TEST OF ENGLISH AS A FOREIGN LANGUAGE

Research Reports

REPORT 66
AUGUST 2001

Effects of the Presence and Absence of Visuals on Performance on TOEFL CBT Listening-Comprehensive Stimuli

April Ginther

ETS Educational Testing Service

Effects of the Presence and Absence of Visuals on Performance on TOEFL® CBT Listening-Comprehension Stimuli

April Ginther

Educational Testing Service
Princeton, New Jersey

RR-01-16

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

Abstract

This study was conducted in order to begin to understand the effects of providing different types of visuals -- such as pictures, diagrams, and still photos of the speakers -- with items on the computer-based Test of English as a Foreign Language (TOEFL®). A nested cross-over design (subjects nested in proficiency, level, and form) was used to examine the effects of language proficiency (high or low), still photos (present or absent), and type of stimuli (dialogues/short conversations, academic discussions, mini-talks with context visuals, mini-talks with content visuals) on performance on TOEFL multiple-choice items. Three two-way interactions were significant: proficiency by type of stimuli, type of stimuli by visual condition, and type of stimuli by time. The weakest of these interactions, type of stimuli by visual condition, was the most interesting; it indicated that the presence of visuals results in facilitation of performance when the visuals bear information that complement the audio portion of the stimulus. The interaction effect between type of stimulus and time suggests that the trend toward the use of longer stimuli is not without consequences, and these consequences may extend to testing situations. Responses to a series of questionnaires indicate that the majority of the subjects preferred the stimuli that were accompanied by visuals.

Key words: Listening comprehension, Computer-based language testing, Multimedia, Context, Content.

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

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

A continuing program of research related to the TOEFL test is carried out under the direction of the TOEFL Committee of Examiners. Its 13 members include representatives of the TOEFL Board, and distinguished English as a second language specialists from the academic community. The Committee meets twice yearly to oversee the review and approval of proposals for test-related research and to set guidelines for the entire scope of the TOEFL research program. Members of the Committee of Examiners serve three-year terms at the invitation of the Board; the chair of the committee serves on the Board.

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

Current (2001-2002) members of the TOEFL Committee of Examiners are:

Lyle Bachman	University of California, Los Angeles
Deena Boraie	The American University of Cairo
Micheline Chalhoub-Deville (*Chair*)	University of Iowa
Jodi Crandall (*Ex Officio*)	University of Maryland, Baltimore
Catherine Elder	University of Auckland
Glenn Fulcher	University of Surrey
William Grabe	Northern Arizona University
Stan Jones	Carleton University
Keiko Koda	Carnegie Mellon University
Richard Luecht	University of North Carolina at Greensboro
Terry Santos	Humboldt State University
Merrill Swain	The University of Toronto
Richard Young	University of Wisconsin-Madison

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

Email: toefl@ets.org

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

Acknowledgments

This study could not have been completed without the support and assistance of many individuals both at Educational Testing Service (ETS®) and at Purdue University. The Test of English as a Foreign Language (TOEFL®) Committee of Examiners' approval of the proposal and financial commitment to the project made the study possible.

Some individuals deserve special mention here because of their special contributions. In particular, Louis Mang and Mike Ecker at ETS authored and programmed the experimental stimuli. Joyce Blanchette, Mark Tolo, and Susan Nissan of GPE Assessment wrote and reviewed the stimuli scripts. Kentaro Yamamoto at ETS provided the experimental design. Kathleen Sheehan of ETS reviewed earlier versions of the report. My thanks as well to Phil Oltman, Ken Sheppard, and Larry Stricker for their helpful comments and assistance throughout the process.

At Purdue University, Heather Allen, Genick Blaise, Jennifer Gerrity, Krishna Prasad, and Alan Redmon, graduate students in the English Language and Linguistics Program, ran subjects. Diana Woollen of the English as a Second Language Program provided administrative assistance. James O'Malley of the Statistical Consulting Service provided invaluable consulting expertise and conducted the statistical analyses of the performance data. My thanks as well to the subjects who participated in the study.

Table of Contents

List of Tables

Page

List of Figures

Page

Introduction

In an effort to use multimedia computer technology to develop a more realistic and valid test of communicative competence, the Test of English as a Foreign Language (TOEFL®) listening-comprehension section now includes visual accompaniments to verbal stimuli. The primary purpose of this study was to examine the effects of the presence or absence of visual accompaniments in association with three different types of listening-comprehension stimuli -- dialogues/short conversations, academic discussions, and mini-talks. Effects of the presence or absence of the visual accompaniments were examined with respect to both subjects' performance on the items and subjects' preference for audio-only versus audiovisual presentation of the stimuli.

Current TOEFL test development efforts involve the production of listening-comprehension items that include still photos, drawings, and pictures. The decision to include visuals on the test was made for several reasons. Presenting examinees with a blank screen in the computer-based testing (CBT) environment was considered inappropriate, and the introduction of visuals was intended to enhance the face validity of the test. Most importantly, item stimuli including visual accompaniments to the audio text are considered better representations of actual communicative situations, so the inclusion of visuals may enhance the measurement of the test-taker's listening comprehension. Although determining precisely how or why the inclusion of visuals effects the measurement of listening comprehension may be difficult, TOEFL CBT listening-comprehension stimuli are different than their pencil-and-paper counterparts, and performance on these items may involve previously untapped aspects of proficiency.

The current study was designed to provide preliminary answers to the following questions:

1. Do subjects perform better on TOEFL test items when visuals accompany the audio text?

2. Does the presence of visuals interact with stimulus type? (If visuals play different roles with different types of TOEFL listening-comprehension stimuli, this result may help identify the functions that visuals can be intentionally designed to perform.)

3. Is an effect on performance related to subjects' English proficiency? (Some theorists have argued that visuals should have the strongest effects on the performance of less-skilled subjects, because the visuals make the task easier. Others have argued that if the complexity of the stimulus is increased with the addition of visuals, less-skilled subjects will be unable to take advantage of their presence and their performance will be debilitated. Experimental results appear to support the former argument. However, "skill" in this context refers to prior knowledge. Can the notion of skill be related to language proficiency?)

4. Do subjects report clear preferences for either the audio-only or audiovisual stimulus conditions?

It is well established in the literature on this subject that the effects of visual stimuli on the comprehension of verbal stimuli depend on several factors, including the task (Weidenmann, 1989), the kinds of visual materials used (Mastropieri, Scruggs, & Levin, 1987; Bauer & Johnson-Laird, 1993), the characteristics of the learners (Hegarty & Just, 1993; Parkhurst & Dwyer, 1983), and on the interaction of these factors (Salomon, 1979; Dwyer & de Melo, 1984; Weidenmann, 1989; Moore & Dwyer, 1994). The consideration of possible interactions among content and visual characteristics creates what Cronbach (1975), in a discussion of the interpretation of interaction effects, refers to as "a hall of mirrors that extends to infinity."

In the case of the TOEFL CBT listening-comprehension items currently under development, uncertainty about the identification of features that might critically effect performance, along with the complexity of the interaction effects, might encourage a conservative experimental approach. Variables of interest could be manipulated and all external variables controlled as carefully as possible. Such experiments, however, would have only limited generalizability, and therefore limited application, to the development of the TOEFL CBT item pool. An alternative approach is to study the effects of visuals in the current CBT item pool, and to identify, albeit broadly, the theoretically important variables in these items. This is the approach that has been adopted for this initial study.

The most frequently used type of visual in the current CBT listening-comprehension item pool provides information about the context in which the verbal exchanges occur. As a group, these visuals are referred to as "context visuals." For example, a photo might portray the participants in a conversation in the setting in which the conversation occurs. Such photos accompany all three types of listening-comprehension stimuli. Broadly, the presence of these visuals serves three purposes: (1) to complement verbally presented information with visually presented information, (2) to set the scene for the verbal exchange, and (3) to cue examinees to a change in speakers in a conversation.

The second type of visual consists of a photo, graph, or drawing that is related to the content of the verbal stimulus. As a group, these visuals are referred to as "content visuals." Content visuals are used *only* as accompaniments to mini-talks; however, their presence is confounded in the TOEFL item pool because content visuals are typically accompanied by one or more context visuals. Furthermore, as TOEFL test developers are reluctant to test "visual comprehension," it may be safe to assume that even when the inclusion of a content visual seems reasonable, a visual depicting the immediate context of the mini-talk may be preferred.

Three combinations of audio and visual conditions represent the CBT listening-comprehension item pool at the present time:

1. dialogues/short conversations with context visuals (a still photo of the speakers)

2. academic discussions with context visuals (still photo(s) of the speakers)

3. mini-talks with context visuals (still photo(s) of the speaker) *and/or* content visuals (photos, diagrams, and/or drawings related to the content of the audio portion of the stimulus)

The majority of the visuals that accompany listening-comprehension audio texts in the TOEFL listening-comprehension pool depict context-related rather than content-related information.

Confounds among types of visuals (context and content) and the type of the audio stimulus are not entirely negative. Current trends reflect test development efforts to capture the complexity of actual situations in which language is used and to exploit the technology available at the present time. It remains, however, that the purposes the visuals *are intended* to serve do not have one-to-one correspondence with particular item types, and there has been no empirically-based investigation of actual as opposed to intended effects.

Fortunately, the three types of listening-comprehension items are accompanied by visuals that provide different kinds of information, thus enabling a quasi-experimental, post-hoc manipulation of the information contained in the accompanying visual stimuli. In order to disambiguate the effects of the type of visual that accompanies mini-talks, separate sets of mini-talks were created for this study: mini-talks with context visuals and mini-talks with content visuals. This differentiation resulted in four types of experimental stimuli:

1. dialogues/short conversations with context visuals (a still photo of the speakers)

2. academic discussions with context visuals (still photo(s) of the speakers)

3. mini-talks with context visuals (still photo(s) of the speaker)

4. mini-talks with content visuals (photos, diagrams, and/or drawings related to the content of the audio portion of the stimulus)

The focus of this study involves a fundamental concern: Does the presence of visuals effect performance on language examinations? The potential of types of visuals to interact with the texts and tasks typical of language examinations is not an important concern if the presence of visuals produces no effect. If the presence of visuals produces a positive effect on performance, there would be a basis for an argument that visual representations of communicative situations play a part in the processing of audio information. However, if the presence of visuals can be associated with debilitation of performance, their presence in high-stakes examinations would be problematic.

Related Literature

As stated above, one of the advantages that computer-based testing is thought to offer the TOEFL examination is the ability to more effectively represent and test communicative competence. In discussions of the nature of communicative competence, emphasis is consistently placed on the language learner's ability to use language effectively within and across different contexts. Nevertheless, in spite of at least 20 years' of concern with the influence of context on language performance, no generally agreed upon characterization of context has emerged. One well known attempt to capture the potentially influential features of context on the performance of language users is provided by Hymes (1974) in the following mnemonic:

Situation: the meaning the participants attribute to the physical and temporal setting, psychological and cultural scene

Participants: the role(s) they think they should take in the interaction

Ends: outcomes and goals they attribute to the exchange

Act sequence: message form and content they think they should attend to

Key: the tone and the manner they think appropriate

Instrumentalities: channels, codes, and registers they think appropriate

Norms: norms of interaction and interpretation they think are called for

Genres: categories of speech events they think they are engaged in

Given this all-inclusive characterization of context as a starting point, it is easy to understand why no consensus has emerged. Researchers are free to emphasize the aspects of context most germane to their particular interests, and the diverse interests of researchers have led to multiple terms and emphases. As Douglas (1997) observes, context has been defined in various ways by researchers and has been associated with such notions as situation, setting, domain, environment, task, content, topic, schema, frames, script, and background knowledge (p.15). In an address to the American Association of Applied Linguistics, Schegloff (1997) adds to the list of potentially relevant features:

> … the ways of formulating the context within which something occurred are multiple. ... [For example,] if one had to characterize what I am doing at the moment, one might say I am presenting a paper, introducing my remarks, reading a text, arguing a point of view, responding to our chair's invitation, gesticulating occasionally and suppressing gesticulation mostly, managing recurrent eye contact with members of the audience, and many others. And a similar stricture can be introduced here: none of these characterizations can get an adequate warrant by saying that it was employed because it is *true* -- even though it *is* true. They are *all true.* (pp.165-166)

But if all aspects of context are, in some sense, true, which aspects of context should be emphasized in representations of those contexts? The TOEFL program's solution has been to emphasize the first two features of Hyme's characterization of context: situation and participants. Representing these features of context can be argued to enhance the "situational authenticity" (Bachman, 1991) of the exam by allowing the examinees to orient themselves to the appropriate domain of language use. It is assumed that having the proper orientation provides information that facilitates comprehension and interpretation. The frequency of context visuals (as compared to content visuals) in the TOEFL CBT item pool appears a reflection of the extent to which both applied linguists and test developers value the representation of context. In actual testing situations, the effects of such representations remain undetermined.

A substantial body of literature exists on the effects of the presence of visuals on written text comprehension (e.g., Mandl & Levin, 1989; Willows & Houghton, 1987; Winn, 1991); however, the focus of this literature is primarily on the effects of visuals that are directly related to the content, as opposed to the context, of written text. Similar to the literature involving investigations of context, the literature that focuses on the effects of content visuals has not led to a consensus or an overarching theoretical framework with respect to the interaction of graphic and textual sources of information (see Mayer, 1993a).

The absence of a clearly applicable theoretical perspective is directly related to the complexity of the combined effects of textual and visual stimuli. Salomon (1989) argues that one of the reasons that no clear theoretical perspective has emerged is that, too often, researchers have attempted to isolate effects, and he states that different effects are produced when different combinations of variables enter into a particular research design. Salomon criticizes the most common research paradigms, explaining that (1) approaches based on empirical findings related to the surface features of materials do not offer a basis for explanation in and of themselves because such findings tend to be inconsistent; (2) approaches that conceptualize independent variables in terms of underlying cognitive processes invite circular reasoning because of the difficulty of distinguishing, for example, the complexity of the stimulus from complexity of the cognitive process invoked; and (3) approaches adopting a purely phenomenological stance based on person-situation interactions make objective classification of stimulus materials impossible.

Salomon (1989) proposes an approach that attempts to explain the effects of pictures in terms of the interactions among surface features of the stimulus materials, underlying cognitive functions, and person and task characteristics. This approach is based on three underlying assumptions:

> (1)... what figures in learning from text and pictures is not the surface features of stimuli or their combinations but the cognitive functions they accomplish ...; (2) texts and pictures can potentially accomplish a variety of cognitive functions as a result of their specific symbolic, informational, and configurational nature ... However, whether different stimulus configurations accomplish these functions in actuality depends not only on the nature of the stimulus materials but also on person, situation, and task variables ...; and (3) a cognitive function accomplished in actuality does not guarantee better learning. [Again,] a variety of person, situation, and task variables [are involved]. (p. 77)

Salomon (1989) continues by stating that at least five clusters of variables are involved in the process of learning from text and pictures. The five clusters are:

> (a) Stimulus variables: the nature of the pictures, the texts, and their interrelations; (b) Cognitive variables: the processes that are called upon, are required, and become employed in actuality; (c) Person variables: the nature of the learner's relevant abilities, prior knowledge, motivations and perceptions of the stimuli, of self, of the situation and the task; (d) the psychological functions accomplished:

> summary, introduction, explication, memory pegs, supplantation, and the like; and (e) Task variables: the nature of the tasks to be accomplished. (p 74)

Salomon (1989) concludes that developing an explanatory model of picture and text comprehension should attempt to take all of these variables into account. Obviously, this could never be accomplished by a single study but rather would require a series of related investigations. Nevertheless, Salomon does present an overarching theoretical concept for the integration of these variables: visual supplantation. He states:

> According to this formulation, explicit pictorial presentations of a process, or of an intermediate state in a process, may overtly model (that is -- supplant) the kind of imagery that learners should have conjured up on their own, assuming of course that such imagery is necessary for the acquisition of the material to be learned. To the extent that learners cannot conjure up such imagery on their own, visual supplantation should facilitate their learning. (p. 77)

It is important to note that in order for visual supplantation to occur and to have a positive effect on performance, there must be a complementary relationship between information presented in the visual portion of the stimuli and the information presented in the audio portion of the stimuli. Thus, the crucial determinant of the effects of the visual stimulus is the relationship between the information contained in the visual and that contained in the verbal (oral or written) text.

Finally, what Salomon (1989) refers to as "person effects" must also be considered. He states:

> ... if particular stimuli are capable of supplanting learners' imagery, then it follows that learners who, for whatever reason, cannot generate their own will benefit from these stimuli; not so for learners who can and do conjure up their own mediating images. Indeed, the results of a series of experiments (Salomon, 1979) consistently showed aptitude-by-treatment interactions: students with a poor mastery of the supplanted skills benefited from the visual supplanting treatment, whereas those with better mastery showed clear signs of debilitation, a possible result of interference. (p. 79)

Salomon's emphasis on the complexity of the possible interaction effects does not comprise a model per se, but it does provide a perspective from which a theoretical framework might eventually be developed.

A series of studies that attempt to unravel the complexities of the potential interactions among text, graphic, and individual variables have been conducted by Mayer and his collaborators (Mayer, 1984; Mayer, 1989; Mayer, 1993a; Mayer, 1993b; Mayer & Anderson, 1991; Mayer & Anderson, 1992; Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Mayer & Gallini, 1990; Mayer & Sims, 1994; Mayer, Steinhoff, Bower, & Mars, 1995). Common to these studies is the examination of methods to improve students' comprehension of scientific text.

These studies develop a perspective called the "generative theory of multimedia learning," which Mayer (1997) summarizes as follows:

> In a generative theory of multimedia learning, the learner is viewed as a knowledge constructor who actively selects and connects pieces of visual and verbal knowledge. The basic theme of a generative theory of multimedia learning is that the design of multimedia instruction affects the degree to which learners engage in the cognitive processes required for meaningful learning within the visual and verbal information processing systems. (p. 4)

Given the focus on understanding scientific explanations, the dependent measure in these studies involves the ability to demonstrate the transfer of learning to the solution of new problems, rather than on recall or comprehension. The success of a student to integrate visual and verbal sources of information is examined in the generation of acceptable and creative solutions in different but related domains.

A consistent finding of these studies is that learners benefit when the presentation of visual and verbal information is contiguous (the "contiguity effect"). Mayer's (1993a) contiguity effect parallels Salomon's (1989) argument that the information presented in text and visual sources must be complementary in order for facilitation to occur. Facilitative effects of the presentation of visual information were reduced when visual information appeared on separate textbook pages or on separate computer screens. Given that the contiguity effect was consistent across text-based or computer-based presentations, Mayer argues that the common research emphasis on the medium of presentation needs to be rethought. The commonalties across media are, most likely, more important than the medium itself.

In further examinations of the contiguity effect, Mayer and Gallini (1990) and Mayer et al. (1995) uncover interactions with what Solomon identified as person effects. In a series of three experiments investigating both the appropriateness and creativity of solutions to problems as related to different multimedia conditions (presentations either possessing or lacking contiguity), Mayer and Gallini found that the effects of multimedia were strong for those subjects without prior knowledge, but weak or nonexistent for subjects with high prior knowledge. That is, when multimedia presentations were contiguous, low-prior-knowledge subjects benefited. When presentations were not contiguous, neither the low-prior-knowledge nor high-prior-knowledge groups benefited. This means that without consideration of presentation effects, the potential facilitation of multimedia presentations might be missed -- and misunderstood.

Mayer et al. (1995) uncover a related interaction effect involving spatial ability. In another series of experiments, they found that subjects with high spatial ability were those who benefited from contiguous multimedia presentations. Low-spatial-ability subjects did not benefit from contiguous multimedia presentations. Mayer (1997) explains this result as follows:

> Students who possess low levels of spatial ability may be less able than high spatial ability learners to take advantage of contiguous presentation of visual and verbal material because they have more difficulty in holding and manipulating the

> visual representation in memory, as is required to integrate the visual and verbal representations. In contrast, students who possess high levels of spatial ability may be more likely than low spatial ability learners to be aided from contiguous presentation because they are more facile at holding and manipulating visual representations in memory … (p. 16)

This series of studies demonstrates that contiguous presentations of visual and verbal information are more effective for students with low levels of prior knowledge and high levels of spatial ability.

One final experiment is perhaps most germane to the present purposes. Mayer (1997) reports the findings of a pilot study that examined the appropriateness and creativity of solutions to problems when the multimedia presentations compared two conditions: text with visuals and narration with visuals. In both conditions, the visuals were animated. Mayer cites Chandler and Sweller (1991), Sweller, Chandler, Tierney, and Cooper (1990), and Baddeley (1992) with respect to split-attention theory and working memory to point out that verbal information may be processed differently when presented as text (visually) or as narration (acoustically). He explains:

> In particular, when text and animation are both presented visually, the learner's visual attention must be split between the animation and the text. When visual attention is overloaded, some of the information may be lost and the process of constructing connections between visual and verbal information will be disrupted. In contrast, when text is presented auditorily and the corresponding animation is presented visually, the learner can process the representation of text within an acoustic working memory and the representation of the animation within a visual working memory, which reduces the load on attention. This situation increases the chances that the learner will be able to construct connections between visual and verbal representations of the causal chain. (p. 17)

Findings of the pilot confirm this prediction. Subjects who were exposed to the animation accompanied by narration produced approximately 50% more creative solutions than did the subjects who were exposed to the animations accompanied by text. While Mayer argues that this finding suggests that narrations may be more effective than texts in multimedia programs, he also argues that researchers' lack of attention to differences in memory load across different types of tasks in multimedia learning environments is a critical lapse.

The majority of the studies examining the effects of visuals on subjects' performance on various experimental tasks use native English speakers as subjects. A notable exception to this trend is a set of related studies conducted by Tang. These studies, which are primarily concerned with the instructional potential of graphics to make text-based information more comprehensible to English as a Second Language (ESL) students, examine 1) the interactions of middle school ESL students with textbook graphics (Tang, 1991), 2) the effects of instruction in the use of graphics on the recall of information by middle school ESL students (Tang, 1994), and 3) the characteristics of graphics in middle school social studies textbooks across four cultures (Tang, 1992).

Tang (1991) employs a variety of ethnographic techniques to examine ESL students' interactions with graphics in the natural setting of two seventh-grade classrooms. She argues that, in spite of the frequency of the presence of graphic accompaniments to text (textbooks were highly illustrated), students tended to pay only nominal attention to graphics. Tang observes, "Few [students] interacted with the graphics they encountered unless the teacher or the assignment required them to study, write about, or reproduce them" (p. 34). Furthermore, students were found to have difficulty exploiting the information presented in graphics effectively. Tang argues that even when explicit instruction in the use of graphics was provided by the teacher, facilitation of understanding associated with the use of graphically presented information occurred only in 10% of the test papers examined. She concludes by suggesting a variety of instructional strategies to help ESL students exploit the knowledge available in graphics in textbooks.

Tang (1992) examines the effects of explicit instruction in the use of graphics on the comprehension and recall of information by middle school ESL students. Using a pretest-posttest nonequivalent-control group quasi-experimental design, Tang found that instruction in the use of graphics facilitated both the comprehension and the recall of information and argues that facilitation can be attributed to an increase in comprehensible input (Krashen, 1985). She argues, "the graphic supplements and enhances the shape of verbally presented text by organizing it, thus giving it additional coherence …" (p. 190) and making it easier to remember.

This finding supports Solomon's idea of visual supplantation and demonstrates that it might be particularly relevant in instructional contexts involving ESL students. Those students who have particular difficulty comprehending information presented in verbal forms may be those who benefit the most from the presence of graphically presented information. However, it appears that explicit instruction is necessary for middle school students to be able to effectively exploit visual sources of information. It should be noted that in the studies conducted by Mayer and his collaborators, subjects did not receive explicit instruction but were able to effectively exploit visual sources of information when that information was contiguous in presentation. This may be a function of the different instructional levels that were involved.

In an examination of widely adopted, seventh-grade social studies textbooks published in Hong Kong, Japan, Mexico, and Canada, Tang (1994) investigated the frequency and type of graphics present in the texts using Mohan's (1986) "knowledge framework," a conceptual framework for integrating language and content. She summarizes Mohan's assumptions and classification system as follows:

> Mohan (1986) suggests that certain knowledge structures are fundamental across the curriculum. They include classification, principles, evaluation, description, temporal sequence, and choice/decision making. We use the same knowledge structures when we classify imports and exports in Social Studies and when we classify flowering and nonflowering plants in Science. Each knowledge structure has distinct linguistic features that set it apart structurally from others. ... However, knowledge structures are built from semantic relations and are not uniquely textual. If they are built on semantic relations, they can exist in the head

> as thought patterns, in textual materials as oral or written discourse, and in graphic and semiotic form. Mohan (1986) has shown that since classification or description or sequence is the same process regardless of content area, knowledge structures are common across content areas, and that graphic representations of each knowledge structure are also common across content areas. (pp. 179-180)

Tang (1994) examined textbooks published in different languages and countries to determine whether graphic representation could also be argued to be common across languages. Using a series of chi-square analyses to uncover differences in frequency in representational graphics (pictures) versus nonrepresentational graphics (maps, diagrams, and graphic organizers), she found only marginally significant differences across texts. The range of representational graphics across texts was from 69% to 77%, while the range of nonrepresentational graphics was from 23% to 31%. When graphics were categorized as classification, description, or sequence (Mohan, 1986), she found the books "remarkably similar;" the chi-square associated with this analysis was not significant at $\alpha = .05$. Graphic descriptions were clearly the predominant type (89% - 94%), with sequence (1% - 7.5%) and classification (1.5% - 5%) decidedly less common. None of the texts used graphics to express principles, evaluation, or choice. While Tang accepts the idea that there are "significant differences between languages in their rhetorical structure" (Kaplan, 1987), she argues that the similarity of graphic representations across texts "indicates that authors of illustrations from different languages and cultures use the same graphic forms and, possibly, the same graphic conventions to represent the same knowledge structures" (p. 183).

Tang argues that the frequency and similarity of graphics in textbooks across the four languages suggests that those students who have participated in formal education in their native countries have been exposed to large numbers of graphics in their first-language contexts. Such exposure may have established knowledge structures or representational conventions that transfer to second-language educational settings. She recommends that ESL specialists train students in graphic literacy in order to activate prior knowledge structures and to facilitate comprehension of English texts in second-language academic contexts.

Finally, a note on the systems used to classify visuals in the studies reviewed above is important. The development of a theoretical framework that would allow consistent categorizations of graphically-presented information is a primary concern of the work of Salomon. The emphasis that his work shares with that of Mayer and Tang is that the classification system is of interest in order to explicate the relationship between visual and textual sources of information and learning.

Because the work of Mayer and his collaborators addresses the relationship between visual and textual sources of information within a paradigm of contiguity, the concern is not so much with the specific role a visual might play, but rather on the extent to which the sources of information are parallel. If the text involves an explication of the sequence of moves that a pump goes through in moving a liquid from one location to another, the visuals are designed to follow this information. Thus, the function of the text determines the functions of the associated visuals.

Tang, on the other hand, is concerned with the provision of a classification system, and her application of Mohan's knowledge framework appears promising, particularly with respect to instructional contexts. However, her concern with classification is not based on the development of a theoretical system: She accepts the viability of Mohan's classifications and uses them to demonstrate that the functions of visuals in textbooks representing very different cultural contexts are remarkably alike. If we accept her provisional claim that the same graphic conventions are used by textbook authors to represent the same knowledge structures, then variation in performance associated with the presence or absence of visuals may be the function of general cognitive strategies rather than language-specific ones.

The emphasis on instruction is problematic for investigations of the effects of visuals on TOEFL listening-comprehension scores. None of the studies reviewed above consider the use of visuals to provide information about the ***context*** of the verbal exchange as opposed to ***content*** information. However, the classification system devised by Levin (1989) may prove useful. According to Levin's classification scheme and terminology, five primary functions of pictures are:

> ... (1) *decoration,* where the pictures are not related to the text and, thus, may be assumed simply to "decorate" it (which includes an author's or a publisher's desire to make a text more attractive, to capture the reader's attention, and to sell more books); (2) *representation*, where the pictures basically overlap with the text ... thereby making the text more concrete; (3) *organization*, where the pictures add structure/coherence to an otherwise poorly or weakly organized text passage; (4) *interpretation*, where the pictures help to make a difficult-to-process text more comprehensible; and (5) *transformation*, where the pictures are designed explicitly to enhance the memorability of a text by transforming it into a more mnemonically powerful form. (p. 85)

While these classifications will need modification in order to derive a conceptual framework for TOEFL CBT listening-comprehension stimuli, it does provide a starting point for a discussion of possible effects. Once again, as the classification system above is intended to capture instructional effects, the classification of context visuals is not addressed. However, in the absence of a more appropriate scheme, the types of visuals associated with TOEFL listening-comprehension stimuli will be provisionally classified according to the specifications above, with context visuals classified as serving a decorative function. Characteristics of pictures, texts, and tasks associated with the three listening-comprehension stimulus types used in the study are discussed below.

Method

Subjects

One hundred and sixty subjects participated in the current study. These subjects were recruited from ESL programs associated with a large state university in the Midwest. Participation was voluntary, and each subject was paid $40.00 after completion of the experiment. Students volunteered by responding to flyers and announcements. All of the subjects

completed the experiment during the spring semester of 1998. Table 1 presents subjects' English-language proficiency, the ESL or student status of subjects within English-proficiency groups, and the gender of the subjects by status.

Given the potential for English-language proficiency to interact with the presence or absence of visuals, recruitment efforts targeted two levels of proficiency, high and low. All 80 subjects classified as high proficiency had been admitted into the university. (The university requires a minimum total TOEFL score of 550 for admission to both undergraduate and graduate programs.) The high-proficiency group consisted of 40 undergraduate and 40 graduate students.

The 80 subjects classified as low proficiency were first recruited from the Married Student Housing ESL Program or were visiting scholars. None of these subjects had been admitted into the university, and none had been in the United States for longer than two months. (It was assumed that these subjects would be lower in English proficiency than the subjects who had been admitted into regular university programs, and the validity of this assumption was confirmed by the significance of the main effect for proficiency.) As virtually every student who was enrolled in the Married Student Housing ESL Program and who had been in the United States for less than two months was recruited and volunteered, there was a deficit of 14 subjects at the low-proficiency level. Fourteen additional subjects who were visiting scholars were then accepted for participation at the low-proficiency level. The nested effect within proficiency -- high proficiency (undergraduate or graduate) and low proficiency (married student housing or visiting scholar) -- is referred to as status and was tested.

Of the 160 subjects, 73 were male and 87 were female. Within the low-proficiency group, females predominated (see Table 1). These subjects may have been primarily female because they may have accompanied their male spouses to the United States where their husbands were enrolled in graduate school. Within all other groups, males outnumbered females.

The range in age of the subjects who were not enrolled in the university was 17 to 52 (md = 29). The age range of visiting scholars was 23 to 43 (md = 33). Ages of undergraduate subjects ranged from 17 – 32 (md = 20), and graduate subjects ranged from 21 – 46 (md = 28).

Subjects were asked to indicate the highest degree they had obtained in their native countries, and these responses are reported in Table 2. Of the subjects who were recruited from the Married Student Housing ESL Program, only nine had not completed a high school degree in their native countries, and an additional 11 had not completed a college degree. The remaining 46 subjects had completed at least a college degree, indicating that, in spite of their assumed lower English proficiency, these were highly-educated subjects.

As expected of the 40 undergraduates in the high-proficiency group, the vast majority (36) reported that they had not completed an undergraduate degree in their native countries. All of the graduates in the high-proficiency group had completed at least an undergraduate degree in their native countries.

The native languages of the subjects in both the low- and high-proficiency groups are presented in Table 3. Native Chinese-, Hindi-, and Indonesian-speaking subjects predominated in the high-proficiency group; native Korean-, Chinese-, and Spanish-speaking subjects predominated in the low-proficiency group. A balanced sample with respect to native language would have been preferable; however, given the dearth of potential subjects at the low-proficiency level, along with the university's prohibition against the recruitment of individuals with specific characteristics, acceptance of the volunteers was the only option available.

Characteristics of TOEFL CBT Listening-Comprehension Stimuli

As stated earlier, there are three types of stimuli in the present configuration of the TOEFL computer-based listening-comprehension item pool: dialogues/short conversations with context visuals, academic discussions with context visuals, and mini-talks with context and/or content visuals. These stimuli differ with respect to the type of accompanying visuals, the length of the stimulus materials, and the number of items presented. The tasks associated with items (most often multiple-choice questions) depend on comprehension of either the "gist" or specific details presented in the verbal portion of the stimulus. When the visuals are related to the content of the message, the relevant information may also be present in the visual part of the message. Items can also be differentiated depending on whether the relevant information is explicitly stated in the verbal message or is nonexplicit and must be inferred. Thus a particular question might be gist/explicit, gist/nonexplicit, detail/explicit or detail/nonexplicit.

Stimuli are followed by six items on the TOEFL examination; they were followed by five items in this study. Listening-comprehension specifications for the TOEFL test require that one item be a gist item and one be based on nonexplicitly stated information. The remaining items can involve any combination of gist/detail and explicit/nonexplicit information. All items are classified according to these specifications, but are not classified in relation to the visuals presented. Verbal portions of the audio text, however, are classified with respect to the presence of context or content visuals.

Dialogues and short conversations. Short conversations and dialogues contain the shortest verbal stimuli, usually a total of five turns by two speakers who are always a male and a female. These stimuli are accompanied by a single still photo of the participants in the setting where the conversation is supposed to be taking place. Short conversations are accompanied by a set of three items, while dialogues are accompanied by a single item. The only information that the visual stimulus provides pertains to the setting of the dialogue.

Given Levin's classification system, discussed earlier, the visuals accompanying dialogues and short conversations could be categorized as either decoration or interpretation. Because the context represented in the visual has no direct bearing on the content of the text, decoration is a likely classification; however, if the context is thought of as providing information that makes difficult-to-process audio texts more memorable by activating scripts associated with particular contexts, the alternative classification would be interpretation.

Academic discussions. Academic discussions can be described as short conversations about a topic related to course material. Verbal stimuli are accompanied by visuals that contain information only indirectly related to the content in the text: Still photos of "speakers" are used to set the context of the communicative event and to mark turns in a conversation. The verbal portions are timed to be about two minutes in length and may involve two or three turns by different speakers. As many as five still photos may appear, but if the identity of the speaker is clearly marked by verbal qualities, such as tone of voice, then only one photo may appear. This is the case even when there is more than a single speaker of the same gender. For example, if a male professor is speaking to two students (one male and one female), one still photo of the scene will suffice if the male voices are distinguished by "maturity."

Given Levin's classification, the visuals associated with academic discussions, like those associated with dialogues and short conversations, can be categorized as either decoration or interpretation. Because the context represented in the visual has no direct bearing on the content of the text, decoration is the likely classification choice; however, if the visuals associated with academic discussions are thought of as making difficult-to-process text more comprehensible by marking turns in the conversation, it is reasonable to classify these visuals as interpretative.

Mini-talks. The verbal stimulus materials classified as mini-talks are based on the broad content areas of physical sciences, social sciences, life sciences, and the arts. The verbal portions of the stimuli last approximately two and one-half minutes and are accompanied by one to four photos.

The function of the context visuals that accompany mini-talk audio texts is to set the scene. As with dialogues/short conversations and academic discussions, these context visuals can be seen to serve a primarily decorative function. Because the speaker never changes, and the visuals never mark a change in speakers, it is unlikely that these visuals serve an interpretative function.

The content visuals that accompany mini-talks (photos, illustrations, and/or diagrams) complement information presented in the text. Furthermore, the content visuals are presented contiguously; that is, the presentation of the content-bearing information occurs simultaneously with the same information in the audio text. The case of mini-talks with content variables marks the only instance in which the visuals can be classified as representational. These visuals may also fit into the category Levin identifies as transformation, because they may enhance the memorability of the audio portions of the stimuli.

Experimental Materials

For the current study, four sets of experimental stimuli were created. Each set was divided into two subsets of stimuli and items so that the sets could be presented with or without visual accompaniments. Two of the subsets were dialogues/short conversations (with or without context visuals) (dialogues and short conversations were combined as they comprised a single relatively short stimuli type); two were academic discussions (with or without context visuals);

two were mini-talks (with or without context visuals); and two were mini-talks (with or without content visuals).

Table 4 displays the characteristics of the stimulus sets as well as the design of the study. Column 1 identifies the stimulus type, subset, and type of visual. Notice that there are four types of experimental stimuli: dialogues/short conversations, academic discussions, and two types of mini-talks -- mini-talks with context visuals and mini-talks with content visuals. Within each stimulus type, there are two subsets of stimuli and associated items (e.g., dialogue/short conversation subset 1 and dialogue/short conversation subset 2). The type of visual identifies whether the visual(s) represented context or content-related information ***when visuals were present***.

Column 2 identifies the subset and associated items by number, and lists item topics. For example, there are five items associated with dialogue/short conversation subset 1, and these items are labeled D/SC 1.1 – D/SC 1.5. Similarly, there are five items associated with dialogue/short conversation subset 2, and these items are labeled D/SC 2.1 – D/SC 2.5. There are two subsets of academic discussions, AD1 and AD2; two subsets of mini-talks with context-related information, MTX1 and MTX2; two subsets of mini-talks with content-related information, MTN1 and MTN2. Each of the eight stimulus subsets was followed by five items.

Design

A nested cross-over design -- which may also be familiar to readers as a partially-crossed Latin Square (see Cochran & Cox, 1957) -- was used for this study. Subjects were nested in form, proficiency, and status. The partially-crossed and nested effects can be understood by considering the number of subjects within the divisions associated with form, proficiency, and status.

A total of 160 subjects participated in the study, and each subject was administered one of 16 TOEFL test forms. The use of 16 forms allowed the effects associated with the order of presentation of stimulus types and visual conditions to be counterbalanced. As column 3 of Table 4 shows, the 80 subjects who were administered forms 1, 2, 5, 6, 9, 10, 13, or 14 were presented with dialogue/short conversation subset 1 in the visual condition (V) and dialogue/short conversation subset 2 in the no-visual (NV) condition. The 80 subjects who were administered forms 3, 4, 7, 8, 11, 12, 15, or 16 were presented with dialogue/short conversations subset 1 in the no-visual (NV) condition and dialogue/short conversations subset 2 in the visual (V) condition. Thus, the presence or absence of visuals is partially crossed.

Within the visual condition, subjects were nested in proficiency -- that is, 40 of the 80 subjects who were administered forms 1, 2, 5, 6, 9, 10, 13, or 14 were from the low-proficiency (LP) group, and 40 were from the high-proficiency (HP) group. Correspondingly, 40 of the subjects who were administered forms 3, 4, 7, 8, 11, 12, 15, or 16 were from the low-proficiency (LP) group, and 40 were from the high-proficiency group.

Within each type of stimulus, five items were accompanied by visuals and five were not. Thus, each subject was administered 10 items in association with each of the four stimulus types,

for a total of 40 items (10 x 4 = 40 items), 20 of which were accompanied by visuals and 20 of which were not. For example, imagine that a low-proficiency subject was administered form 1; he or she would have answered the 40 items presented in the first of the four "Visual Condition" columns in Table 4. Another low-proficiency subject, administered form 3, would have answered the 40 items presented in the third of the four "Visual Condition" columns in Table 4.

The presence or absence of visuals in accompaniment to the stimulus types for any individual subject was determined by the assigned form (see Appendix A, Test Forms, for the actual configuration of each test form). Finally, the stimulus subsets within each type of stimulus were intended to have comparable content and difficulty. All analyses of these data were conducted through an application of the General Linear Model on SAS (SAS Institute, 1998). Using Cronbach's alpha, the estimated reliabilities for the scores on the 10 items associated with each type of stimuli are as follows: dialogues/short conversations = .80; academic discussions = .72; mini-talks (context) = .76; mini-talks (content) = .70. Reliabilities were calculated by collapsing across visual conditions.

Administration of the Experimental Materials and the Questionnaires

In addition to performance data, subjects were administered a series of questionnaires concerning their attitudes toward the presence or absence of visual accompaniments to the verbal stimuli. All of the experimental materials and the questionnaires were administered by computer in the following manner:

Subjects were not alerted to the fact that half of the stimuli would be presented with accompanying visuals and half would not; however, after completion of each set of items, they were asked (1) whether they found the stimuli and items interesting and (2) whether they found the stimuli and items difficult. They were asked these questions with the expectation that a pattern might emerge in association with the presence or absence of visuals; that is, subjects might have indicated that they consistently found the sets of stimuli that were accompanied by visuals more interesting than those that were not presented with visuals. In answering the questions, subjects were asked to indicate the strength of their agreement or disagreement with two statements -- "the questions were interesting," "the questions were difficult" -- on a five-point Likert scale. Responses to these questions are presented separately in Table 6 and Table 7 and will be discussed in the Results section.

After completion of all 40 items, subjects were presented with a questionnaire comprised of six direct statements about their preferences concerning the presence or absence of visuals. Again, subjects were asked to indicate the strength of their agreement or disagreement with statements on a five-point Likert scale. Responses to these questions are presented in Table 8 and will also be discussed in the Results section.

Observations

Subjects were observed as they worked through the experimental materials and any irregularities were noted. One subject was excluded from the analysis because he fell asleep

before he was able to complete the experiment. No other irregularities were noted. The time the subjects required to complete the entire experiment ranged from 59 to 114 minutes.

Exit Interviews

Ten subjects were asked to participate in an exit interview in which they elaborated on the preferences they indicated in Questionnaire III. Four were graduate students, two were undergraduates, three were not enrolled, and one was a visiting scholar. The interviews were all completed in less than 15 minutes and only addressed the preferences that the subjects had already indicated. Responses tended to be general and will be briefly discussed in the Results section.

Results

Performance Data

The model that was tested is presented in Table 5. The dependent variable was each subject's score on the five items presented with each stimulus subset (refer to Table 4). The complexity of the design requires that Numerator Mean Squares and Denominator Mean Squares be presented in the source table (see columns 4 and 5 of Table 5). In addition, eta square (column 8) and partial eta square (column 9) have been included as measures of effect size or the proportion of variance accounted for by population membership (Cohen, 1988).

The significance of the main effect for proficiency generally confirms assumptions about the proficiency levels of visiting scholars and subjects recruited from the Married Student Housing ESL Program, as compared to the proficiency levels of undergraduate and graduate subjects enrolled in the university. However, as stated above, there were different kinds of subjects within the low- and high-proficiency groups; that is, the high-proficiency group consisted of university-enrolled undergraduate and graduate subjects, and the low-proficiency group consisted of not-enrolled ESL students and not-enrolled visiting scholars. This nested effect, referred to as status, was not significant. Because proficiency was involved in a significant interaction with stimulus type, this effect will be discussed further only in the context of the interaction.

The main effect for form was not significant, indicating that the counterbalancing afforded by the use of 16 forms was adequate. The significance of the main effect for subjects indicates that a considerable amount of variation -- actually the bulk of the variance accounted for by the model -- was not accounted for by the other significant main effects and their interactions, and thus remains unexplained. The main effect of stimulus type (dialogues/short conversations, academic discussions, mini-talks with context visuals, and mini-talks with content visuals) was significant, but was involved in a significant interaction with proficiency.

The main effect for time was significant and tested whether there was a practice or fatigue effect associated with order of presentation within type of stimulus. While the effects of presentation associated with visual accompaniments (visual versus no visual) and subset (subset

1 followed by subset 2 versus subset 2 followed by subset 1) were counterbalanced, the integrity of the stimulus type was not violated. That is, subsets within each type of stimulus (whether subset 1 was followed by subset 2, or subset 2 was followed by subset 1) were presented consecutively (see Table 4 and Appendix A). Thus, the presence of a practice/fatigue effect within each type of stimulus could be tested. Again, because the effect of time is involved in a significant interaction with type of stimulus, these effects will be discussed in terms of their interaction.

The main effect for the presence or absence of visuals was not significant. Because the focus of the study was on the possible interaction among levels of proficiency, stimulus type, and the presence or absence of visuals, this was the only three-way interaction that was tested, and it was not significant. However, given the patterns of results, it may have been more germane to test the significance of the three-way interaction among type of stimuli, visual condition, and time. As the effect of time was not anticipated, and therefore not included in the research questions, this interaction remains untested. However, the potential of certain types of visuals to ameliorate the effects of longer stimuli is an interesting possibility and will be briefly discussed in the following section.

While it may appear that the tested model is incomplete because of the absence of the interactions time-by-proficiency and time-by-visual-condition, these interactions are confounded due to the partially-crossed nature of the design. That is, at time 1, half of the subjects were low proficiency and half were high proficiency; correspondingly, at time 1, half of the subjects were presented with stimuli accompanied by visuals and half were not. Because of the confounds, these effects were not tested.

The two-way interaction, proficiency-by-stimulus-type, which averages across visual conditions, was significant. A graph of the interaction is presented in Figure 1. Reading Figure 1 from the top down, note that all of the sets were easier for the high-proficiency group and that the rank of difficulty for both low- and high-proficiency groups remains the same: dialogues and short conversations were the least difficult (low proficiency, $\bar{x} = 3.78$, sd = 1.38; high proficiency, $\bar{x} = 4.87$, sd = .39), followed by mini-talks with content visuals (low proficiency, $\bar{x} = 3.32$, sd = 1.37; high proficiency, $\bar{x} = 4.51$, sd = .85), academic discussions (low proficiency, $\bar{x} = 2.79$, sd = 1.39; high proficiency, $\bar{x} = 4.09$, sd = .94), and mini-talks with context visuals (low proficiency, $\bar{x} = 2.39$, sd = 1.41; high proficiency, $\bar{x} = 3.91$, sd = 1.12). However, the differences in the means for academic discussions, as compared to mini-talks with context visuals, is less pronounced for the high-proficiency group than for the low-proficiency group. If the level of proficiency of the high group were extended, the rank order of academic discussions and mini-talks with context visuals would eventually reverse. This difference can be considered the source of the significant interaction.

The two-way interaction, stimulus-type-by-visual-condition, which averages across levels of proficiency, was significant. A graph of the interaction is presented in Figure 2. The presence or absence of visual accompaniments had virtually no effect when the content of the stimuli materials involved dialogues and short conversations (no visuals, $\bar{x} = 4.31$, sd = 1.15; visuals, $\bar{x} = 4.34$, sd = 1.15). Both the mini-talks with content visuals (no visuals, $\bar{x} = 3.79$, sd = 1.35;

visuals, $\bar{x}$ = 4.03, sd = 1.21) and academic discussions (no visuals, $\bar{x}$ = 3.39, sd = 1.33; visuals, $\bar{x}$ = 3.49, sd = 1.36) were slightly less difficult in the visual condition. Mini-talks with context visuals (no visuals, $\bar{x}$ = 3.24, sd = 1.44; visuals, $\bar{x}$ = 3.06, sd = 1.52) were slightly more difficult when accompanied by visuals.

The two-way interaction, stimulus-type-by-time, was significant. This interaction collapses proficiency and visual conditions. Time was included in the model because subjects were always presented with two subsets of stimuli within each type of stimulus. The effect of time was tested to determine whether a practice effect was associated with order of presentation. A graph of the interaction is presented in Figure 3.

As Figure 3 shows, a practice effect appears associated only with dialogues and short conversations -- the shortest stimuli type. When the dialogue/short conversation sets were presented to subjects, they performed better on the second set (first presentation, $\bar{x}$ = 4.18, sd = 1.33; second presentation, $\bar{x}$ = 4.48, sd = .90). The reverse is true for all of the longer stimuli sets. When the mini-talks with content visuals were presented to subjects, they performed better on the first set (first presentation, $\bar{x}$ = 4.32, sd = .95; second presentation, $\bar{x}$ = 3.49, sd = 1.43). When academic discussions were presented to subjects, they performed better on the first set (first presentation, $\bar{x}$ = 3.54, sd = 1.30; second presentation, $\bar{x}$ = 3.34, sd = 1.39). When mini-talks with context visuals were presented to subjects, they performed better on the first set (first presentation, $\bar{x}$ = 3.48, sd = 1.29; second presentation, $\bar{x}$ = 2.81, sd = 1.59).

It is important to note that in order for an effect to be considered small, the values associated with eta square and partial eta square should be at least .01 (Cohen, 1988). Accepting this criterion means that the effects associated with the interactions between proficiency and visual condition and between stimulus type and visual condition are negligible at best. However, the interaction between stimulus type and time is substantial.

Questionnaire I: How Interesting and Difficult Were the Stimuli Sets?

As mentioned earlier, subjects were asked to indicate the extent to which they found the stimuli sets interesting and difficult. That is, they were asked to indicate their level of agreement or disagreement with two statements -- "the items were interesting" and "the items were difficult" -- after they completed each set. It may have been the case that the subjects would consistently find one of the visual conditions more interesting or more difficult in either the visual or no visual condition. Chi-square analyses of the response patterns associated with the presence or absence of visuals were conducted, and the results of these analyses are reported in Table 6 and Table 7. In addition, **w**, a measure of effect size for use with chi-square (Cohen, 1988), is reported in association with each of the analyses.

Of the 16 chi-square analyses of subjects' responses to questions about interest and difficulty, only the responses to one stimuli were significantly different with regard to the visual condition: academic discussion, subset 2 (see Table 6). The topic of this discussion was engineering materials.

Interestingly, an examination of subject responses in this single significant case reveals that subjects were more likely to agree that the questions were interesting when the stimulus was presented *without* a visual. That is, 21% of the subjects strongly agreed that the questions were interesting when the stimulus was presented without the accompanying visual, as compared to 10% of the subjects who strongly agreed that the questions were interesting when the stimulus was presented with the accompanying visual. In addition, subjects were more likely to adopt a neutral position when the visual was present (43%), as compared to the no-visual condition (21%). This result suggests that subjects find some aspects of particular visuals unattractive and/or distracting. The source of this response pattern, given that there were five still photos presented, is unknown. None of the other analyses was significant.

Questionnaire II: Preference for the Visual- or No-Visual Condition

When directly asked which visual condition they preferred, most subjects indicated that they preferred the stimuli sets that were accompanied by visuals. The questions that comprised this questionnaire, the responses, and results of chi-square analyses are reported in Table 8. In this case, all chi-square analyses conducted were significant, and examination of the response patterns indicates that subjects were more likely to strongly agree that they preferred visual accompaniments (Question 1). Furthermore, subjects were more likely to strongly agree that the visuals made the speakers easier to understand across all stimuli types (Questions 2, 4, 5, and 6). Finally, subjects were more likely to strongly disagree that, in general, the visuals were distracting (Question 3). Effect sizes ranged from .31 to .63.

Discussion

Proficiency by Stimulus Type

The significance of the two-way interaction between proficiency and stimulus type is not unexpected or surprising. The consistency of the high-proficiency subjects' ability to perform at higher levels than their low-proficiency counterparts does provide evidence that the selection procedure, although indirect, was adequate in distinguishing high- and low-proficiency groups. It should be noted that the high-proficiency group performed close to the top of the scale across all of the stimuli sets, suggesting that the sets and items were relatively easy for these subjects, all of whom were enrolled in regular academic courses and had already scored at least 550 on the TOEFL examination.

Although the texts were intended to be comparable with respect to content and difficulty, the mini-talks *Aquifers* and *Amber* (mini-talks with context visuals, subsets 1 & 2) functioned differently than mini-talks *Prairie Dogs* and *Anthropology* (mini-talks with content visuals, subsets 1 & 2). *Prairie Dogs*, a discussion of animal communication, and *Anthropology*, a discussion of early Polynesian trade routes, were more difficult than *Aquifers*, an explication of geological strata and water tables, and *Amber*, a discussion of the characteristics of amber and amber inclusions. It is impossible to determine from this study whether the differences in difficulty are the result of differences in topic or differences in the items in each subset. Unless more is understood about the levels of difficulty associated with text types and items independent

of visual accompaniments to those texts, the interaction between text and visual sources of information will remain ambiguous.

Stimulus Type by Visual Condition

The pattern of results associated with the significant two-way interaction between stimulus type and visual condition confirms the predictions in the literature about the facilitative effects of visuals when those visuals can be argued to perform the function of supplanting subjects' visual imagery. The stimulus conditions which have the greatest potential for complementary mapping of visual and audio portions of the stimulus materials are precisely those in which the means of the subjects were higher. When context visuals accompanied academic discussions, and when visuals bearing content information accompanied mini-talks, the effect of the visuals was slightly facilitative. However, visuals that served only to set the scene, appearing to serve primarily a decorative function, produced either no effect or a slightly debilitating effect. When visuals accompanied dialogues and short conversations, the effect was nonexistent, and when visuals bearing contextual information accompanied mini-talks, the effect was slightly debilitating.

If we accept Tang's (1992, 1994) notion that the knowledge structures represented by visual presentations of information are common across curricular domains -- and appear to be common across languages -- the lack of significance of the three-way interaction with proficiency and the significance of the two-way interaction between type of stimulus and visual condition are not unexpected. The ability to exploit the information represented by visuals may not be a function of language proficiency, but may rather involve graphic literacy and, as Mayer and Gallini (1990) argue, prior knowledge. If this is the case, we would expect both the low- and high-proficiency language groups to be equally aided, or not aided, by the presence or absence of visuals. The significance of the two-way interaction between type of stimulus and visual condition supports this interpretation.

Tang (1992) found with her subjects that explicit instruction in the interpretation of visuals aided middle school ESL students' ability to exploit visually presented information. However, the subjects who participated in this study, whether of high or low language proficiency, were all highly educated. All are assumed to have completed high school in their native countries and only 29% had not completed a college-level degree as well; 61% of the sample had already completed at least an undergraduate degree prior to their residence in the United States. It may be that these subjects are already highly proficient with respect to the interpretation of graphics and, in addition, may have more than adequate prior knowledge of the topics represented in TOEFL listening-comprehension stimulus materials.

However, another interpretation exists for the significance of this two-way interaction. Although it seems safe to assume that the subjects were graphically literate and did possess adequate levels of background knowledge, they may not have possessed these characteristics. While Tang argues that students must be explicitly instructed to exploit visual sources of information, the conditions she investigated involved written texts rather than audio texts. Because the texts in this case were audio texts, and the presentation of visual information was

contiguous, the study may have created the conditions that are most facilitative when visual sources of information are under consideration. By avoiding problems with potential conflicts associated with dual coding (written texts with visuals as opposed to audio texts with visuals, see earlier discussion of Mayer, 1997), the probability that subjects would attend to the visuals was increased. In fact, it would be difficult to imagine subjects not attending to the visually-presented information, unless they completed the experiment with their eyes closed. The presentation of the visuals and the text, in essence, forced subjects to attend to the visual information and may have ameliorated the effects of graphic literacy and prior knowledge.

The visuals accompanying dialogues and short conversations were provisionally classified as serving a decorative purpose, according to Levin's scheme. However, given that Levin's classification was designed to capture instructional purposes displayed in visuals, the categorization was recognized as inadequate and unable to account for the potentially important information that an understanding of the context of a conversation might provide. The function of the visuals accompanying dialogues and short conversations has been identified by TOEFL test developers as setting the scene. It may be that the presence of a visual bearing contextual information activates certain scripts. For example, knowing that the dialogue occurs in a restaurant might generate expectations regarding the content of the verbal exchange. However, TOEFL items are not presently designed to take advantage of such contextual information, and in most cases, the content of the dialogues is only marginally related to the dialogue's location. The dialogues usually occur in school-related settings -- like a dining hall, a corridor, or a classroom -- and typically involve school-related topics -- such as selecting classes or studying for exams. Given the contextual information presented in the visuals, it seems equally likely that the conversation could involve a visit to a doctor's office or a discussion of a work assignment. Any conversation related to any institutional setting would be equally appropriate.

Furthermore, there are no contextual restrictions that prohibit speakers from engaging in a discussion about swimming or mountain climbing in any of the scenes depicted. That is, it is not necessary to be in a pool to discuss swimming. While context may lead to the activation of particular scripts, the actual topic of conversation may supercede the context and probably, in normal conversation, usually does. It may be the case, with respect to dialogues and short conversations, that in order for the contextual information provided by the visual to actually facilitate comprehension, the information provided by both sources would have to be closely matched (e.g., perhaps talking about swimming in a pool). The complementary nature of visual and verbal information is argued by Salomon (1989) to be a necessary condition for visual supplantation to occur. Given the very generic nature of both the visual and the verbal information presented in TOEFL dialogues and short conversations, it is hardly surprising that the effect of visuals in this condition was virtually nonexistent.

Like the visuals associated with dialogues and short conversations, the visuals that accompanied academic discussions did not present contextual information that was directly related to the content of the verbal portion of the message; however, their presence did serve to facilitate comprehension. One reason that subjects might have an advantage when visuals are present in association with academic discussions is related to testing conditions. Because academic discussions involve a series of turns among several speakers of the same gender, and

because they are longer than the dialogues and short conversations, changing still photos in association with changes from speaker to speaker may have made the discussions easier to parse.

Given Levin's classification, the function of visuals in the case of academic discussions was provisionally identified as decoration or as interpretation. While the visuals that accompany academic discussions are similar to the visuals that accompany dialogues and short conversations, it is likely that the visuals provided with academic discussions are slightly facilitative because they mark turns in the conversations rather than because they "set the scene." Thus, in this case the more appropriate classification would be interpretation.

The effect associated with the presence or absence of visuals with mini-talks provides more evidence for the idea that audio portions of stimuli interact differently with different types of visual information. Mini-talks with content visuals appeared to have the greatest potential for complementary mapping of information between the verbal and visual portions of the stimuli. This mapping was expected to result in visual supplantation and stronger performance on these items, and it did. While the presence of visuals bearing content information might be thought to enhance the complexity of the stimulus, subjects apparently were able to use the information presented in the visual to help create a mental representation of the text.

The content visuals that accompanied mini-talks were provisionally classified, according to Levin's scheme, as serving the function of representation, because these visuals did overlap with the text and can be argued to make the text more concrete. These visuals may have also served the function of transformation by enhancing the memorability of the text. It is likely that, given a still photo of a mosquito in a piece of amber, it might be easier to answer a question concerning the types of animals typically found in amber inclusions than if the subject had only heard a verbal description.

When mini-talks were presented with visuals bearing only contextual information, the presence of the visuals had a slightly debilitating effect on performance. This is the only condition in which subjects actually performed better when viewing a blank screen. Once again, the information in the context visuals that accompanied mini-talks -- like that contained in the visuals accompanying dialogues and short conversations and academic discussions -- bore generic information about the context in which the communicative event occurred -- information that was, at best, only tangentially related to the content of the text. It may be that when subjects are presented with a short lecture that involves academic content, they are distracted by the presence of visuals that have little or no bearing on the content of the talk. For example, when the subject is presented with an audio text discussing Polynesian trade routes, a map depicting those routes would be expected to make the direction or location of the routes easier to remember. If presented with a visual depicting the lecturer, the relationship between the content information in both sources is certainly less direct.

It might be possible to devise visuals in which a gesture or an expression marks the relative importance of different pieces of information, but the visuals in this study were not designed to investigate that possibility. Given that the visual information provided by context visuals in association with mini-talks requires processing but provides no directly-related

complementary information, the processing load may be increased and the memorability of a particular piece of information may be decreased. Considering the interaction effect across each of the stimulus types, this scenario appears likely.

Stimulus Type by Time

Given that the experimental context was not a high-stakes testing situation, the drop in performance in association with the longer sets of stimuli may have occurred because of boredom or fatigue. Subjects may not have been sufficiently motivated to overcome these effects in the experimental context. If subjects had been told that they would receive only partial payment if they were unable to perform at a predetermined level of performance, these effects might have disappeared. An alternative explanation is that the effect is the product of memory limitations and attentional deficits that would occur in association with longer stimuli sets in any context.

Mayer (1997) argues that one variable that has been consistently overlooked in the literature concerning the effects of visuals in instructional contexts is memory load. As TOEFL listening-comprehension materials are consistently being lengthened in an attempt to more accurately reflect the kinds of materials to which students are exposed in academic settings, the significance of this interaction suggests that the effect of the length of listening-comprehension stimuli deserves special attention. A reasonable argument in support of the inclusion of visuals bearing content related information is their potential to ameliorate the effects of memory load on performance. While it may be argued that the TOEFL listening-comprehension section is not intended to test visual literacy, it is also not intended to test memory. The benefits of the inclusion of visuals bearing content-related information may help offset the cognitive cost of the greater memory load required by increasing the length of the verbal stimuli.

Proficiency and Visual Condition

Stimulus type and visual condition did not interact with proficiency. Furthermore the significance of the interaction between stimulus type and visual condition (which collapses proficiency levels) suggests that proficiency did not effect the potential of facilitation to occur. The performance of both groups was facilitated when the presence of visuals complemented the audio portions of the stimuli. As Mayer and Gallini (1990) suggest, the potential for facilitation to occur may be the function of prior knowledge rather than of language proficiency. It is important to note that, in spite of different levels of English proficiency, the majority of the subjects in this study were highly educated and may be assumed to have requisite levels of background knowledge with respect to the topics presented in TOEFL stimuli.

A Review of the Research Questions

Do subjects perform better on TOEFL test items when visuals accompany the audio text?

The main effect for the presence of visuals was not significant, but the presence of visuals did interact with stimulus type.

Does the presence of visuals interact with stimulus type?

Salomon (1989) argues that for visual supplantation to occur, the information in the visuals and in the text must have a complementary relationship. In the present study, when the visuals presented information that complemented the audio texts -- as was the case with academic discussions and mini-talks with content visuals -- their presence facilitated performance. In Levin's classification, the visuals that accompanied academic discussions can be argued to serve an interpretative function by marking turns in the conversation and making difficult-to-process text easier to comprehend. The visuals that accompanied mini-talks with content visuals can be argued to serve a representational function, because they overlapped with the text and made the text more concrete. These visuals may also have served a transformational function by making the text more memorable. Facilitation in these conditions is predicted by previous findings in the literature and was the expected result.

When the visuals presented information that could only be related to the context of the communicative event, either there was no effect (dialogues and short conversations) or there was a slightly debilitating effect (mini-talks with context visuals). While Levin's classification does not take into account the potentially facilitative effects of contextual information, for lack of a better system the visuals in association with these stimulus types were classified as serving a decorative function. Because of the generic nature of the contexts involved and their only tangential relationship to the content of the audio portion of the stimuli, the absence of facilitation in these conditions is not unexpected.

It might be possible to create context visuals that produce facilitation, but it appears that the context would have to bear a more direct relationship to the content of the audio portion of the stimuli than it does now. Test designers might provide visual information specifically designed to activate particular scripts (e.g., ordering from a menu when the setting is a restaurant). However, as the visuals become more specific in their representation of contextual information, they might also become more culturally specific. The advantage of providing generic contexts is lowering the possibility of the introduction of cultural bias.

In their current form, the presence of cultural bias cannot be discounted. Perhaps facilitation did not occur because the expectations of the subjects were violated by the contexts that were represented. This appears unlikely. While it is possible, it seems unlikely that TOEFL examinees are not familiar with the hallways and classrooms depicted in the context visuals. The contexts seem consistent with the representation of any hallway or classroom that exists in a building in any context.

Is an effect on performance related to subjects' English proficiency?

Stimulus type and visual condition did not interact with proficiency. Furthermore, the significance of the interaction between stimulus type and visual condition (which collapses proficiency levels) suggests that proficiency did not effect the potential of facilitation to occur. That is, the performance of both high- and low-proficiency groups was facilitated when the presence of visuals complemented the audio portions of the stimuli. As the research of both

Mayer and Gallini (1990) and Tang (1992, 1994) suggest, the potential for facilitation to occur may be the function of prior knowledge and visual literacy rather than of language proficiency. The facilitative effects of visuals bearing context information may also be the result of presentation. Because visual and text information were complementary, presentation of visual information was contiguous, and given that audio texts rather than written texts were used when content visuals were involved, the test designers may have created the most felicitous conditions with respect to the potential for facilitation.

It is interesting to note that Tang (1992, 1994) resists the trend toward an emphasis on cultural differences in the presentation of information in order to emphasize commonalties across cultures -- at least with respect to the frequency and function of graphics in middle-school social studies textbooks. If it is the case that representational structures (Mohan's knowledge structures) are similar at other instructional levels and across other instructional domains, their use in language testing raises some interesting questions.

If the presence of graphics activates prior knowledge structures that are common across cultures and languages, it may be the case that a particular examinee's effective use of a graphic would allow the examinee to perform at a higher level than if the verbally-presented information had appeared alone. This clearly raises questions about the nature of the construct being tested. To the extent that it is possible to test language without testing cognition -- that is, if test developers want to capture language proficiency without testing concomitant knowledge structures -- it may be advisable to exclude some types of graphics from test materials. On the other hand, it can be argued that the inclusion of graphics is important with respect to representational fidelity. In Tang (1994), all of the texts were highly illustrated (graphics were present on 71% to 94% of the pages); introductory college-level texts may also be. Furthermore, if we emphasize "bias for best" (Duran, 1989) -- that is, the creation of stimuli and items that allow for the strongest rather than the weakest performance -- it may also make sense to include potentially facilitative graphics whenever possible.

Do subjects report clear preferences for either the audio-only or audiovisual stimulus conditions?

The absence of a significant difference in response patterns in association with the presence or absence of visual accompaniments (Tables 6 and 7) suggests that subjects were not attending to visual conditions while they were in the process of completing the experiment. They may have been attending to different aspects of the presentation of the stimuli (e.g., the actual content of the verbal portion of the stimuli) when they considered how interesting and difficult the items were.

Subjects clearly preferred audiovisual stimuli (Table 8). Debilitating effects of the presence of visual accompaniments cannot be discounted for the small portion of the subjects who preferred the no visual condition, but these subjects were in the minority, and their performance was not examined separately. Given the preference for the audiovisual condition by the majority of the subjects, the use of visuals could be justified by subjects' preferences alone.

Directions for Future Study

It is important to bear in mind that even though significant interaction effects were found, the interactions account for a very small portion of the variance in subjects' performance. Given these exceedingly weak effects, research focusing on the examination of the unexplained portions of the variance is required.

If we ever hope to fully explicate the effects of different types of visuals on subjects' performance, it would be extremely useful to develop a model for the difficulty of TOEFL listening-comprehension items that is independent of the effects of visuals. If the features affecting difficulty were more clearly understood, the characteristics of experimental materials could be more carefully controlled. This would allow the interactions with visual sources of information to be examined in a more systematic fashion. It would be possible to devise experimental manipulations of visual information that could be used to unravel the effects of the presence of different types of context and content visuals.

It is difficult to imagine an accurate model of listening-comprehension item difficulty that did not take into account the memory requirements associated with different topics and lengths of audio texts. Given the different ranks of difficulty in association with the different general topics presented by the mini-talks in this study, it appears that both topic and item characteristics should be considered.

While the visuals in this study were provisionally classified according to Levin's scheme, there clearly were problems with his classification system. The first was that no appropriate classification exists for context visuals. It is possible to describe the function of certain types of context visuals as interpretative when they mark turns in a conversation; however, the designation of context visuals that set the scene as decorations seems inadequate. Identifying the provision of contextual information as decoration underplays its potential to provide relevant, perhaps even important, construct-related information -- in spite of the absence of facilitation in association with these visuals found here. If the presence of context visuals improves the face validity of the test for both examinees and language specialists, it seems reasonable to include them if their presence does no harm. The slight debilitation of performance when context visuals were presented in association with mini-talks is problematic, but remains a weak effect. In any case, this type of visual needs a better descriptor -- one that captures its intended purpose and theoretical value even though it may not reflect performance effects.

If the presence of visuals is intended to "bias for best," the characteristics of content visuals require continued investigation. The potential of content visuals to ameliorate the effects of increased difficulty that may occur with the use of longer audio stimuli is an interesting possibility. While content-related visuals appeared only with mini-talks, they could easily be provided with academic discussions as well. An examination of the effects of context-related visuals marking turns in the conversation, as opposed to content-related visuals presenting complementary visual information, might lead to better understandings of not only how content visuals function, but also when and why contextual information can be used to facilitate comprehension. Academic discussions lend themselves well to this type of investigation.

Finally, it should be noted that the provisional classifications of the visuals used in this study were intended to categorize the sets of visuals that accompanied the stimulus types in the visual condition. One of the problems with classifying a particular set of visuals as representational is that other classifications, particularly interpretation and transformation, appear to apply as well. A second difficulty is that each of the individual visuals may serve different functions within each set. Clearly, test developers need a classification system that allows them to accurately categorize the visuals that are currently in use.

Given the weak effect of the stimulus-type-by-visual interaction, the most accurate summary is that the presence of visuals appeared, generally, to do no harm. Indeed, visuals had a slightly beneficial effect on performance when the visuals bore information related to the content of the audio portions of the stimuli. It may be disappointing to some that the presence of context visuals as accompaniments to dialogues and short conversations, or as accompaniments to mini-talks, produced either no effect or a slightly debilitating effect, given the value placed on context in discussions of communicative competence. However, if the presence of context visuals improves the face validity of the test for both examinees and language specialists, it does not appear unreasonable to include them. The slight debilitation of performance when context visuals were presented in association with mini-talks is problematic, but remains a negligible effect. It may, however, not be reasonable to emphasize the importance of representations of context if those representations cannot be associated with an improvement of the measurement of the underlying construct that goes beyond increased face validity.

Tables

Table 1. Proficiency by Status by Gender

Proficiency							
Low = 80				High = 80			
Status							
Married Student ESL Program		Visiting scholar		Undergraduate		Graduate	
66		14		40		40	
Gender							
M	F	M	F	M	F	M	F
17	49	8	6	25	15	23	17

Table 2. Highest Degree Earned by Proficiency by Status

Highest degree earned in native country	Proficiency			
	Low		High	
	Status			
	Married Student ESL Program	Visiting scholar	Undergraduate	Graduate
Elementary	9			
High school	11	1	36	
Associate's	1			
B.A./B.S.	26	6	3	26
M.A./M.S.	14	5	1	12
Ph.D.	3	2		2
M.D.	1			
Law	1			
Total	66	14	40	40

Table 3. Native Language by Proficiency

Native language	Proficiency	
	Low	**High**
Arabic	5	1
Bambara	1	
Bengali		1
Bulgarian		1
Chinese	19	21
Dutch		1
Farsi	1	
German		4
Greek		1
Hindi	2	9
Indonesian	2	9
Italian	1	1
Japanese	4	4
Korean	22	6
Luo		1
Malay	1	4
Memde		1
Ndebele		1
Persian	2	
Polish	1	
Portuguese	2	2
Romanian		1
Russian	2	
Serbian		1
Setswana		2
Spanish	13	4
Telugu	1	
Turkish		1
Ukranian	1	
Urdu		1
Total	80	80

Table 4. Stimulus Type by Visual Condition by Proficiency

		Visual Condition			
Stimulus type Subset *Type of visual*	Item/topic	Forms: 1, 2, 5, 6, 9, 10, 13, 14 (n = 80)		Forms: 3, 4, 7, 8, 11, 12, 15, 16 (n = 80)	
		Proficiency			
		LP (n = 40)	HP (n = 40)	LP (n = 40)	HP (n = 40)
Dialogue/Short Conversations Subset 1 *Context*	D/SC 1.1: The wind D/SC 1.2: Biology class D/SC 1.3: Library card D/SC 1.4: Library card D/SC 1.5: Library card	**V**	**V**	NV	NV
Dialogue/Short Conversations Subset 2 *Context*	D/SC 2.1: Stat Course D/SC 2.2: Borrow book D/SC 2.3: Dinner party D/SC 2.4: Dinner party D/SC 2.5: Dinner party	**NV**	**NV**	**V**	**V**
Academic Discussion Subset 1 *Context*	AD 1.1: Art History AD 1.2: Art History AD 1.3: Art History AD 1.4: Art History AD 1.5: Art History	**V**	**V**	NV	NV
Academic Discussion Subset 2 *Context*	AD 2.1: Engineering AD 2.2: Engineering AD 2.3: Engineering AD 2.4: Engineering AD 2.5: Engineering	NV	NV	**V**	**V**
Mini-talk Subset 1 *Context*	MTX 1.1: Prairie Dogs MTX 1.2: Prairie Dogs MTX 1.3: Prairie Dogs MTX 1.4: Prairie Dogs MTX 1.5: Prairie Dogs	**V**	**V**	NV	NV
Mini-talk Subset 2 *Context*	MTX 2.1: Anthropology MTX 2.2: Anthropology MTX 2.3: Anthropology MTX 2.4: Anthropology MTX 2.5: Anthropology	NV	NV	**V**	**V**
Mini-talk Subset 1 *Content*	MTN 1.1: Aquifers MTN 1.2: Aquifers MTN 1.3: Aquifers MTN 1.4: Aquifers MTN 1.5: Aquifers	**V**	**V**	NV	NV
Mini-talk Subset 2 *Content*	MTN 2.1: Amber MTN 2.2: Amber MTN 2.3: Amber MTN 2.4: Amber MTN 2.5: Amber	NV	NV	**V**	**V**

Note: LP = low proficiency; HP = high proficiency; D/SC = dialogues/short conversations; AD = academic discussions; MTX = mini-talks with context visuals; MTN = mini-talks with content visuals; V = visual provided; NV = no visual provided.

Table 5. Results of the Study

Source	Num. SS	Num. Df	Num. MS	Den. MS	F	*p*	η^2	partial η^2
Proficiency	325.50	1	325.50	4.92	66.15	.00		
Status (proficiency)	11.53	2	5.77	4.92	1.71	.31		
Form	49.11	15	3.27	4.92	.67	.82		
Subjects (proficiency X status X form)	693.80	141	4.92	.76	6.51	.00		
Stimulus type	258.42	3	86.14	.76	113.92	.00		
Time	39.55	1	39.55	.76	52.31	.00		
Visual	.60	1	.57	.76	.75	.39		
Proficiency X stimulus type	8.12	3	2.71	.76	3.58	.01	.003	.009
Proficiency X visual	.09	1	.09	.76	.12	.72		
Stimulus type X visual	7.09	3	2.36	.76	3.12	.03	.003	.008
Stimulus type X time	62.11	3	20.70	.76	27.40	.00	.025	.06
Proficiency X stimulus type X visual	1.15	3	.38	.76	.51	.68		
Error	833.27	1102	.76					
Corrected Total	2487.96	1279						

Table 6. The Questions Were Interesting

The questions were interesting.		V (n = 80)	NV (n = 80)	⊠^{2}	df	sig	w
Dialogue/short conversations Subset 1	1. Strongly disagree	1%	5%	5.18	4	.21	--
	2.	9%	9%				
	3.	24%	28%				
	4.	43%	48%				
	5. Strongly agree	24%	11%				
Academic discussion Subset 1 ***Art History***	1. Strongly disagree	3%	15%	8.63	4	.07	--
	2.	16%	15%				
	3.	33%	28%				
	4.	39%	30%				
	5. Strongly agree	10%	13%				
Mini-talk w/ content Subset 1 ***Prairie Dogs***	1. Strongly disagree	1%	4%	2.56	4	.64	--
	2.	4%	6%				
	3.	25%	24%				
	4.	41%	45%				
	5. Strongly agree	29%	21%				
Mini-talk w/ context Subset 1 ***Aquifers***	1. Strongly disagree	1%	6%	5.16	4	.27	--
	2.	4%	8%				
	3.	19%	25%				
	4.	44%	39%				
	5. Strongly agree	31%	23%				
		NV	**V**				
Dialogue/short Conversations Subset 2	1. Strongly disagree	3%	4%	1.96	4	.74	--
	2.	6%	4%				
	3.	31%	26%				
	4.	39%	38%				
	5. Strongly agree	21%	29%				
Academic discussion Subset 2 ***Engineering***	1. Strongly disagree	10%	11%	10.35	4	.04	.53
	2.	19%	13%				
	3.	21%	43%				
	4.	29%	24%				
	5. Strongly agree	21%	10%				
Mini-talk w/ content Subset 2 ***Anthropology***	1. Strongly disagree	5%	5%	5.25	4	.26	--
	2.	26%	16%				
	3.	18%	30%				
	4.	34%	36%				
	5. Strongly agree	18%	13%				
Mini-talk w/ context Subset 2 ***Amber***	1. Strongly disagree	3%	8%	4.90	4	.30	--
	2.	9%	8%				
	3.	28%	16%				
	4.	33%	34%				
	5. Strongly agree	29%	35%				

Table 7. The Questions Were Difficult

The questions were difficult.		V (n = 80)	NV (n = 80)	$\boxtimes^2$	df	sig	w
Dialogue/short conversations Subset 1	1. Strongly disagree	30%	35%	2.49	4	.65	--
	2.	39%	28%				
	3.	20%	24%				
	4.	10%	11%				
	5. Strongly agree	1%	3%				
Academic discussion Subset 1 ***Art History***	1. Strongly disagree	4%	13%	5.39	4	.25	--
	2.	20%	21%				
	3.	33%	29%				
	4.	29%	20%				
	5. Strongly agree	15%	18%				
Mini-talk w/ content Subset 1 ***Prairie Dogs***	1. Strongly disagree	14%	13%	.23	4	.99	--
	2.	33%	35%				
	3.	33%	30%				
	4.	19%	20%				
	5. Strongly agree	3%	3%				
Mini-talk w/ context Subset 1 ***Aquifers***	1. Strongly disagree	19%	19%	1.83	4	.77	--
	2.	29%	34%				
	3.	26%	28%				
	4.	20%	13%				
	5. Strongly agree	6%	8%				
		NV	**V**				
Dialogue/short conversations Subset 2	1. Strongly disagree	39%	41%	6.38	4	.17	--
	2.	31%	35%				
	3.	19%	16%				
	4.	11%	4%				
	5. Strongly agree	0%	4%				
Academic discussion Subset 2 ***Engineering***	1. Strongly disagree	9%	9%	2.43	4	.66	--
	2.	29%	21%				
	3.	28%	26%				
	4.	23%	33%				
	5. Strongly agree	13%	11%				
Mini-talk w/ content Subset 2 ***Anthropology***	1. Strongly disagree	8%	11%	1.61	4	.80	--
	2.	21%	23%				
	3.	31%	34%				
	4.	29%	21%				
	5. Strongly agree	11%	11%				
Mini-talk w/ context Subset 2 ***Amber***	1. Strongly disagree	10%	16%	7.08	4	.13	--
	2.	18%	30%				
	3.	35%	26%				
	4.	29%	18%				
	5. Strongly agree	9%	10%				

Table 8. Preference for Visuals

Questions:		(n=160)	☒^{2}	df	Sig	w
1. I preferred the exercises with visuals.	1. Strongly disagree 2. 3. 4. 5. Strongly agree	3% 6% 24% 31% 36%	68.00	4	.00	.63
2. It was easier for me to understand the speakers when they were accompanied by visuals.	1. Strongly disagree 2. 3. 4. 5. Strongly agree	4% 15% 16% 29% 36%	50.81	4	.00	.56
3. I found the visuals distracting.	1. Strongly disagree 2. 3. 4. 5. Strongly agree	29% 24% 29% 11% 8%	32.00	4	.00	.45
4. The visuals made it easier for me to understand the speaker(s) in an academic lecture.	1. Strongly disagree 2. 3. 4. 5. Strongly agree	9% 19% 23% 27% 22%	14.94	4	.00	.31
5. The visuals made it easier for me to understand the speaker(s) in an academic discussion.	1. Strongly disagree 2. 3. 4. 5. Strongly agree	9% 19% 25% 30% 17%	18.88	4	.00	.34
6. The visuals made it easier for me to understand the speaker(s) in a short conversation.	1. Strongly disagree 2. 3. 4. 5. Strongly agree	7% 11% 18% 33% 31%	45.00	4	.00	.53

Figures

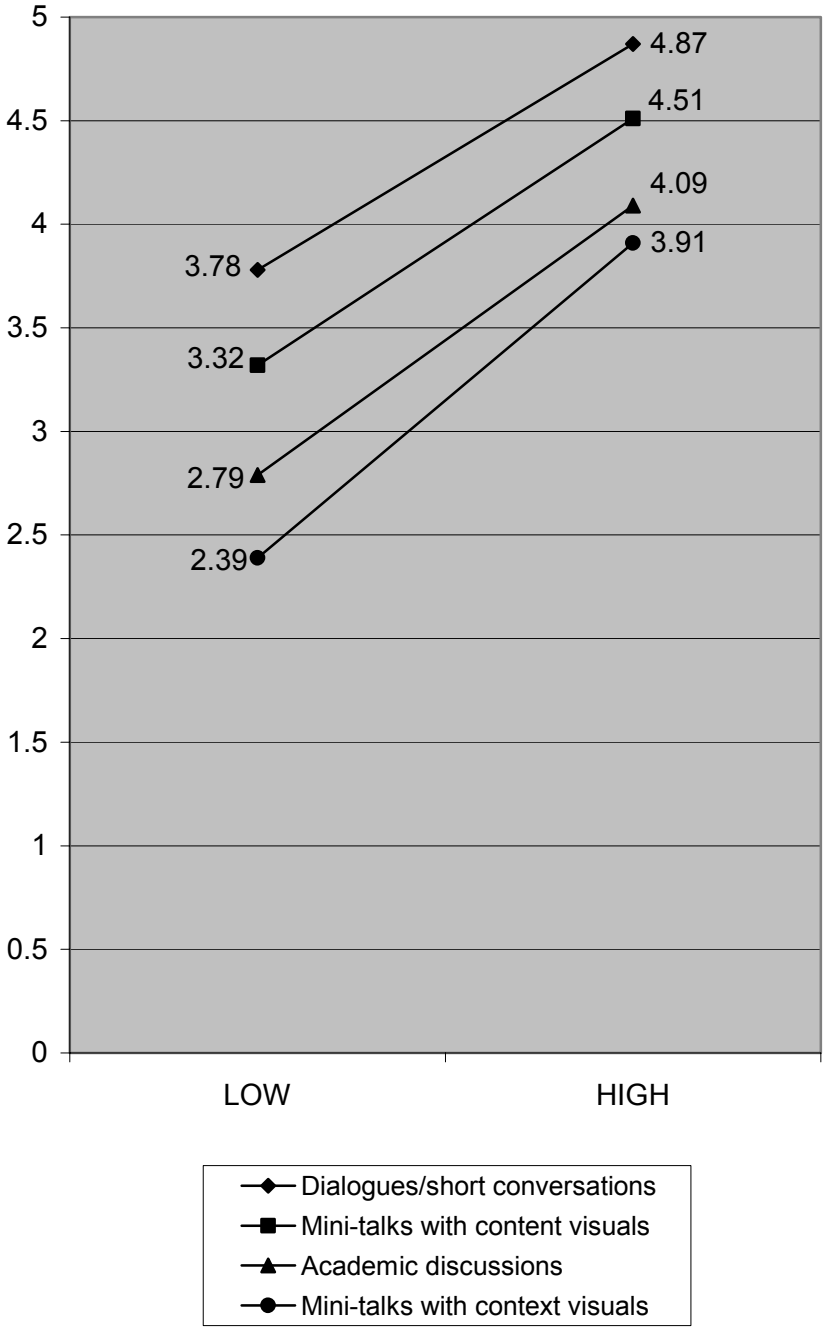

Figure 1. Proficiency by stimulus type.

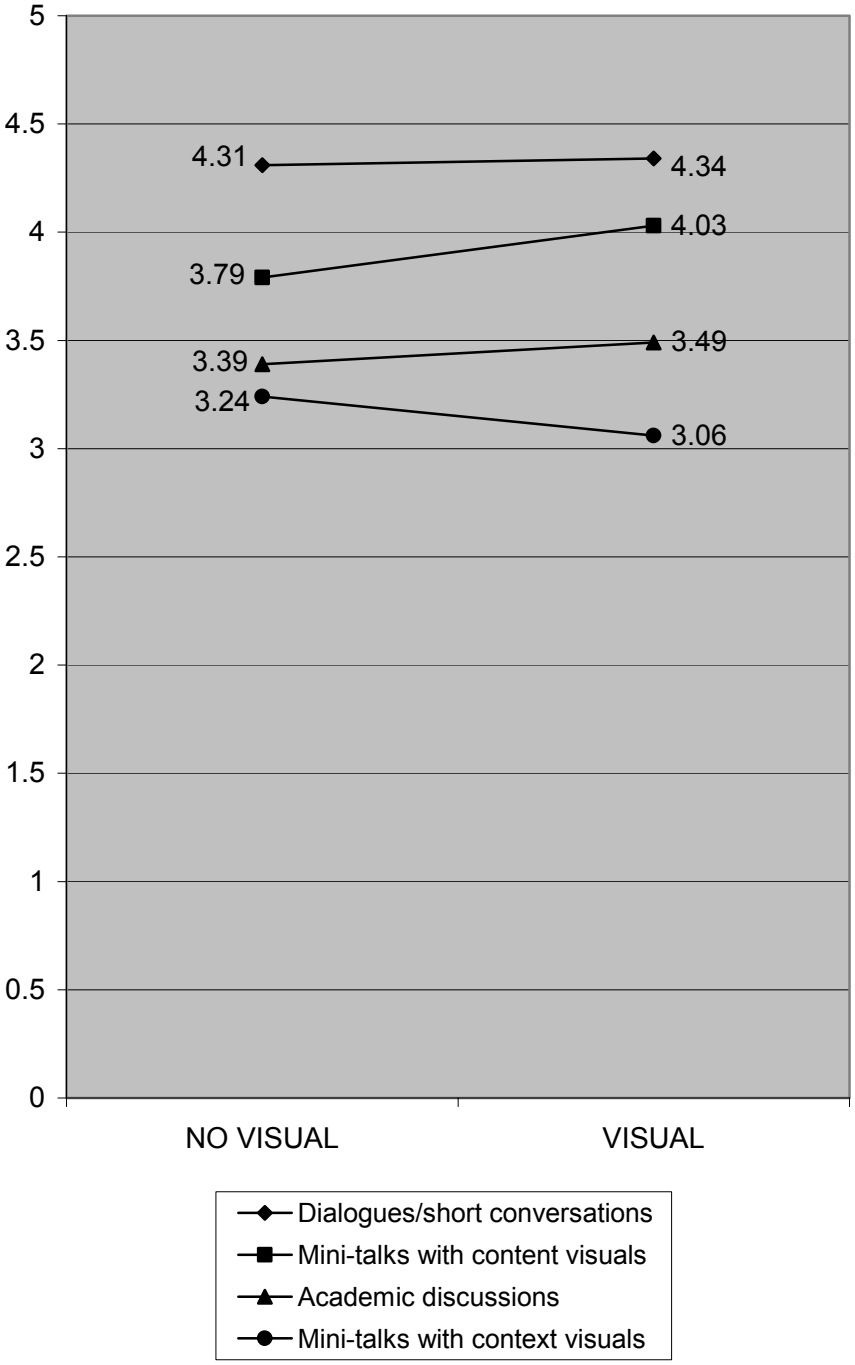

Figure 2. Stimulus type by visual condition.

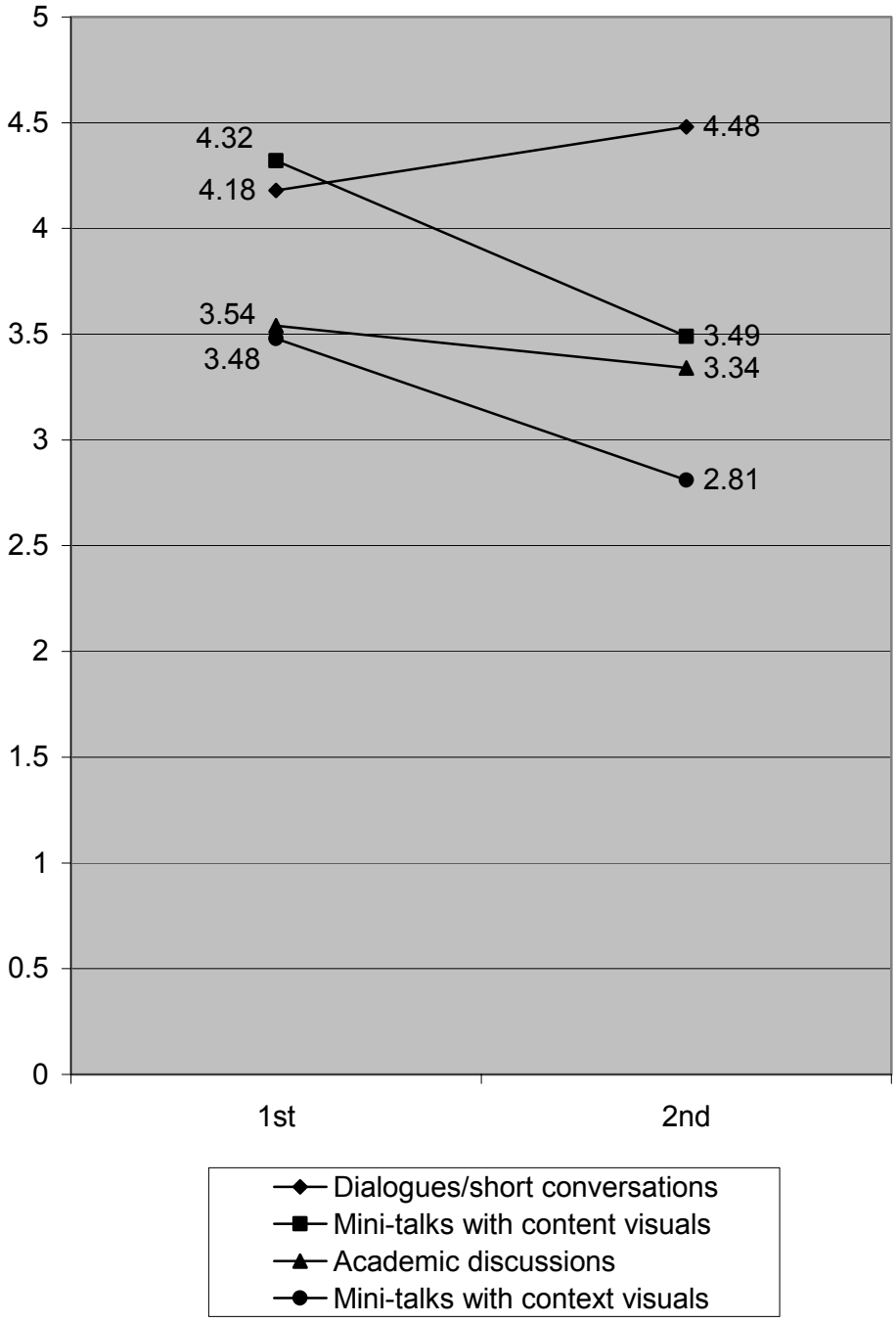

Figure 3. Stimulus type by time.

References

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly 25* (4), 671-704.

Baddeley, A. (1992). Working memory. *Science, 255*, 556-559.

Bauer, M. I., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science, 4,* 372-378.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

Cochran, W. G., & Cox, G. M. (1957). *Experimental designs*. Toronto, CA: John Wiley & Sons.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*, 116-126.

Douglas, D. (1997). *Testing speaking ability in academic contexts: Theoretical considerations* (TOEFL Monograph Series No. 8). Princeton, NJ: Educational Testing Service.

Duran, R. P. (1989). Testing of linguistic minorities. In R. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 573-587). New York: American Council on Education: Macmillan.

Dwyer, F. M., & de Melo, H. (1984). Effects of mode of instruction, testing, order of testing, and cued recall on student achievement. *Journal of Experimental Education, 52*, 86-94.

Hegarty, M., &. Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*, 717-742.

Hymes, D. (1974) *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania.

Kaplan, R. (1987). Cultural thought patterns revisited. In U. Connor & R. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 9-21). Reading, MA: Addison-Wesley.

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

Levin, J. R. (1989). A transfer-appropriate-processing perspective of pictures in prose. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 83-100). New York: Elsevier Science.

Mandl, H., & Levin, J. R. (Eds.). (1989). *Knowledge acquisition from text and pictures*. New York: Elsevier Science.

Mastropieri, M. A., Scruggs, T. E., & Levin, J. R. (1987). Learning disables students' memory for expository prose: Mnemonic versus nonmnemonic pictures. *American Journal of Educational Research, 24*, 505-519.

Mayer, R. E. (1984). Aids to prose comprehension. *Educational Psychologist, 19*, 30-42.

Mayer, R. E. (1989). Models for understanding. *Review of Educational Research, 59*, 43-64.

Mayer, R. E. (1993a). Comprehension of graphics in text: An overview. *Learning and Instruction, 3*, 239-246.

Mayer, R. E. (1993b). Illustrations that instruct. In R. Glaser (Ed.), *Advances in instructional psychology*: *Vol. 5* (pp. 253-284). Hillsdale, NJ: Lawrence Erlbaum.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32 (1)*, 1-19.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 84*, 444-452.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*, 444-452.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*, 64-73.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*, 715-726.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389-401.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development, 43 (1)*, 31-44.

Mohan, B. A. (1986). Language and content. Reading, MA: Addison-Wesley.

Moore, D. M., & Dwyer, F. M. (1994). Effect of cognitive style on test type (visual or verbal) and color coding. *Perceptual and Motor Skills, 79*, 1532-1534.

Parkhurst, P. E., & Dwyer, F. M. (1983). An experimental assessment of students' IQ level and their ability to profit from visualized instruction. *Journal of Instructional Psychology, 10*, 9-20.

Salomon, G. (1979). *Interaction of media cognition and learning*. San Francisco: Jossey-Bass.

Salomon, G. (1989). Learning from texts and pictures: Reflections on a meta-level. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 73-82). New York: Elsevier Science.

SAS Institute. (1998). Statistical Analysis Software, Version 8. Cary, NC: Author.

Schegloff, E. A. (1997). Whose text? Whose context? *Discourse and Society, 8*, 165-187.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structure of technical material. *Journal of Experimental Psychology: General, 119*, 176-192.

Tang, G. (Winter, 1991). The role and value of graphic representation of knowledge structures in ESL student learning: An ethnographic study. *TESL Canada Journal/Revue TESL Du Canada*, *9* (1), 29-41.

Tang, G. (1992). The effect of graphic representation of knowledge structures on ESL reading comprehension. *SSLA, 14*, 177-195.

Tang, G. (1994). Textbook illustrations: A cross-cultural study and its implications for teachers of language minority students. *The Journal of Educational Issues of Language Minority Students, 13*, 175-194.

Vacca, R. T. (1981). Content area reading. Boston: Little Brown & Company.

Weidenmann, B. (1989). When good pictures fail: An information processing approach to the effect of illustrations. In H. Mandl &. J. R. Levin (Ed.), *Knowledge acquisition from text and pictures*. New York: Elsevier Science.

Willows, D. M., & Houghton, H. A. (Eds.). (1987). *The psychology of illustration: Volume 2. Instructional Issues*. New York: Springer-Verlag.

Winn, W. (1991). Learning from maps and diagrams. *Educational Psychology Review, 3*, 211-247.

Appendix A. List of Test Forms

Section	Form 1	Form 2	Form 3	Form 4
1	D/SC1V D/SC2NV	D/SC2NV D/SC1V	D/SC1NV D/SC2V	D/SC2V D/SC1NV
2	AD1V AD2NV	AD2NV AD1V	AD1NV AD2V	AD2V AD1NV
3	MTX1V MTX2NV	MTX2NV MTX1V	MTX1NV MTX2V	MTX2V MTX1NV
4	MTN1V MTN2NV	MTN2NV MTN1V	MTN1NV MTN2V	MTN2V MTN1NV

Section	Form 5	Form 6	Form 7	Form 8
2	AD1V AD2NV	AD2NV AD1V	AD1NV AD2V	AD2V AD1NV
3	MTX1V MTX2NV	MTX2NV MTX1V	MTX1NV MTX2V	MTX2V MTX1NV
4	MTN1V MTN2NV	MTN2NV MTN1V	MTN1NV MTN2V	MTN2V MTN1NV
1	D/SC1V D/SC2NV	D/SC2NV D/SC1V	D/SC1NV D/SC2V	D/SC2V D/SC1NV

Section	Form 9	Form 10	Form 11	Form 12
3	MTX1V MTX2NV	MTX2NV MTX1V	MTX1NV MTX2V	MTX2V MTX1NV
4	MTN1V MTN2NV	MTN2NV MTN1V	MTN1NV MTN2V	MTN2V MTN1NV
1	D/SC1V D/SC2NV	D/SC2NV D/SC1V	D/SC1NV D/SC2V	D/SC2V D/SC1NV
2	AD1V AD2NV	AD2NV AD1V	AD1NV AD2V	AD2V AD1NV

Section	Form 13	Form 14	Form 15	Form 16
4	MTN1V MTN2NV	MTN2NV MTN1V	MTN1NV MTN2V	MTN2V MTN1NV
1	D/SC1V D/SC2NV	D/SC2NV D/SC1V	D/SC1NV D/SC2V	D/SC2V D/SC1NV
2	AD1V AD2NV	AD2NV AD1V	AD1NV AD2V	AD2V AD1NV
3	MTX1V MTX2NV	MTX2NV MTX1V	MTX1NV MTX2V	MTX2V MTX1NV

Note: V = visual present; NV = no visual present

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

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

Phone: 609-771-7100
E-mail: toefl@ets.org
Web site: http://www.toefl.org

57906-005535 • Y81M.7 • Printed in U.S.A.
I.N.275623