Journal of Educational Psychology, 65*, 321-325.

Flowerdew, J. (1994). Conclusion. In J. Flowerdew (Ed.), *Academic listening: Research perspectives (*pp. 290-297). New York: Cambridge University Press.

Frank, B. M. (1984). Effects of field independence-dependence and study technique on learning from a lecture. *American Educational Research Journal, 21*, 669-678.

Freedle, R.F., & Kostin, I. (1999). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing, 16,* 2-31.

Ganske, L. (1981). Notetaking: A significant and integral part of learning environments. *Educational Communication & Technology, 29*, 155-175.

Gernsbacher, M. (1990). *Language comprehension as structure building.* Hillsdale: Erlbaum.

Hale, G., & Courtney, R. (1994). The effect of notetaking on listening comprehension in the Test of English as a Foreign Language. *Language Testing, 11,* 29-47.

Hansen, C., & Jensen, C. (1994). Evaluating lecture comprehension. In J. Flowerdew (Ed.), *Academic listening* (pp. 241-268). Cambridge: Cambridge University Press.

Hartley, J. (1983). Notetaking research: Resetting the scoreboard. *Bulletin of the British Psychological Society, 36,* 13-14.

Hartley, J., & Davies, I. K. (1978). Notetaking: A critical review. *Programmed Learning and Educational Technology, 15*, 207-224.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.

Hughes, C. A., & Suritsky, S. K. (1994). Notetaking skills of university students with and without learning. *Journal of Learning Disabilities, 27*, 20-24.

Hult, R. E., Cohn, S., & Potter, D. (1984). An analysis of student notetaking effectiveness and learning outcomes in the college lecture setting. *Journal of Instructional Psychology, 11*, 175-181.

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (1997). *TOEFL 2000 framework: A working paper*. Princeton, NJ: Educational Testing Service.

Jarvella, R. J. (1970). Effects of syntax on running memory span for connected discourse. *Psychonomic Sequence, 19,* 235-236.

Kiewra, K. A. (1984). The relationship between notetaking over an extended period and actual course-related achievement . *College Student Journal, 17*, 381-385.

Kiewra, K. A. (1987). Notetaking and review: The research and its implications. *Instructional Science, 16*, 233-249.

Kiewra, K. A. (1988). Cognitive aspects of autonomous notetaking: Control processes, learning strategies, and prior knowledge. *Educational Psychology, 10,* 378-86.

Kiewra, K. A. (1989). A review of notetaking: The encoding-storage paradigm and beyond. *Educational Psychology Review, 1*, 147-172.

Kiewra, K. A., Benton, S. L., Christian, M., Kim., S. I., & Lindberg, N. (1989). A more equitable account of the notetaking functions in learning from lecture and from text. *Instructional Science, 18,* 217-32.

Kiewra, K., Benton, S., Risch, N., & Christensen, M. (1995). Effects of notetaking format and study technique on recall and relational performance. *Contemporary Educational Psychology, 20*, 172-187.

Kiewra, K. A., DuBois, N., Christian, D., McShane, A., Meyerhoffer, M., & Roskelley, D. (1991). Notetaking functions and techniques. *Journal of Educational Psychology, 83*, 240-245.

Kiewra, K. A., & Fletcher, H. J. (1984). The relationship between levels of notetaking and achievement. *Human Learning, 3*, 273-280.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal, 29,* 303-323.

King, P. (1994). Visual and verbal messages in the engineering lecture: Notetaking by postgraduate L2 students. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 219-238). New York: Cambridge University Press.

Klatzky, R. (1980). *Human memory: Structures and processes* (2nd ed.). San Francisco: W. H. Freeman & Co.

Martin, M. (1978). Memory span as a measure of individual differences in memory capacity. *Memory & Cognition, 6,* 194-98.

Masson, M. E., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology, 75,* 314-18.

Maxwell, S., & Delaney, H. (1990). *Designing experiments and analyzing data: A model comparison perspective* (pp. 504-507). California: Brooks/Cole.

McKenna, E. (1987). Preparing foreign students to enter discourse communities. *English for Specific Purposes, 6,* 187-202.

Meyer, B.J.F., & Freedle, R. F. (1984). The effects of different discourse types on recall. *American Educational Research Journal, 21,* 121-43.

Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Mishra, S., Ferguson, B., & King, P. (1985). Research with the Wechsler digit span subtest: Implications for assessment. *School Psychology Review, 14,* 37-47.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Palmatier, R. A., & Bennett, J. M. (1974). Notetaking habits of college students. *Journal of Reading, 18,* 215-218.

Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141-183). New York: Wiley.

Rickards, J. P., Fajen, B. R., Sullivan, J. F., & Gillespie, G. (1997). Signaling, notetaking, and field independence-dependence in text comprehension and recall. *Journal of Educational Psychology, 89*, 508-517.

Sattler, J. M. (1981). *Assessment of children's intelligence and special abilities* (2nd ed.) Boston: Allyn and Bacon.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.

Spink, A., & Goodrum, A. (1996). A study of search intermediary working notes: Implications for IR system design. *Information Processing and Management, 32,* 681-695.

TOEFL. (1998). *TOEFL Computer-based TOEFL Score User Guide.* Princeton, NJ: Educational Testing Service.

Turner, M. L., & Engle, R. W. (1986). Working memory. *Proceedings of the Human Factors Society, 30,* 1273-1277.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127-154.

Van Meter, P., Yokoi, L., & Pressley, M. (1994). College students' theory of notetaking derived from their perceptions of notetaking. *Journal of Educational Psychology, 86*, 323-338.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72,* 89-104.

Wechsler, D. (1944). *The measure of adult intelligence.* Baltimore: Williams & Watkins.

Wechsler, D. (1974). Wechsler Intelligence Scale for Children-Revised. San Antonio: The Psychological Corporation, Harcourt Brace.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale (3rd ed.) San Antonio: The Psychological Corporation, Harcourt.

Weener, P. (1974). Notetaking and student verbalization as instrumental learning activities. *Instructional Science, 3*, 51-74.

Appendix A
Biodata Form

Institution____Participant Number__________

ETS Study of Listening Comprehension and Notetaking
Biodata Questionnaire

Full Name (please print)____

Home Country____

Native Language____

Other Languages (in addition to English)____

Age (to nearest year)________________

Sex (check one): ______Male ______Female

Length of Time in the USA: _________Years _________Months

Length of Time Spent Studying English: ________ Years ________Months

Academic Level: ____Pre-university ____Undergraduate ____Graduate ____Other

Field of Study:____

Self Assessment of Listening Comprehension (circle one):

1	2	3	4	5
Very low	Low	Intermediate	High	Very High

Appendix B

Debriefing Questionnaire

Survey Questionnaire

Directions: We'd like to give you the opportunity to give your views about notetaking and the computer-based TOEFL. Read each of the following statements and indicate your agreement or disagreement with the statement. **Circle** the number (5, 4, 3, 2, or 1) that best describes your opinion about the statement.

5 = Agree strongly **4** = Agree **3** = Neither agree nor disagree **2** = Disagree **1** = Disagree strongly

1. Taking notes helped me to answer the questions better than if I had not been able to take notes.	5 4 3 2 1
2. Taking notes made it easier to remember the lecture information.	5 4 3 2 1
3. I felt more at ease when I could take notes than when I could not.	5 4 3 2 1
4. Taking notes made answering the test questions more difficult.	5 4 3 2 1
5. Taking notes helped me **listen** carefully to the lectures.	5 4 3 2 1
6. Taking notes helped me to understand the lectures.	5 4 3 2 1
7. Taking notes distracted me from paying close attention to the information in the lectures.	5 4 3 2 1
8. I wanted more time to review my notes before answering the test questions.	5 4 3 2 1
9. I found it difficult to listen to the test questions and at the same time to look for the specific information in my notes.	5 4 3 2 1
10. I used my notes when answering the test questions.	5 4 3 2 1
11. It was difficult to locate the information in my notes and then to mark the answer on the computer.	5 4 3 2 1
12. Taking notes would have helped me more if I had had to answer an essay question.	5 4 3 2 1
13. I had enough time to take as many notes as I wanted.	5 4 3 2 1
14. The questions were about things I had written down in my notes.	5 4 3 2 1

15. The lectures were too short for notetaking to help me very much. 5 4 3 2 1

16. I remembered enough of the lecture to answer the test questions without taking notes. 5 4 3 2 1

17. The lectures were too easy for notetaking to help me very much. 5 4 3 2 1

18. Taking notes was important for me to get a better score on the test. 5 4 3 2 1

19. Taking notes on a sheet of paper interfered with my concentration during the computerized lectures. 5 4 3 2 1

20. The talks were too long for notetaking to help me very much. 5 4 3 2 1

21. I relied on my memory more than my notes to answer the test questions. 5 4 3 2 1

22. I have had training in developing notetaking skills in English. 5 4 3 2 1

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
Email: toefl@ets.org
Web site: http://www.toefl.org

73313-010266 • Y82E.75 • Printed in U.S.A.
I.N. 990631