

# Research Forum

## Evaluating Learner Self-Assessment

**Colin Painter**

*Prefectural University of Kumamoto*

This exploratory study examines Pearson product-moment correlations between learner and teacher-assessment in a CAI (Computer Assisted Instruction)-based communicative English course for Japanese university students. It also explores the validation of the program-specific tests used for self-assessment through correlation of the students' self-assessed test scores with their TOEIC scores. Although the self-assessment scores did not correlate significantly with all parts of the TOEIC, significant correlations of self-assessment were observed with teacher assessment, suggesting the reliability of the self-assessment procedure.

本論では、日本の大学におけるコンピュータを利用した英語コミュニケーション授業場面を取り上げ、学習者自身による評価と教師による評価とのピアソン相関係数の分析を行った。さらに、学習者の自己評価得点とTOEIC得点の相関分析を行い、この自己評価に用いたテストの妥当性を検証した。その結果、自己評価点は、TOEICの全部のパート別得点に有意な相関関係があるわけではないが、自己評価得点と教師による評価との間には有意な相関が認められた。このことは、自己評価の信頼性の高さを示唆している。

**T**his exploratory study examines the following aspects of learner self-assessment: (1) whether learner and teacher assessment have positive correlations, thus indicating the reliability of the learners' self-scoring; and (2) whether the role-play tests used for assessment have positive correlations with a standardized test. The study also examines whether the number of self-assessment tests increased compared with the number of teacher-assessed tests reported previously (Painter, 1995).

The following review explores the positive results of studies on learner self-assessment and addresses the necessity of establishing the reliability and validity of the program-specific test used for self-assessment activities.

### Learner Self-Assessment

Studies on learner self-assessment are relatively few but report generally positive results. From 1967 to 1998 *TESOL Quarterly* published only one article containing "self-assessment" in the title (LeBlanc and Painchaud, 1985). This paper examined students' ability to self-assess levels in French and English as a Second Language using a questionnaire for placement purposes. Pearson product-moment correlations between a proficiency test and two types of self-assessment questionnaires were .80 and .82. Thus, the authors concluded that self-assessment was valuable as a placement instrument.

Since its founding in 1985, *Language Testing* has published seven papers relevant to the area of self-assessment (Bachman & Palmer, 1989; Blanche, 1990; Heilenmann, 1990; Janssen van Dieten, 1989; Oscarson, 1989; Ross, 1998; Shameen, 1998). One of the most recent (Ross, 1998) includes a meta-analysis of the correlations contained in a number of studies made since 1978 (Bachman & Palmer, 1981, 1982; Blanche, 1990; Buck, 1992; Ferguson, 1978; Janssen van Dieten, 1989; LeBlanc and Painchaud, 1985; Milleret, Stansfield & Mann-Kenyon, 1991; Wongsotorn, 1981). These included research across the four language skills within a wide range of second and foreign language contexts. The criterion Ross employed to select these studies for analysis was the presence of "an empirical basis for evaluating the relationship between self-assessment and a second or foreign language criterion variable" (p. 2). Examining the Pearson product-moment correlations between self-assessment and speaking skills, Ross found the average to be .55 ( $p < .05$ ) for the 29 self-assessments of speaking within the ten studies. Looking at the total of 60 self-assessments across the four language skills, Ross found a correlation of .63 ( $p < .05$ ). Thus, Ross concluded that self-assessment typically offers "robust" concurrent validity with criterion variables.

Other researchers have also made a case for self-assessment. Murphey (1994) noted the ability of a test not only to measure but to stimulate learning. He requested that his students make their own tests and test each other. Believing that there is insufficient time to test everyone orally, he sacrificed teacher control and encouraged students to test each other, inside or outside the classroom.

Computer-assisted Instruction (CAI) is also suggested to engender a learning environment which promotes learner autonomy. Peterson (1997) believes that computer-mediated instruction (CMI) promotes learner autonomy in that it provides a less restrictive learning environment than the traditional language classroom. Citing Cooper and Selfe (1990),

Peterson feels CMI is compatible with personal learning styles and encourages the learner to take control of the learning process.

Following the positive views of both self-assessment and CAI, this exploratory study argues for the reliability of student self-assessment made using course-specific tests given in a CAI class for communicative English. Correlational evidence is provided showing a positive relationship with teacher assessment and with some sections of a well-known test of English language proficiency.

### Test Types and Criterion-Related Validity

Validity issues usually concern two types of test, Criterion Referenced Tests (CRTs) and Norm Referenced Tests (NRTs). Brown (1995) discusses several characteristics which distinguish CRTs from NRTs, and suggests that the most fundamental is the purpose of the test. He notes that CRTs foster learning and are typically used by teachers to encourage students to study, review, or practice the material in a course. On the other hand, the basic purpose of NRTs is to spread students' performances out so that they can be classified for admission or placement (Brown, 1995, p. 13; 1998). CRTs are more likely used to discover how much of a given level of ability or content domain the test-takers have learned, for example, when a teacher gives a test at the end of a unit of language study. The focus of the CRT, then, is on the relationship between the learner/test-taker *and the material*, whereas the focus of the NRT is on comparing the learners' performances with *one another*.

The CRT, which is based on the syllabus of a course, is likely to have beneficial washback effect on the learners, encouraging them to take the syllabus seriously. After the test, teachers can go through the test questions with the learners, making it a teaching tool. However, NRT test-takers may never learn their mistakes since the NRT paper is less likely to be returned to test-takers. In fact, there may be no direct connection between the multiple-choice questions in the NRT and the syllabus of the course. An important question, then, is whether different CRTs are valid measures of the learners' language skills in general.

Among the different types of validity, *criterion-related validity* is particularly important since it indicates the extent to which scores on one test will estimate or predict performance on other tests measuring the same ability. The primary way of establishing criterion-related validity is by correlating the test in question with another test which is well established and measures the same ability. Although a major issue in test design is the extent to which syllabus-based CRTs can be used as valid indicators of learners' proficiency, Brown (1988, 1995) notes that it is

often not possible to use an NRT to validate a CRT since they measure different things, the CRT testing mastery of specific course content and the NRT being a more global measure of language proficiency.

Complicating the validation process of specific CRTs is the lack of a CRT which is well established and is thus appropriately representative of the ability criterion. Bachman (1990) points out that there is a strong need to develop valid criterion-referenced measures of communicative language ability. He feels there is a need for a "common yardstick" (p. 334) and that CRTs would fulfil this need. A recent paper by Nakamura (1995) laments the absence of a relevant CRT which could be used for establishing *concurrent validity* (p. 129), that is, the extent to which results on two tests administered at the same time correlate significantly with each other. He used students' grades in conversation classes and compared them with teacher estimates of their speaking ability to investigate concurrent validity.

Thus, although varied learning situations and their accompanying syllabuses cause difficulties in defining a common level of ability, making the "common yardstick" elusive, both NRTs and CRTs have an important role in program evaluation (Lynch, 1992) and in measuring learning. Mindful of the difficulty of using an NRT to validate CRTs, this exploratory research nonetheless uses an well-known NRT to test the validity of the type of CRT assessment test used in this study.

#### *Validity of the TOEIC*

The Test of English for International Communication (TOEIC), developed by The Educational Testing Service (ETS), is an example of an NRT used in language education. Although it does not directly test oral skill, the TOEIC is a well-established language test. MacGregor (1997) suggests that both the TOEIC and the TOEFL are regarded as valid instruments because ETS regularly publishes reliability and validity reports on their use. She cites Wilson (1993) on the link between TOEIC listening scores and the scores on the Language Proficiency Interview (LPI), a direct assessment of oral language proficiency developed by the Foreign Service Institute of the US government. The correlation between the LPI and the TOEIC listening was a consistently high .83, "suggesting that both tests are, as they claim, effective measures of the ability to understand and use spoken English" (p. 32). MacGregor also cites Woodford (1992) who reports that, "in 1989 and 1990, test reliability for TOEIC using the KR-20 formula was .96" (p. 35).

In this report, correlational analysis of learner self-assessment is conducted, using the TOEIC to assess the criterion-related validity of the self-assessment process.

## The Study

This exploratory study investigates learner self-assessment during three years of a university CAI oral communication program, 1995-1997. A previous report (Painter, 1995) described how the program aimed at the development of oral communication using computers and how paired learners requested testing through role play after they had completed a unit of functionally-based language activity. The role-play test scores were analyzed for both test-retest reliability and intra-rater reliability (Painter, 1997b) and in both cases the Pearson product-moment correlation coefficient was .88 ( $p < .05$ ), indicating a significant test-retest correlation (see Painter, 1997b for details). Moreover, test validity was indicated since (1) the ability domain was based on the course outline, and (2) the test scores, as well as the number of tests requested by the students, correlated significantly with cloze test scores (Painter, 1997b). However, it was suggested that further correlation studies of the role-play tests would provide more convincing evidence of criterion-related validity. The participants of the study provided this opportunity when they subsequently took part in the TOEIC, allowing for comparison of the role-play test scores with their TOEIC scores.

### *Research Focus*

Three areas regarding learner self-assessment are explored in this limited report:

- (1) Investigation of how self-scored testing affects the pace of learning, as reflected in the number of tests taken during the years of self-assessment compared with the number taken during the period of teacher-assessment.
- (2) Investigation of the reliability of the course-specific role-play tests by examining the relationship between learner and teacher scoring.
- (3) Investigation of the criterion-related validity of the role-play tests by correlating learner self-assessment scores with a widely used reliable and valid test, the TOEIC.

## Method

### *Participants*

Learners at the Prefectural University of Kumamoto, Faculty of Administration are of mixed gender (M:F; 46:54). Classes are ninety minutes in length and the CAI Oral English class is offered once weekly for first-year learners and once biweekly for second-year learners. A total of 151 stu-

dents participated in this study, and five of the six groups took the TOEIC test, as shown in Table 1.

### *Description of the Program, Testing, and Test Scoring*

#### *The CAI Program*

First-year learners begin the CAI program using a situational/functional English software program titled *Nova City, Beginner* (Milward, 1993), containing five units and tests. The units included such topics as "At the Airport," "Checking into a Hotel," and so forth. The second-year learners used the next course in the series, *Nova City, Intermediate*, containing 20 units and tests.

#### *Scoring of the Assessment Tests*

The twenty-five performance tests used in the CAI program were CRTs in the form of role-plays derived from the material studied in class (see Painter, 1996, for a full description of the test development process). Pairs of students were requested to perform a role-play based on the material they had just studied. In 1995, the first year of the program, all tests were administered and scored by the teacher. The scoring procedure used during teacher assessment went as follows:

1. Communication was meaningful and grammatically correct:  
2 points for each section
2. Communication was meaningful but contained grammatical errors:  
1 point for each section
3. Communication was meaningless:  
0 points for each section

Table 1: Participants in the Study

Year	Students' year	Number of classes	Learners completing 2 semesters of CAI	Learners taking TOEIC (N= 151)
1995	1st	26	48	22
	2nd	13	48	none*
1996	1st	26	49	29
	2nd	15	43	17
1997	1st	27	47	45
	2nd	16	50	38

\*The 1995 second-year learners did not take the TOEIC

Here a "section" refers to a section of dialogue, such as an initiating remark, question, response, or closure. This scoring method attempted to reduce the items the assessor needed to keep track of during the test (Underhill, 1987).

A subsequent study (Painter, 1997b) indicated that learners sometimes had to compete for the chance to test, possibly dampening the positive effects of autonomy and slowing down the assessment process. To learn more about the relationship between performance opportunities and proficiency it was felt necessary to provide unrestrained opportunity for testing. It was thus suggested (Painter, 1997b) that further research should include self-testing and self-grading by learners. This would enable learners to move through the program at their own pace, without any impediment caused by the teacher-administered testing process.

#### *Learner Self-Assessment*

Since 1996, learners have graded themselves upon finishing their role-play test at the end of a unit. Since learners were both participants as well as assessors of the test, it was impossible to score sections of the test without interrupting the testing process. Therefore scoring took place after each test. Following the teacher scoring guidelines above, the learners were required to estimate an accuracy level for "Meaningful Communication," then estimate "Grammatical Accuracy." These terms were carefully explained in a guide and exemplified by the teacher at the beginning of the course. The learners were informed that 20% of their final grade would come from the self-assessed test scores.

A one-page English-language *Procedure Guide* was issued to the learners from the first semester in 1995. A revised five-page English-language guide was issued in 1996, and in 1997 the *Procedure Guide* was issued bilingually (Painter, 1997a).

#### *Correlational Analysis*

For the purpose of comparison between learner and teacher-assessment, simultaneous scoring began in 1996. Twenty-three categories were used for analysis, as shown in Figure 1. Some categories, such as "grade" and its components such as "attendance," are self-correlated. However, in the interest of comprehensive investigation, all categories were recorded for comparison. Spreadsheets with Pearson's product-moment correlation matrixes were produced representing the data from each of the learner groups. Only a small portion of this data is generated for the present report.

The learners' TOEIC test results were used for the purpose of comparing self-assessment with a validated test. Data was recorded over the six semesters covered by the study, 1995-1997. Two groups of first-

Figure 1: Correlation Categories

- 
1. Learner self-assessed performance (1 time only, 7/1996)
  2. Teacher scored performance (1 time only, 7/1996)
  3. TOEIC listening score
  4. TOEIC reading score
  5. TOEIC overall score
  6. Cloze score, first semester
  7. Cloze score, second semester
  8. Cloze score, average
  9. Learner self-assessed average performance score, first semester
  10. Learner self-assessed average performance score, second semester
  11. Learner self-assessed average performance score
  12. Performance test quantity, first semester
  13. Performance test quantity, second semester
  14. Performance test quantity, total
  15. Homework quantity, first semester
  16. Homework quantity, second semester
  17. Homework quantity, total
  18. Attendance, first semester
  19. Attendance, second semester
  20. Attendance, average
  21. Grade, first semester
  22. Grade, second semester
  23. Grade, average
- 

year learners were studied in both semesters of 1995. However, the TOEIC was not taken by the 1995 second-year learners, therefore only basic data appears for them. Two groups of first and second-year learners were studied in both semesters of 1996. Also, two groups of first and second-year learners were studied in both semesters of 1997. The data for TOEIC-takers from identical learner-year groups is combined for the purpose of the correlation study. Pearson product-moment correlation matrixes were made for all learner groups. The data contained in the tables below is derived from the matrixes, and a descriptive statistics table appears in the Appendix. Space limitation prevents the display of the matrixes themselves.

## Results

### *Test Quantity and Self-Assessment*

During 1995, the period of teacher-assessment, the first-year learners took an average of nine assessment tests, these scored by the teacher



(Table 2). In 1996, with self-assessment, there were 12 tests per first-year learner, an increase of 33%, and in 1997, these learners took 13 tests. Interestingly, the average score of tests remained the same, at about 79%, regardless of whether assessment was made by the teacher or the learners. Second-year learners receiving teacher assessment took only four tests, but when conducting self-assessment in 1996, they took an average of six tests, with an average score of 75%, an increase in output of 50%. The average scores of the 1997 second-year learners were almost the same at 77%, while test quantity was the same, at six tests during the year. Thus, both first- and second-year learners took more tests when self-assessing, and the self-assessment procedure did not appear to result in inflated scoring.

Table 2: Influence of Self-Assessment on Test Quantity & Average Score

Year	Year	Average Test Score**	Number of Tests Taken**
1995*	1st	79	9
1996	1st	79	12
1997	1st	80	13
1995	2nd	74	4
1996	2nd	75	6
1997	2nd	77	6

\* Only teacher-assessment was used in 1995

\*\* Values for test scores and number of tests taken have been rounded

### *Teacher and Learner Assessment Compared*

In the first semester of 1996, 68 tests were scored simultaneously, both by learner self-assessment and by the teacher. To compare the reliability, a one-time correlational analysis of self-assessment and teacher-assessment using the tests given in July, 1996 was performed, and the results are shown in Table 3. First-year learner self-assessment and teacher-assessment correlated significantly at .53 ( $p < .05$ ). The correlation of  $r = .66$  ( $p < .05$ ) for the second-year assessments was also significant.

### *Correlational Analysis of Learner Assessment Scores with the TOEIC*

Table 4 shows first-year and second-year learners' scores correlated with the TOEIC for 1996 and 1997, first-semester and second-semester tests, and the two sets of scores for each year combined and re-correlated.

Table 3: One-Time Correlation of  
Learner Self-Assessment and Teacher-Assessment

Year	Year of Study	Number of Students	Correlation
1996	1st	29	.53*
	2nd	17	.66*

\* Significant ( $p < .05$ )

In the first semester of 1996, the first-year learners' self-assessment indicated a weak non-significant correlation with TOEIC Overall, as shown in Table 4 below. However, the second-year learners' scores had significant correlations with TOEIC Listening, Reading and Overall Total, at  $r = .46$  ( $p < .05$ ),  $r = .42$  ( $p < .05$ ) and  $r = .54$  ( $p < .05$ ) respectively.

The second-year 1997 learners' TOEIC scores dated from 18 months prior to their participation in the CAI program, and there was no significant correlation between those scores and the scores obtained in the program (Table 4). However, for the first semester of 1997, the first-year learners' self-assessment average correlated significantly with both TOEIC Listening, at  $r = .35$ , and TOEIC Overall Total at  $r = .29$ .

Only eight significant correlations out of 36 were observed between the TOEIC and the self-assessment scores of the learners, with three of the eight coming from the larger number of tests represented in the combined first and second semester scores. Therefore, the validity of learner self-assessment receives only slight support from correlation with the learners' TOEIC scores.

Table 4: Correlation of Self-Assessed Average Performance Scores  
with TOEIC

Year	1996						1997					
	First			Second			First			Second		
Learner year of study Semester of self-assessment	1	2	1+2	1	2	1+2	1	2	1+2	1	2	1+2
N	29	29	29	17	17	17	45	45	45	38	38	38
TOEIC listening	.22	.18	.24	.30	.46*	.41*	.35*	.24	.30*	-.06	.05	.01
TOEIC reading	.13	.28	.25	.29	.42*	.38	.17	.08	.13	-.02	.19	.09
TOEIC total	.18	.26	.27	.36	.54*	.48*	.29*	.18	.24	-.06	.13	.04

\*Significant ( $p < .05$ )

### Discussion

In the CAI program, completing a unit of study was a pre-condition for taking a role-play assessment test. Consequently, the number of tests taken implies the pace of study. With sizeable groups of learners, having the teacher assess every learner pair's role-play is impractical and is believed to slow down the learners' progress (Painter, 1997b). In this program, the transition to self-assessment resulted in an increased pace of learning without an accompanying inflation of grades through the self-scoring procedure. The increase of between 33% and 50% in the number of tests taken, with stability of scoring maintained, observed under self-assessment suggests that self-assessment has a positive influence on the pace of learning.

However, the increased number of tests taken without inflated self-grading, in itself, is not sufficient to establish the reliability of the self-assessment procedure. It is also desirable that learner self-assessment be significantly correlated with teacher-assessment. In this study, first-year and second-year learner self-assessment scores on one test correlated significantly with teacher-assessment, suggesting reliability in self-assessment. Clearly, however, wider correlational studies are necessary.

Concerning validity, self-assessment was examined for correlation with the TOEIC, a validated NRT. As noted, the purposes of NRTs such as the TOEIC, and CRTs, which are program-specific tests measuring learner mastery of what has been taught, are quite different and one should not necessarily expect significant correlations. In this study, only a few significant correlations were observed. Further research is also necessary in this area.

### Conclusions

The results of this exploratory study suggest that self-assessment enhances the output of performance while retaining stability of scoring. Reliability of the self-assessment process was suggested by the significant correlation between learner and teacher scoring procedures on a single test. Only limited confidence, however, is suggested concerning the criterion-related validity of the self-assessment test due to the small number of significant correlations between parts of the TOEIC and the self-assessed role-play tests.

Further research should consider the need for larger groups, perhaps assembled by combining results from several classes of learners being taught by similarly interested teachers. A training period would be necessary in which learners are first tested on their grasp of the criteria for

self-assessment, followed by a period to harmonize their self-assessment ratings. In this way, reliable results could be produced from subsequent correlation studies. Teacher-researchers are encouraged to try out self-assessment in their teaching situations.

The learners in this study were certainly enthusiastic about the opportunity to assess themselves and the washback effect was evidenced by the 33%-50% increased output noted. Tying self-assessed scores to a modest percentage of the grade, such as the 20% in this study, convinces learners that they are being taken seriously.

#### Acknowledgements

*This is a version of a paper presented at the Japan Association of College English Teachers (JACET), 36th Annual Convention Program, Waseda University, Tokyo. The author is grateful for advice given at the beginning of the program, particularly by Dr. Thomas Robb, Chair, English Department, Kyoto Sangyo University and Dr. John Shillaw, Tsukuba University. Thanks are due to the two anonymous JALT Journal reviewers for their valuable suggestions, as well as to the students who participated in the study. Gratitude is expressed toward colleagues for their support.*

*Colin Painter* is an Associate Professor at the Prefectural University of Kumamoto. He has taught at universities in Asia for the last 16 years. His interests include language acquisition, curriculum development, and computer-assisted language learning.

#### References

- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. & Palmer, A. (1981). The construct validity of the FSI Oral Proficiency Interview. *Language Learning*, 31, 67-86.
- Bachman, L.F. & Palmer, A. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16 (4), 449-65.
- Bachman, L.F. & Palmer, A. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, 6 (1) 14-29.
- Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: The DLIFLC study. *Language Testing*, 7 (2), 202-229.
- Blanche, P & Merino, B. (1989). Self-assessment of foreign language skills: Implications for teachers and researchers. *Language Learning*, 39, 323-340.
- Brown, J.D. (1988). *Understanding research in second language learning*, Cambridge: Cambridge University Press.
- Brown, J.D. (1995). Differences between norm-referenced and criterion-referenced tests. In J.D. Brown & S.O. Yamashita (Eds.). *Language Testing in Japan* (pp. 12-19). Tokyo: The Japan Association of Language Teaching.

- Buck, G. (1992). Listening comprehension: Construct validity and trait-characteristics. *Language Learning*, 42, 313-57.
- Cooper, M.M. & Selfe, C.L. (1990). Computer conferencing and learning: Authority, resistance and internally persuasive discourse. *College English*, 52 (8), 847-869.
- ETS (Educational Testing Service). (1992). *Guide to SPEAK*. Princeton, NJ: Educational Testing Service.
- Ferguson, N. (1978). Self-assessment of listening comprehension. *International Review of Applied Linguistics*, 16, 146-156.
- Heilenmann, L.K. (1990). Self-assessment of second language ability: The role of response effects. *Language Testing*, 7 (2), 174-201.
- Janssen van Dielen, A. (1989). The development of a test of Dutch as a second language: The validity of self-assessment by inexperienced subjects. *Language Testing*, 6 (1), 30-46.
- LeBlanc, R. & Painchaud, G. (1985) Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19 (4), 673-687.
- Lynch, B. (1992). Evaluating a program inside and out. In J.C. Alderson & A. Beretta (Eds.). *Evaluating second language education* (pp. 61-99). Cambridge: Cambridge University Press.
- MacGregor, L. (1997). The *Eiken* test: An investigation. *JALT Journal*, 19 (1), 24-42.
- Milleret, M., Stansfield, C. & Mann-Kenyon, D. (1991). The validity of the Portuguese speaking test for use in a summer study abroad program. *Hispania*, 74, 778-787.
- Milward, M. (1993). *Nova City*. (CD-ROM) Tokyo: Nova Information Systems.
- Murphey, T. (1994). Tests: learning through negotiated interaction. *TESOL Journal*, 4 (2), 12-16.
- Nakamura, Y. (1995). Making speaking tests valid: Practical considerations in a classroom setting. In J. D. Brown & S. O. Yamashita (Eds.). *Language Testing in Japan* (pp. 126-133). Tokyo: The Japan Association of Language Teaching.
- Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6 (1), 1-13.
- Painter, C. (1995). Developing oral communication using computers: Computer assisted language learning. *Administration*, 2 (3), 109-150.
- Painter, C. (1996). *Performance Tests*. Kumamoto: Prefectural University of Kumamoto, Foreign Language Education Center.
- Painter, C. (1997a). *Procedure Guide For Using Software* (Bilingual) Mimeograph. Kumamoto: Prefectural University of Kumamoto, Foreign Language Education Center.
- Painter, C. (1997b). Continuous assessment facilitated by CAI. In S. Cornwell, P. Rule & T. Sugino (Eds.). *On JALT 96, Crossing Borders* (pp. 119-125). Tokyo: The Japan Association for Language Teaching.
- Peterson, M. (1997). Language teaching and networking. *System*, 25 (1), 29-37.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15 (1), 1-20.
- Shameen, N. (1998). Validating self-reported language proficiency by testing

- performance in an immigrant community: The Wellington Indo-Fijians. *Language Testing*, 15 (1), 86-108.
- Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.
- Wilson, K. (1993). Relating TOEIC scores to oral proficiency interview ratings. *TOEIC Research Summaries 1*. Princeton, NJ: Educational Testing Service.
- Wongsotorn, A. (1981). Self-assessment in English skills by undergraduate and graduate students in Thai universities. In J. Read (Ed.) *Directions in language testing* (pp. 240-260) Singapore: Singapore University Press.
- Woodford, P. (1992). A historical overview of TOEIC and its mission. *The 35th TOEIC seminar* (pp. 10-15). Tokyo: The Institute for International Business Communication.

(Received October 5, 1997; revised December 21, 1998)

## Appendix: Descriptive Statistics Table

		1 L Perf 1 time	2 T Perf 1 time	*3 TOEIC Listen	4 TOEIC Read	5 TOEIC Total	6 Cloze sm 1	7 Cloze sm 2	8 Cloze Ave	9 L Perf sm 1	10 L Perf sm 2	11 L Perf Ave	12 Perf 1 tests	13 Perf 2 tests	14 Perf Total
1995-6 First Year	n			22	22	22	22	22	22	22	22	22	22	22	22
	Administered	N/A	N/A	10/96	10/96	10/96	7/95	12/95		~7/95	~1/96		sm 1	sm 2	
	SD			39.0	44.9	64.5	6.9	8.4	6.6	8.4	9.4	7.1	0.7	0.6	1.1
	Mean			194	135	329	85	73	79	80	78	79	4	4	9
1996-7 1st Year	n	29	29	29	29	29	29	29	29	29	29	29	29	29	29
	Administered	7/96	7/96	5/96	5/96	5/96	7/96	12/96		~7/96	~1/97		sm 1	sm2	
	SD	16.0	17.0	30.2	48.3	72.4	9.0	7.9	7.8	8.7	6.9	6.8	1.0	1.5	2.1
	Mean	71	77	197	155	352	80	72	78	76	82	79	7	6	12
1996-7 2nd Year	n	17	17	17	17	17	17	17	17	17	17	17	17	17	17
	Administered	7/96	7/96	10/96	10/96	10/96	7/96	12/96		~7/96	~1/97		sm 1	sm2	
	SD	15.2	16.1	47.7	45.7	76.7	7.2	11.4	8.4	11.1	13.1	11.5	0.7	1.1	1.5
	Mean	69	83	217	164	382	53	54	53	75	75	75	3	3	6
1997 1st Year	n			45	45	45	45	45	45	45	45	45	45	45	45
	Administered	N/A	N/A	5/97	5/97	5/97	7/97	12/97		~7/97	~1/98		sm 1	sm2	
	SD			46.5	49.7	85.9	7.1	12.2	8.4	8.8	10.6	9.3	1.0	1.5	1.9
	Mean			221	150	371	87	75	81	78	81	80	6	7	13
1997-8 2nd Year	n			38	38	38	38	38	38	38	38	38	38	38	38
	Administered	N/A	N/A	5/96	5/96	5/96	7/97	12/97		~7/97	~1/98		sm 1	sm2	
	SD			43.2	34.9	64.8	12.1	11.8	10.4	7.6	5.7	5.8	0.8	0.9	1.6
	Mean			209	172	381	62	59	60	76	78	77	1.9	3.9	5.6

sm: Semester  
Perf: Performance  
L: Learner  
T: Tests: Test Quantity  
\*TOEIC administration may pre- or post-date year of study

SHARE THE REWARDS OF A UNIQUE INVESTMENT OPPORTUNITY

paradigm:  
FUND

For the first time retail investors can access a unique investment opportunity previously only available to institutional investors.

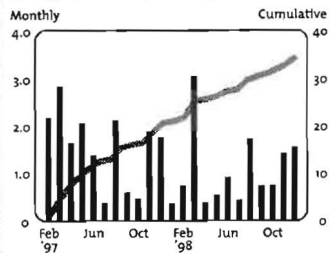
The Paradigm Fund – a superior retail investment product from Banner Japan – accesses expert management to realise the return potential of a formerly exclusive investment sector.

The strategy underpinning the Fund has demonstrated solid, positive returns since inception through varying investment conditions.

Investing only in AAA-rated US Mortgage Backed Securities, the Fund seeks to generate high returns through sophisticated management of this sector's unique risk and return profile.

Utilising the cutting-edge analytics of a pre-eminent US investment house, the Fund aims to deliver a dynamic, risk-controlled investment strategy and a tax effective investment.

Strategy Performance (% Increase)



■ Cumulative Performance ■ Monthly Performance

Note - past performance is no guarantee of future returns.

For a brochure, or more information, contact:

Banner Overseas Financial Services

Tel (03) 5724 5100 Fax (03) 5724 5300

Email [banner@gol.com](mailto:banner@gol.com)

[www.paradigmfund.com](http://www.paradigmfund.com)

BANNER