

# Design and Development of a Video-Mediated Test of Communicative Proficiency

**Paul Gruba**

*University of Melbourne*

This paper discussed the development of the Kanda English Proficiency Test (KEPT) as a communicative test of English as Foreign Language for Japanese university students. The paper first considers the framework of communicative testing, examines each section of the test, and discusses issues related to the role of video in testing.

## コミュニケーション能力測定のための ビデオを使用したテストのデザインと開発

本稿は日本人大学生の為のコミュニケーション重視試験として開発された KEPT (神田英語学習熟度テスト) について論じる。先ず、コミュニケーション重視テストの枠組みを述べた後、KEPT の各項を検討し、試験におけるビデオ教材の役割に関する焦点を論ずる。

In McLuhan's (1964) phrase "The medium is the message" lies a challenge for language teaching and testing personnel: does the manner in which information is presented truly affect the way it is understood? Benson (1993) seems to think so, and calls on those using video for teaching to use the same medium to test student progress. For Benson, alternative learning styles are activated by visual media and video presents a range of information (such as sociolinguistic markers and views of the target culture) not available in any other form. Despite a range of materials on *how* to use video in the classroom (e.g., Cooper, Lavery & Rinvolutri, 1991; Stempleski & Tomalin, 1990), Tatsuki (1993) points out, "basic research into how learners learn language and how specific media affect language learning is needed" (p. 24). Gruba (1993) notes that, particularly in language testing, there is a need to explore the applications of video for evaluation purposes.

Established in 1989, the Kanda English Proficiency Test (KEPT) project aims to create a video-mediated, Japan-based communicative test of English proficiency. Though still in a developmental phase, the KEPT has been administered ten times. Following each administration, it has been revised. In order to look at the use of video in testing, this paper provides an overview of communicative language testing, explains key design concepts of the KEPT, and discusses issues of the use of video in language testing.

### Toward a Definition of Communicative Proficiency Testing

*Overview.* Regarding communicative tests, Heaton (1988) writes that they should result in a profile of the learner's abilities and be adapted for specific purposes, while not losing sight of purposeful applications of language use. Further, Heaton argues that communicative test content must be "totally relevant for a particular group of testees" and that test tasks "should relate to real-life situations, usually specific to a particular group or culture" (p. 20). Test developers ignorant of the examinees' culture may unintentionally introduce cultural bias and reduce test reliability, Heaton points out. A recognition of student needs, important in curriculum development (Nunan, 1988), should also figure prominently in communicative test development, according to Heaton.

To better distinguish communicative tests from exams that use language in a decontextualized framework, Bachman (1991, p. 678) noted four distinct points that set them apart:

1. An "information gap" is created that requires test takers to process information from several sources.
2. The successful completion of tasks in one section builds upon the content and completion of tasks in earlier sections.
3. Test tasks and content are integrated within a given domain of discourse.
4. A broad range of language abilities are measured, including knowledge of cohesion, functions, and sociolinguistic appropriateness (as opposed to tests which focus on grammar, vocabulary, and pronunciation).

### Definition of KEPT

Definitions of tests are difficult to frame and may not capture particular attributes of a given instrument (Henning, 1987). In order to set the

boundaries of the KEPT, however, key design concepts of the test are:

1. It is based on communicative principles of language learning, in that responses are meant to show a purposeful *exchange of information* in a specific domain of language.
2. Awareness of student needs is recognized in its design and development within the context of Japan.
3. Test content is integrated throughout, and only seen as a whole can it be fully understood.
4. Information is video-mediated.

To sum up, the KEPT is intended to be a video-mediated, integrated proficiency test of communication in English as a foreign language, based on preferences and needs of Japanese students, and is intended for use by Japanese students at universities and comparable foreign language institutions.

### Structure of KEPT

*Overview:* Following a number of modifications to the original series of developmental prototypes, the KEPT now has six sections: (a) an introductory video; (b) a reading section that has both a factual memo and an extended fictional narrative; (c) a cloze section to test grammar; (d) a video-mediated listening section that consists of two monologues and a dialogue; (e) a writing section that offers a choice of two prompts; and (f) an oral interview section (see Appendix A). The entire test is based on one integrated storyline and takes a total of one hour and 45 minutes to complete: the 90-item multiple choice section takes one hour, the composition a half hour, and the oral examination 15 minutes. Test instructions are printed both in English and Japanese on the test booklets, and are presented on the video tape itself.

*The Integrated Storyline:* Though the use of an integrated storyline appears to match communicative testing parameters, employing such a design is not without criticism. Low (1986) reported that the use of integrated storylines in language test development is faced with several key problems, chiefly: (a) the possibility of lowered reliability due to a violation of assumptions of item independence; (b) the possible introduction of a content bias founded upon a single theme, which cannot be moderated through the use of a variety of topics within one test; (c) possible initial student "shock" in adapting to an integrated test format; (d) the increased difficulty in specifying design and test specifications; and (e)

the lack of a clear definition of what is meant by "line of development" and thus the complications arising out of this for validation purposes.

Low (1986) concluded that investigation into the use of an integrated storyline in full scale examinations is "largely an unknown field" (p. 34), and one that requires rigorous evaluation. KEPT developers, aware of these problems, decided to adopt an integrated design while attending to student preferences.

According to a university student needs analysis conducted by Widdows and Voller (1991), the most common reason Japanese students would like to acquire English proficiency is to be able to cope with everyday situations in English abroad. A similar study (Busch et al., 1992) reconfirmed the findings that the KEPT theme of international travel, originally established in 1989, was appropriate.

The KEPT is based on the actions of five characters, employed at an international trading company in Tokyo, who must travel abroad as a team. Each character is of a different nationality: American, Australian, British, Canadian, and Japanese. There are two women and three men. The same five characters are seen in each of the four versions of the test.

In each version, the basic premise is essentially the same. The team has a meeting before leaving Tokyo, arrives in the new country and begins to explore. They are soon faced with a problem (e.g., someone becomes lost before a crucial meeting, or a corporate business decision is questioned) and two conflicting sides of the issue emerge. Both sides are presented and the team is faced with making a critical decision under deadline pressure. As the multiple choice section of the test ends, test candidates in the productive phase (writing and speaking) present arguments and suggest solutions to the dilemma.

*Introductory Sequence.* As the KEPT begins, an introductory sequence plays for five minutes to build context, introduce characters, and allow examinees time to adjust to English as spoken by native speakers. Candidates are told that no information given in the introductory video will be directly tested. The set for the video is a meeting room; the twelve characters reacquaint themselves with each other and discuss the upcoming project. The meeting is the only one held in Tokyo before the characters travel together. Technically, the video is simple: it is taped in real time using one camera from the front and contains no complex graphics, special lighting, or sophisticated sound track.

*Reading Proficiency:* Immediately following the introductory sequence, examinees are instructed to open their test booklets and start the read-

ing section. They first encounter a factual memo of 250 words. The memo is linked to the introductory sequence (one team member has promised to provide background information about the country). The format of the memo is standard American business style, containing a brief header of recipients, the sender, a topic, and date. Ten multiple-choice items, written in “wh-” format as recommended by Popham (1990), follow the memo to test skimming and scanning abilities in addition to the comprehension of factual information.

Because reading is an interactive process between the reader and the page (Grabe, 1991), the reading section of the KEPT aims to encourage students to bring much of their learning outside the classroom to the task of deciphering text within the contextual framework. Test items 11-30 of this section focus on inferential reading skills. Questions here, for example, may ask candidates to infer a character's emotional state from a written description of a gesture. As such, the second half of the reading section is fictional, written from an omniscient point of view, and is approximately 750 words long. In this section, the characters have recently arrived in the country and are beginning to explore. The total time allowed for examinees to complete the entire reading section of 30 multiple-choice items is 30 minutes, an average pace for reading tests.

*Grammatical Proficiency:* This section recognizes the role of grammar in a communicative framework; that is, to stress discourse features of the text and “where a number of possible choices will be correct or possible in terms of grammatical form alone, but only one will be correct when the meaning derived from context is applied” (Johnson et al., 1991, p. 22). Contextual comprehension is seen as an interdependent link to grammatical proficiency. Both Heaton (1988) and Savignon (1991) note the importance of grammar in communicative materials.

As the fictional story continues, the text is presented to examinees as a cloze test that employs rational (as opposed to fixed interval) deletions for a total of 30 items. Instructions advise the students to read through the text once for understanding before attempting to choose the correct word or phrase out of four possible choices.

*Listening Proficiency:* Prior to the introduction of communicative methodologies, listeners were viewed as passive receptors of messages; current models of communication stress that listeners are part of an interactive and interpretive process in which they “engage in a dynamic construction of meaning” (Murphy, 1991, p. 56).

There are three separate video-mediated listening passages in the listening section: two monologues and one dialogue. The problem, developed in the reading and grammar sections, is such that there are two possible solutions. The first monologue presents one side of the argument; the second argues the opposing point of view. The dialogue has two actors attempting to resolve the dilemma. All video passages have the actors talking at natural speed to the group; no special lighting, graphics or sound effects are used, and each video segment lasts to three minutes. After a one-minute preview of the items and note taking while viewing/listening, examinees are allowed six minutes to answer 10 four-distractor multiple-choice questions in each of the three sections.

To attempt to draw on some of the benefits of using video, questions in the first listening section employ graphics of gestures as part of the information to be tested. Efforts to further maximize the advantages of the video (e.g., facial expressions, movement, show of emotions) have proved difficult to incorporate because of the complexity of describing such actions in a multiple choice format.

*Writing:* To increase interactional authenticity in a communicative testing framework, particularly in the area of goal setting, Bachman (1991) suggests that examinees be given choices when faced with writing tasks. The writing section of the KEPT requires students to comment on what the team should do with their problem (analytical skills), or to continue writing the story from the point of view of one of the characters (creative skills). Compositions are later marked by trained raters using a standard scale (see Appendix B).

*Speaking Proficiency:* Oral examinations are held separately from the multiple-choice and composition sections and are held on an appointment basis. In the speaking section, examinees are given a choice of tasks to complete, similar to the ones in the composition section.

Upon entering the examination room, four examinees are seated in chairs facing each other. Two trained oral evaluators are present, one of whom acts as a guide in the exam while the other sits quietly. The four examinees sit in pairs directly facing each other while the evaluators sit parallel at either end. After introductions are made, the lead evaluator hands a mounted photograph (a scene from the video) to each of the candidates. Each photo is accompanied by a 30- to 35-word description of the scene, such as "Karen listened carefully to the others and took notes. She wondered what they would decide as a group. Tom looked out the window, hoping that he would eventually be able to see some

of the sights of Mexico City." The evaluator then instructs the examinees in English that they will be required to read the text aloud and describe for the group the actions in the photograph. The examinees are allowed one minute to read the text and examine the photograph.

After one minute of silent reading, each candidate reads the caption aloud. Following the task, the group is told to discuss the story and to solve the problem collectively. Both evaluators then sit back and take notes while the candidates have a conversation. Each of the group examinations lasts fifteen minutes.

Categories for oral evaluation include pronunciation, speech flow, expressiveness, and grammar (see Appendix C). Each of the four categories ranges from a top of five to a low of one; a total of 20 points can be awarded. Four points is the minimum score. The two evaluators may compare notes and discuss student performances immediately following the exam.

### KEPT Administration-Spring 1993

*Administration:* In early April 1993, the KEPT was administered. Out of the 1,545 English majors at Kanda University, 1,128 students took the examination, for a total of 73% of the available population. As the test is still in its developmental stages, participation in the examination was voluntary and students were informed that scores would not be added to permanent students records. After data processing, however, student scores were made available to students and faculty campus wide. A full report of the administration, including detailed statistical and item analysis of each section, was also written (Gruba, 1993b).

### Reliability

According to Bachman (1990), internal consistency measures to estimate test reliability (e.g., KR20, KR21, split half, or Cronbach's alpha) are thought to overestimate the reliability coefficient in integrated proficiency examinations. Accordingly, KEPT developers used "test-retest" procedures for reliability estimates.

Following the advice of Henning (1987), the test-retest procedure was conducted within two weeks of the first KEPT administration to minimize the effects of both test familiarity and further instruction. By the second week of classes, two random groups of second- and third-year students ( $n = 42$ ) were selected to take the multiple-choice section of the KEPT a second time. Using Pearson product moment correlation

procedures, the KEPT test reliability estimate was found to be .85. Ideally, reliability indices should range from the high .80s to the .90s (Hatch & Lazaraton, 1991).

Concerning the oral section, Gruba (1992) found KEPT inter-rater reliability estimates to be .92 in one test administration. Modification of the oral rating procedure in subsequent administrations allowed raters to discuss individual performances immediately following interviews before making a score, possibly violating assumptions of independence needed to use inter-rater reliability estimates. The shared two-rater procedure was done to encourage constant "renorming," so that raters were continually seeking agreement during difficulty rating sessions.

Based on Heaton (1988), rating scales for judgment of the writing section (Appendix B) of the test are holistic, as opposed to the analytical/categorical evaluations of oral abilities. The two contrasting styles of rating were adapted to see discrete aspects of oral proficiencies closely but to view writing as an integrated whole. All compositions are read by two readers who are allowed to discuss a paper before assigning a final score. Particularly troublesome compositions are given to a third reader. Gruba (1992) found inter-rater reliability estimates on compositions to be .88 during one administration, though subsequent changes in the procedure may affect this figure.

### Validity

In its present developmental stages, the KEPT cannot be considered a valid examination of English proficiency. Construct validity studies, for example, as well as item revisions, must be completed. In its relation of Kanda University, the KEPT is yet to be placed in a proper operational context within the University, its scores possibly being made part of the permanent student records. University administrators, supportive of the test, are awaiting faculty-wide discussions of the examination. In short, the KEPT is seen as a research project and is not yet recognized as a test of proficiency for the students.

### Conclusion

Set within a communicative framework, the KEPT is an attempt to provide an integrated storyline within a video-mediated format as a way to test English proficiency. The examination is still in a developmental stage, and there is ample proof of Weir's (1990) warning that such tests are "difficult and time consuming to construct, require more resources



to administer, demand careful training and standardization of examiners, and are more complex and costly to mark and report results on" (p. 35). Weir claimed much research is needed to develop true communicative tests of language ability. Clearly though, as the Japanese Ministry of Education further encourages communicative teaching (Goold, Madeley, & Carter, 1993), test developers in Japan must seek to make examinations that are themselves responses to changes in the field.

The use of video testing (along with its technological equivalent, computer assisted testing) comes with several nagging issues. The first concerns are practical ones: the writing of video scripts, the justification to budget committees of the additional cost of video, the selection of actors (age, gender, and race issues may come into play here), sufficient equipment at a testing site to show the video, and the case of producing copies of the test for later sale. The second set of concerns touches on more theoretical issues: test specifications needs to define clear standards of cultural appropriateness, gestures, and other complex sociolinguistic markers, and, finally, construct validation of various language proficiencies when a candidate is examined by a visual medium ("visual proficiency" as opposed to simple listening proficiency?).

Before addressing issues of video-mediated test development, one must first justify that the effort is worth the benefit. For the KEPT project, developers tackled this question in a straightforward manner: they use video in a communicative framework to teach students English proficiency in the classroom (as opposed to audio tapes or simple printed media in a grammar-translation approach) and therefore feel obliged to evaluate the same students using similar methods. Since adopting this approach, other evaluation methods appear, to project members, to be somewhat odd and increasingly, outdated. To continue to test English proficiency in Japan using only printed media and audio tapes in a communicative era for an MTV generation is indefensible.

Practical problems in the KEPT development proved surmountable. Ideas for writing video scripts came from close study of readily available commercial movie scripts, including the formatting of scripts and the provision of standard actor and stage directions. Actor selections were made out of the faculty of a sister institution (Kanda Institute of Foreign Languages), and actors were given the scripts weeks in advance to study. The actual video taping was done in a small on-campus television studio, with backdrops and simple props providing a realistic setting. No students complained about the quality of the production. As for the actual presentation of the test, Kanda University, like many universities in Japan, has a video monitor in nearly every classroom.

As noted above, efforts to exploit totally the advantages of using video by testing all the complexities of real, communicative language (e.g., facial expressions and gestures) have been impeded by the difficulties implicit in such an approach. Test developers could not guarantee that all candidates were watching the video during the "listening" section (video tapes made of the students taking the test later confirmed that many were not continuously viewing the KEPT video but were rather simply listening to the audio portion of it). Additionally, attempts to write items indicating facial gestures, for example, met with unsatisfactory results. At present, the video portion is a complement to the audio portion of the tape, a point that may be resolved when video-mediated tests are brought to the personal computer via multi-media presentations.

Further theoretical issues continue to plague the development of the KEPT. When a listening score is presented, critics have charged that students have done much more than "just listen"—but surely that is the point of using video as a test medium in the first place. Discussions on how to report candidate scores, section by section, remain very much alive. Quite simply, what does an English proficiency score on the KEPT mean? In particular, is a listening score a reflection of "pure" listening abilities, or a much more complex awareness of appropriate gestures, sociolinguistic markers, and other visual cues? Though the test is not yet complete, KEPT developers are beginning to create a defense of the reported scores by first pointing out that scores indicate, to a large extent, the level of ability of a student taught under a communicative paradigm. Since the test is based on a simple storyline, the overall score points to an ability to use information from a variety of sources in a sustained manner. Composition scores reflect an ability to draw on learned experiences in order to build a coherent argument or create a logical extension of an event. Oral scores point to an ability to talk in a group, for a purpose, and not simply to an ability to answer questions in a one-to-one interview.

finally, testing biases created by the inclusion of visuals may need to be investigated. Possible variations of "English proficiency" test scores due to a student's perception of an actor's style of dress, gender, race, and overall appearance may need to be investigated. Who, for example, is a prototypical "native speaker" in style and appearance—certainly the choice of an educated, middle-aged white male would be subject to controversy! Much more work, of course, is needed to finalize the KEPT as a communicative test of foreign language proficiency.

*The author would like to acknowledge the substantial contribution to the KEPT Project by the late Professor Yamada, in addition to the work of all others at Kanda University including the International Studies students who took the examination.*

*Paul Gruba, a former research associate at Kanda University of International Studies, is now studying at the University of Melbourne.*

### References

- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. (1991). What does language testing have to offer? *TESOL Quarterly*, 26, 671-704.
- Benson, V. (1993). Testing made-for-ELT-videos. *The Language Teacher*, 17 (9), 15-19.
- Busch, M., Elsea, M., Gruba, P., & Johnson, F. (1992). Report of the needs analysis committee. Unpublished manuscript, Kanda University of International Studies, Chiba.
- Cooper, R., Lavery, M., & Rinvoluceri, M. (1991). *Video*. Oxford: Oxford University Press.
- Goold, R., Madeley, C., & Carter, N. (1993). The new Monbusho guidelines. *The Language Teacher*, 17 (6), 3-5.
- Grabe, W. (1991). Current developments in second language reading research. , 375-406.
- Gruba, P. (1992). Kanda English proficiency test; Interim report # 3, April 1992. Unpublished manuscript, Kanda University of International Studies, Chiba.
- Gruba, P. (1993). A comparison study of audio and video in language testing. *JALT Journal*, 15, 85-88.
- Gruba, P. (1993b). Kanda English proficiency test; Interim report #6, April 1992. Unpublished manuscript, Kanda University of International Studies, Chiba.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.
- Heaton, J. (1988). *Writing English language tests*, (New ed.). New York: Longman.
- Henning, G. (1987). *A guide to language testing*. New York: Newbury House.
- Johnson, F., Chun, V., Elsea, M., Higa, H., Miller, S., Moren, B., Smillie, B. (1991). Research project #2: English proficiency testing. Interim report #1: June, 1991. Unpublished manuscript, Kanda University of International Studies, Chiba.
- Low, G. (1986). Storylines and other developing contexts in use-of-language test design. *Indian Journal of Applied Linguistics*, 7, 15-35.
- McLuhan, M. (1964). *Understanding media: The extension of man*, (2nd ed.). New York: McGraw-Hill.

- Murphy, J. (1991). Oral communication in TESOL: Integrating speaking, listening and pronunciation. *TESOL Quarterly*, 25, 51-74.
- Nunan, D. (1988). *The learner centered curriculum*. Cambridge: Cambridge University Press.
- Popham, W. J. (1990). *Modern educational measurement* (2nd. ed.). Englewood Cliffs, NJ: Prentice Hall.
- Savignon, S. (1991). Communicative language teaching: State of the art. *TESOL Quarterly*, 25, 261-277.
- Stempleski, S., & Tomalin, B. (1990). *Video in action*. Englewood Cliffs: NJ: Prentice Hall.
- Tatsuki, D. H. (1993). Interactive video and hypermedia: Where's the beef? *The Language Teacher*, 17, 19-24.
- Weir, C. (1990). *Communicative language testing*. London: Prentice Hall International.
- Widdows, S., & Voller, P. (1991). PANSI: A survey and ELT needs of Japanese university students. *Cross Currents*, 18, 127-141.

## Appendix A

### KEPT Outline

- |   |   |
|---|---|
| <p>1. <i>Title page</i><br/>University name<br/>KEPT title<br/>Booklet number</p> <p>2. <i>Instructions</i> (In English, administered orally by proctors)<br/>Example of card marking<br/>Explain sections<br/>Explain total time<br/>Explain regulations</p> <p>3. <i>Introduction</i> (video)*<br/>All five characters introduced<br/>Business meeting setting<br/>Explain country location<br/>Explain purpose of trip<br/>Explain each member's role<br/>*No items (introduction only)</p> <p>4. <i>Memo/Factual Passage</i><br/>Link to intro video<br/>Added business context<br/>Date/time<br/>Format for literal comprehension<br/>Explain country</p> <p>5. <i>Items 1-10</i> (reading items related to memo)</p> <p>6. <i>Narrative Passage</i><br/>Arrive in country<br/>Tourist activity<br/>Problem introduced</p> | <p>7. <i>Items 11-30</i> (reading items related to the narrative passage)</p> <p>8. <i>Cloze Passage</i><br/>Problem continues<br/>Business decision attempted</p> <p>9. <i>Items 31-60</i> (Grammar)</p> <p>10. <i>Listening 1 Monologue</i><br/>Problem explained from one point of view</p> <p>11. <i>Items 61-70</i> (with graphics)</p> <p>12. <i>Listening 2 Monologue</i><br/>Problem explained from second point of view</p> <p>13. <i>Items 71-80</i> (Listening)</p> <p>14. <i>Listening 3 Dialogue</i><br/>Problems discussed by two team members; dilemma</p> <p>15. <i>Items 81-90</i> (Listening)</p> <p>16. <i>Composition</i> (Writing)<br/>Prompt to continue story<br/>Prompt to analyze story</p> <p>17. <i>Student Survey Items</i><br/>Impression of KEPT<br/>Likes/dislikes</p> <p>18. <i>Oral Interview</i> (Speaking)<br/>Prompt to explain event in video<br/>Description of photo from video<br/>Read aloud<br/>Discussion of solutions</p> |
|---|---|

### *Main Characters*

Mike Brennan, 29, American. Mike served in the US Army in Okinawa and learned to speak Japanese during his military service. His hobby is photography and he is obsessed with health and fitness.

Karen Sampson, 34, Canadian. With an MBA in international finance, Karen is quickly gaining a reputation in the company as an intelligent and efficient business woman.

Tomoko Watanabe, 27, Japanese. After completing an economics degree in the States, Tomoko returned to Japan hoping to start her own business. She is currently gaining experience while trying to save money.

Jim Garrick, 32, British. A child of a British diplomat, Jim has grown up in a number of different countries and speaks a variety of languages.

David Lu, 23, Australian. He recently completed a degree in fine arts and came to Japan to earn money before going on to get a Masters in fine Arts.

### Appendix B

Kanda English Proficiency Test Composition Scoring Guidelines		
18-20	Excellent	Virtually free of errors, complex and appropriate structure, task addressed with facility and creativity
16-17	Very Good	Minimal errors, clearly competent, varied use of structure, task addressed with support
12-15	Good	Occasional errors; simplistic vocabulary; task sufficiently addressed but not fully supported
8-11	Acceptable	Multiple errors, flawed structure; distracting vocabulary; task marginally addressed
5-7	Weak	Serious errors; task completion inadequate; inappropriate vocabulary usage
0-4	Very Poor	Errors bar communication; lack of basic structure; insufficient vocabulary; task not addressed

Compositions are to be read impressionistically (as opposed to analytically). A reading pace of two to three minutes per composition is recommended for a total of 20 to 30 compositions per hour per reader. Effort should be made to give equal time to each paper. Breaks should be taken in case of fatigue to minimize error in judgment. Readers should clearly mark their initials to an assigned score for identification.

Markers will award scores based on subjective impressions, but should take care in assigning a mark merely based on (a) what is believed the student wanted to write; (b) general appearance of the students' handwriting; and (c) previous knowledge of the student. firm understanding of the composition task, adequate composition marking training, and frequent reevaluation of scores in conjunction with colleagues will increase substest reliability.

Each composition will be read by a minimum of two readers; a difference of three or more points between raters will require a rating of a third reader. A

score will be considered settled if the third reader is in agreement with one other reader within a spread of two points. A total score will be assigned to each composition based on the sum of readers' scores divided by the total number of readers.

### Appendix C

Kanda English Proficiency Test Oral Rating Scales					
Rating	Speech Flow	Pronunciation Accuracy	Grammatical Accuracy	Expressiveness	Total
5	High degree of fluency, effortless, smooth, natural rhythm	Rarely mispronounces; l/r/lh/v/f problems	Uses high level discourse structure; occasional errors	Interacts as a leader; body/facial expressions, genuinely at ease	
4	Speaks with facility and confidence; paraphrases; occasional pauses	Accent rarely disturbs native listener	Full range of basic structures; mistakes do not interfere with communication	Frequent eye contact; some body language; interacts with facility	
3	Hesitant speech; some paraphrasing; noticeable errors in speech rhythms	Often faulty but intelligible	Meaning expressed in accurate simple sentences; complex grammar avoided	Self-conscious, formal; responds with caution; waits for prompting	
2	Slow, strained, except for routine expressions	Frequent errors; intelligible to NS accustomed to NNS	Errors frequent but intelligible to NS accustomed to NNS patterns	Stiff; fixed posture; hesitant to contribute; requires prompting	
1	So halting conversation is nearly impossible	Unintelligible without effort	Fragmented phrases; single words	No outward signs of willing participation	

Oral ratings will be assigned to examinees on the basis of both reading aloud and conversation with three other examinees in a period of fifteen minutes.

It is recommended that ratings for speech flow be assigned to all four examinees immediately following individual read aloud sessions. Other categories should be assessed in the course of the conversation, with the score for expressiveness being the last to be assigned. It is imperative that raters try to allow each examinee equal attention before scoring oral abilities.

Oral evaluation forms should be complete filled out, including total scores, and initialized before a new group of examinees is allowed to begin.