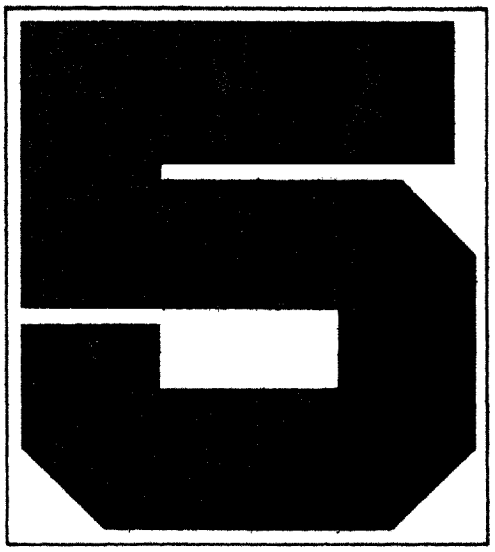
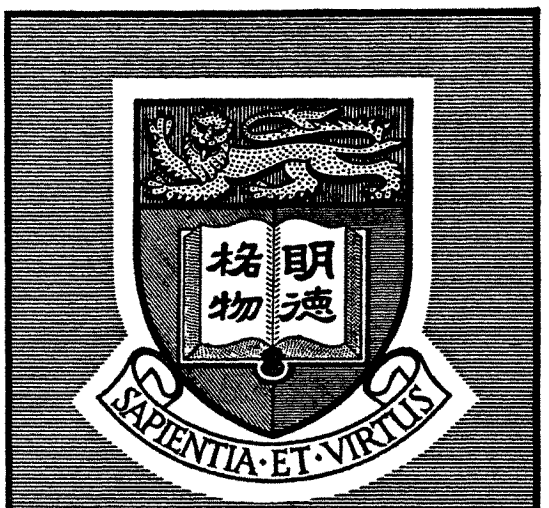


working papers in linguistics and language teaching

university of hong kong, language centre feb. 1982



nos.5-6
(1982)

CONTENTS

ACKNOWLEDGEMENTS

The Editors are greatly indebted to the assistance rendered by Ms. Amy Lam (printing and coordination) and Miss Connie Kwok and Miss Alice Tse (typing), as well as to the University Printing Unit.

POLICY

Working Papers in Linguistics and Language Teaching publishes work in a number of areas, including: general linguistics, language teaching methodology, evaluation of teaching materials, language curriculum development, language testing, educational technology, language and language teaching surveys, language planning, bilingual education. Articles on Chinese and Chinese language teaching may be published in Chinese.
Working Papers is aimed primarily at Hong Kong and, as such, is intended to be informal in character, but it is hoped that it may also be of interest to specialists in other parts of the world, especially in Asia. There are two issues per year.

SUBSCRIPTION

The 1982 issues will cost a nominal HK$5 each or HK$10 for the year (US$2.50 overseas, including postage for two issues). If you are interested in receiving them, please fill in the accompanying slip and return it as directed, or write directly to the Editors.

CONTRIBUTIONS

Contributions in the form of articles, reports, in-depth reviews or simply comments on earlier articles are welcome. They should conform to the style sheet on pages iii and iv and should be addressed to:

The Editors,
Working Papers in Linguistics and Language Teaching,
c/o The Language Centre,
University of Hong Kong,
Hong Kong.

STYLE SHEET

1. Manuscripts should be typewritten, preferably on A4 size paper. Typing should be double-spaced and on one side of the paper only.

2. Items to be italicised should have single underlining. These include the following:

   a. Section headings and subheadings (which should not be numbered).

   b. Words or phrases used as linguistic examples.

   c. Words or phrases given particularly strong emphasis.

   d. Titles of tables, graphs and other diagrams.

   e. Titles or headings of other books or articles referred to or cited.

3. Capitals (no underlining) should be used for the following

   a. Title of article or review. (The author's name(s) may be in smaller type).

   b. Headings of NOTES and REFERENCES sections.

4. Single inverted commas should be reserved for

   a. A distancing device by the author (e.g. This is not predicted by Smith's 'theory'.....).

   b. A method of highlighting the first mention of terms specially coined for the paper.

5. Double inverted commas should be reserved for verbatim quotations.

6. The first page should contain the title of the article at the top of the page, in capitals, with the name of the author(s) immediately below and centred. A reasonable amount of blank space should separate these from the start of the text. Headings such as Introduction should be underlined and located at the left-hand side of the text. There should be two blank spaces between the subheading and the start of the first sentence of the text, which should be indented 5 spaces.

7. Tables and diagrams should each be numbered sequentially and their intended position in the text should be clearly indicated. Diagrams should be on separate sheets. All such graphic displays should have single underlining. Capitals should only be used for the initial letter of the word Table or Diagram and for the first word in the following sentence (e.g. Table 2. Distribution of responses).

8. Footnotes should not be used. Reference in the text should be to author's name, year of publication and , wherever applicable, page or pages referred to (e.g. "This is refuted by Smith (1978a: 33-5). However, several authors take a different view (Chan 1978:13; Green 1980)").

9. Notes which require explanation should be indicated by superscript numerals in the body of the article and should be grouped together in a section headed NOTES (in capitals) at the end of the text. The number and quantity of notes should be kept to a minimum.

10. References should be listed in alphabetical order in a section headed REFERENCES (in capital letters), immediately following the NOTES section.

11. In cases of joint authorship, the name of the main author should be placed first. Where each author has taken an equal share of the work, the names should be sequenced alphabetically. The fact that the names are in alphabetic order may, if so desired, be pointed out explicitly in a note.

12. Journal articles should be referenced in the following way:

    Oller, J.W. and Streiff, V. 1975. 'Dictation: A test of grammar-based expectancies.' English Language Teaching Journal 30(1):25-36.

13. Books and pamphlets should be referenced in the following way:

    Foss, B. (ed.) 1974. New Perspectives in Child Development Harmondsworth: Penguin.

14. Articles in books should be referenced in the following way:

    Kvan, E. 1969. 'Problems of bilingual milieu in Hong Kong: Strain of the two language system.' In Hong Kong: A Society in Transition, edited by T.C. Jarvie and J. Agassi, pp. 327-343. London: Routledge and Kegan Paul.

CONTRIBUTORS

Evelyn Cheung	Language Instructor at the Language Centre, University of Hong Kong
Angela Fok	Lecturer at the Language Centre, University of Hong Kong.
Lee Yick Pang	Lanquage Instructor at the Language Centre University of Hong Kong.
Graham Low	Lecturer at the Language Centre, University of Hong Kong.

CONTENTS OF WORKING PAPERS IN LANGUAGE AND LANGUAGE TEACHING, NOS. 1-4.

No. 3. July 1980

No. 4. June 1981

MEASURING READING ACHIEVEMENT IN A BILINGUAL SITUATION

Angela Fok
Language Centre
University of Hong Kong

Introduction

The purpose of this study is to survey the reading achievement of children studying in different types of schools in Hong Kong[1]. As is often acknowledged, reading is the keystone of the arch of education. It is closely related to learning and personal enjoyment, and the success of the school system depends heavily on the child's ability to interpret the printed page. In this survey, reading achievement is considered as a reflection of the success of a number of school types in helping the child to acquire his/her basic language skills.

Studies of reading achievement tend to be performed on primary school children when they are still in the early stages of reading. A number of research studies have been published in recent years which relate reading attainment to methods of teaching, school organization (Morris 1966), intelligence (Cockburn 1973), parent's occupation and parental interest in the child's education (ILEA 1966). Amongst these surveys is the study reported by Morris (1966) who tried to compare the reading attainment of primary school children in Kent with that of children in Britain as a whole, and to relate these attainments to methods of teaching, type of school organization and non-verbal ability of pupils concerned. The survey confirmed the superior reading standard of the students under observation as well as establishing a strong relationship between reading attainment and socio-economic conditions. Moreover, Morris' results also suggest greater success for whole-word methods than for phonic methods, the difference being statistically significant however only after adjusting for school differences in non-verbal ability. Morris also investigated the effects of the type of school organization (junior without infants versus junior with infants), but this variable is largely confounded with that of teaching method.

Turning to second language reading and testing in bilingual education, both Modiano and Osterberg discovered that experiences gained in the first language are clear assets to second language acquisition. In Mexico, Modiano (1968) found that teaching Indian children to read their own Indian language before introducing Spanish texts led to superior achievements in reading Spanish than was the case if initial teaching was purely in the second language -- Spanish. Similarly Osterberg (1961) in Sweden discovered that children taught to read in a local dialect first and then transferred to standard Swedish afterwards were more successful at reading Swedish that those children who were introduced to standard Swedish alone from the beginning. Besides the need for first language experiences, Barbara Murphy (1980) argues that teachers should also pay attention to bridging the gap between first and second language. She points

out that one must not assume that students who can understand and handle concepts in their native tongue can necessarily apply them to the second language. She also stresses the need for experiences in the target language which allow students to relate them to their own language and eventually train them to think and manipulate situations in the target language.

Discussions on advanced reading abilities either in the first or second language situations are very difficult to locate (Brumfit 1977). Even at the basic level much remains to be discovered about the cognitive skills which are closely assoicated with linguistic competence. Comprehension models deriving from recent psycholinguistic research offer the beginnings of an account of what is involved in understanding a written text (Massaro 1975, Oden and Massaro 1978). However, the ways these models can be related to the needs of a learner coming from a different cultural and linguistic situation still await exploration. Moreover, very little investigation has been undertaken into the problems confronted by foreign readers using advanced texts.

The Present Survey

This study is directed at finding out the reading abilities of 14-year-olds studying in various schools in Hong Kong using different media of instruction. My intention is to examine the effect of the teaching medium on reading achievement in both English and Chinese. In the Hong Kong situation, this problem is acute since a majority of the Chinese students are learning through English -- the weaker language. This exerts great pressure on the student population, since they have to cope with the language medium as well as the subject under study, and in many cases it is felt that weakness in one seems likely to result in weakness in the other. Moreover, teachers have to teach through their second language and sometimes explanations and descriptions might not be as clear and precise as they would like them to be. In addition to the above problems, there are schools which insist on using English as the only language permissible in class (except for lessons in Chinese and other languages). At the other extreme, some of these so-called English-medium schools limit the use of English to textbooks and school work, and all explanations and instructions are given in Cantonese.

Another type of secondary school in Hong Kong is the Chinese-medium type. The majority of them operate in Cantonese while a few use what is purported to be Mandarin. Students studying in these schools encounter English only in their English classes.

In addition to the above two types of school, there also exist, at the other end of the scale, a few English-medium schools for native speakers of English.

In view of the variety of practices adopted by the different schools in Hong Kong, it is very difficult to measure the precise effect of the medium used in relation to the language proficiency of the child. However, gross categories can be characterized by

referring to the kind of language used in class and text-books selected. Form Two students (14-year-olds) were chosen since this group, at the time of testing, had already been exposed to their respective teaching medium for at least 18 months and,if these different practices have any effect at all on the development of the first and second language of the child, it seems a reasonable assumption that it should be reflected in their ability to read.

The Reading Passages

The English passages used in this study are those developed by the International Association for the Evaluation of Educational Achievement (IEA) to measure the reading ability of 14-year-olds (Population II). The following sub-skills were sampled:

A) *Vocabulary Skills* - This includes the ability to decode unfamiliar words from context. To do this, the student may use:

1) configuration cues - recognizing the word by its overall shape,

2) context cues - getting the meaning from the sentence in which the word is used, and

3) word structure cues - splitting the word into its component parts and working out the meaning from the structural or referential meaning of the components.

B) *Comprehension Skills*

1) At a more primitive level of understanding, comprehension means noting what is overtly stated in the passage. This requires the students to derive what might be called 'literal' meaning from sequential words and their grammatical relations to each other in the sentences and paragraphs. This is what Edgar Dale described as 'reading the lines'. When designing a test at this level, some reinterpretation of the text by the testee is probably unavoidable, if we are to be really confident that meaningful comprehension has, in fact, taken place. Thus not only must the testee decode the ideas presented and answer questions which are specifically answered in the passage, but he will have to do so without parroting the author's words.

2) At a slightly higher level of understanding, the testee is asked to go beyond the literal meaning of a passage. This requires the students to 'read between the lines' and determine the writer's purpose, intent and point of view. This will often involve the ability to distinguish between facts and opinion. This level of comprehension also requires the reader to recognize and interpret many literary devices such as metaphor and irony.

3) A still higher level of comprehension again requires the reader to go beyond the printed words and draw generalizations and inferences not overtly stated by the author. On this level, the reader may have to rearrange the author's ideas into new patterns and extend the implications of these ideas to experiences not stated in the passage.

The above skills as described are taken to constitute the core skills involved in advanced reading and hence constitute the skills sampled in this survey.

In addition to the English tests, a translated Chinese version was also used to provide a common basis for the comparison of results. However, it was also noted that translated pieces would raise serious questions in regard to the difficulty level between the finished products and the original passages. Despite this, translated texts and items were used because it appears to be about the only means of bringing the two languages on to comparable ground. Moreover, there is some evidence that difficulties related to translation were not as severe as one might anticipate. In the UNESCO study completed in 1962 (Foshay 1962), a short reading test translated into the national language of each participating country was included and it was found that the difficulty of the test was quite uniform for thirteen-year-olds across twelve countries. In fact, the reading test showed less variation from country to country than either an arithmetic or a non-verbal intelligence test. Furthermore, even the relative difficulty of specific passages and specific items was highly stable from country to country and language to language.

In the IEA survey upon which this study is based (Thorndike 1973), all the passages and items chosen were translated into the different local languages of the fifteen countries. Over 50 different passages and 445 items were piloted and among the items retained were those which were found to be highly discriminating in two or more languages.

The Translation

As the validity of the Chinese version of the tests hinges on the degree of perceived naturalness of the Chinese passages, extra care was taken in the translation of the stories and items. Several measures were taken to ensure that the resulting work was as close to actual Chinese writing as possible. First the stories and other items were translated by experienced translators, and then the results were worked on by a teacher of Chinese who corrected and modified the expressions used. Lastly the passages were presented to a professional writer, who functioned exclusively in Chinese, for comments and improvement. At each stage, he would put his expertise into shaping the passages into acceptable and appropriate texts. During the translating and improving processes, however, some problems were observed. Many of these problems hinged, not very surprisingly, on the differences between the two languages. At the level of vocabulary, some of the

proper nouns were difficult to translate. When coining new lexis for new objects, animate or inanimate, there seems to be a greater tendency in Chinese than in English to give unfamiliar things some kind of semantic clue in the form of an overt class marker or super-ordinate term. For example, in Section C, the first passage, the meaning of *auk* could only be gathered by the fact that it had feathers, hence the item asking for the meaning of *auk* (item number 2) was not as explicit as it seemed. But on the other hand, the Chinese translation of *auk* is 海雀 (meaning 'sea bird') and the name itself gave away the answer directly. With respect to this item, half of the students were given a phonetic equivalent of the English word so that the task of getting the clue from context would remain the same as in the English case, and the other half of the students were given a new word 鶇 with the radical 'bird' incorporated. In this way the task of the latter group of students would be more akin to the task which would be involved when they came up with a new word in reading Chinese: they would have to guess at the meaning of the word by analysing the character in question as well as by using context. Another major problem experienced with the translation was with the different semantic mapping of the words in the two languages. Thus words like *authority, induction, deduction* and *major premises* in the third passage of Section C did not seem right when translated, partly because terms like *induction* and *deduction* were foreign, and partly because speakers of Chinese seldom seem to go into explicit discussions of this kind. Cultural interests and values did affect translation and consequently reading achievement.

The structural organization of the passages also posed problems as regards translation. In many cases, compound and complex sentences were found to be difficult to translate. The discourse markers which relate a number of concepts to the main idea in different hierarchical orders were difficult to reproduce in Chinese. For example, in the second paragraph of the first passage in Section C, no matter how one translated the paragraph, it was difficult to delineate the temporal order of the actions together with the descriptions or comments of its various objects or phenomena.

The creation of a general atmosphere for the texts was also difficult, for the selection of details in the English passages might not be appropriate if one wanted to give the same impression to a Chinese reader. Problems like this would undermine the validity of several test items.

Sampling

The focus of this survey was on the effect of the medium of instruction on reading achievement and therefore presupposed sampling by school types. The following four types of schools were indentified:

Type 1 Medium of instruction - English except in non-English language classes (e.g. French or Spanish).

Students - Native speakers of English.

Textbooks - English except in non-English subjects (e.g. French, Spanish etc.)

Type 2 *Medium of instruction* - English except in non-English (e.g. French, Chinese etc.) language classes.

Students - Native speakers of Chinese

Textbooks - English except in non-English subjects (e.g. French, Chinese etc.)

Type 3 *Medium of instruction* - English in non-English (e.g. Chinese) language classes. But teachers were encouraged to explain in Chinese whenever necessary[2].

Students - Native speakers of Chinese.

Textbooks - English except in non-English subjects (e.g. Chinese).

Type 4 *Medium of instruction* - Chinese except in English language classes.

Students - Native speakers of Chinese.

Textbooks - Chinese except for the subject of 'English language'.

One co-educational school was randomly selected from each of the above school types for the tests. In the case of the type 1 school, only the English version of the test was administered whereas in the case of the type 2, 3 and 4 schools, students were given both English and Chinese versions of the tests.

The age limit, as reported above, was 14 years. All students who were between 13 years 11 months and 14 years 11 months at the time of the test were chosen because it coincided with one of the IEA populations, and it was noted that this population gave a wider dispersion of abilities than students at the higher levels. Furthermore, at the time the tests were administered this group of students had all been exposed to their respective school types for nearly 2 academic years. If the teaching medium does affect the reading ability of students, significant differences may be detected at this stage.

Another factor considered along with sampling was the intelligence factor. It seemed very likely, and indeed might be the case, in Hong Kong that the brighter students all clamour to get into Anglo-Chinese schools leaving the less able students to the Chinese-medium schools. However, this was a factor very difficult to off-set, for all that could be obtained was a composite secondary school entrance grade and even here there is the problem that the native speakers of English do not take the examination. In the light of all the complexities, one gross common factor was used as a

basis, and that was: all the students were of the same age and had been studying for at least eight years prior to taking the tests. For each school type, the sample was divided in the same way over the three achievement bands[3]. Over 600 students took the tests and 237 of them were included in the samples selected.

Administration of the Tests

Before the tests were administered, the teachers-in-charge were introduced to the IEA reading comprehension study so that they would prepare the students for what was going to happen. Before the actual test began, students were again informed about the purpose of the tests and the importance of treating the whole exercise seriously.

Each class was then divided randomly into two groups of about the same size so as to form two equivalent samples. This was achieved sometimes by taking alternate rows and sometimes alternate columns of students. Care was also taken to see that there were about the same number of boys and girls in each group. One group of students was then given the Chinese tests and another the English tests (with the type 1 school, only one test form was used and hence all the students in the class did the same test).

Analysis and Discussion of Results

a) Reliability of the Tests

The reliability of the total test score in both English and Chinese versions of the tests is presented in Tables 1 and 2. The estimate used the Kuder-Richardson formula 20 (Kuder-Richardson 1973). This is based on the internal consistency of all the items in the test.

School Type	ANOVA Reliability Coefficient
Type 1	0.843
Type 2	0.866
Type 3	0.795
Type 4	0.541

Table 1. Reliability of reading comprehension tests -- English version

School Type	ANOVA Reliability Coefficient
Type 1	----
Type 2	0.893
Type 3	0.846
Type 4	0.627

Table 2. Reliability of reading comprehension tests -- Chinese version

The values in both tables indicate that the tests are providing a reasonable and precise estimate of reading performance in most school types. All coefficients fall between 0.79 and 0.9 except those from school type 4 which is 0.541 for its English test. One major factor appears to have been that the tests were too difficult for students studying in school type 4. An examination of item statistics with reference to school type 4 showed that only 33 percent of the items were correctly answered and over 25 percent of the items were non-discriminating[4] (Tables 3 and 4). As for the Chinese test, the average score of students belonging to school type 4 is higher than the average score from their English test[5], yet the reliability coefficient is still far below the general average (see Table 2). The result indicates that the translated version of the tests is highly reliable when applied to Anglo-Chinese students (school types 2 and 3) but less reliable when applied to Chinese middle school students (school type 4). Thorndike (1973), in his IEA Reading Comprehension project, also found that the test in its translated form was highly reliable with European countries but less reliable with India and Iran.

Discrimination Index	School Type			
	Type 1	Type 2	Type 3	Type 4
.70 - .79	0	2	0	0
.60 - .69	3	6	4	0
.50 - .59	7	5	3	6
.40 - .49	17	10	11	2
.30 - .39	9	7	9	6
.20 - .29	7	10	12	14
.10 - .19	5	8	4	9
.00 - .09	4	3	5	10
Negative	-	1	4	5

Table 3 Discrimination indices (point bi serial correlation) by school type for reading comprehension tests -- English version

Discrimination Index	School Type			
	Type 1	Type 2	Type 3	Type 4
.70 - .79	-	0	2	0
.60 - .69	-	3	5	0
.50 - .59	-	10	6	4
.40 - .49	-	13	8	8
.30 - .39	-	17	8	12
.20 - .29	-	4	10	7
.10 - .19	-	4	8	7
.00 - .09	-	0	4	9
Negative	-	1	1	5

Table 4 Discrimination indices (point bi serial correlation) by school type for reading comprehension tests -- Chinese version

Thorndike suggested three reasons for this. Firstly, the test even in its translated form, was too difficult for students from India and Iran. Secondly, these students might lack familiarity with multiple-choice tests and, thirdly, the passages were contributed by European countries or the United States and they might have presented more difficulty to students from a non-European background. With the present survey, the first two reasons do not apply as the mean scores of both groups are roughly the same. In fact, results from the Chinese middle school group were higher (see Table 8), and all Hong Kong students are familiar with the multiple-choice format. One of the major differences between the Anglo-Chinese school students and the Chinese middle school students in this study is one of familiarity with English in the school environment. This seems to be a major factor contributing to the differences in the reliability indices of the tests. Erratic performance in this case may be due to the fact that, although the tests were carefully translated to resemble original Chinese texts, yet the approach, the selection of details, the development of the arguments and the points of emphasis in writing descriptions are inherently western. Hence students more accustomed to a western approach behaved consistently and students unfamiliar with English writing behaved more erratically, when confronted with the translated texts and items. This explanation may also contribute to the Indian and Iranian cases as the two languages give the impression of being more unlike English than the other European languages included in the IEA study[6]. If the above speculations are true, then it seems that one cannot confidently say that a comparison of reading achievement based on translated pieces compares only the reading achievement ability of students in their first language.

% Correct	School Type			
	Type 1	Type 2	Type 3	Type 4
90 - 99	7	4	5	1
80 - 89	13	4	4	1
70 - 79	10	5	11	1
60 - 69	10	6	7	1
50 - 59	4	8	12	4
40 - 49	6	9	7	5
30 - 39	1	8	1	11
20 - 29	0	7	4	14
10 - 19	1	1	1	14
0 - 9	0	0	0	1

Table 5 Distribution of item difficulty indices by school type for reading comprehension tests -- English version

% Correction	School Type			
	Type 1	Type 2	Type 3	Type 4
90 - 99	-	5	5	9
80 - 89	-	8	6	7
70 - 79	-	8	9	11
60 - 69	-	7	7	8
50 - 59	-	12	10	3
40 - 49	-	6	9	6
30 - 39	-	3	2	3
20 - 29	-	1	1	2
10 - 19	-	1	3	1
0 - 9	-	1	0	1

Table 6 Distribution of item difficulty indices by school type for reading comprehension tests -- Chinese version

School Type	Mean	SD	SE
Type 1	39.73	6.65	2.632
Type 2	28.03	8.35	3.055
Type 3	31.63	6.79	3.076
Type 4	17.15	4.56	3.087

Table 7 Mean, standard deviation and standard error for reading comprehension tests -- English version

School Type	Mean	SD	SE
Type 1	----	----	----
Type 2	32.47	8.99	2.943
Type 3	31.81	7.71	3.021
Type 4	34.03	4.76	2.906

Table 8 Mean, standard deviation and standard error for reading comprehension tests -- Chinese version

Test scores obtained may also reflect the structural differences[7] between languages. This distance may either involve superficial differences like differences in style or the order of presentation of facts, or more-deep rooted differences lodged in the processes of decoding and encoding the different language groups. This is an area which needs to be explored by more cross-national surveys since most of the studies completed are on European and related languages (e.g. Foshay 1962; Kumbaraci 1966).

b) *Item Difficulty and Discrimination*

The statistical properties of the English and Chinese reading tests can be obtained from Tables 3, 4, 5 and 6. Tables 5 and 6 show the distribution of item difficulties expressed as a percentage in getting the items right for each school type. Tables 3 and 4 show the item discrimination indices expressed as point-biserial correlations between items and total scores on the test[8].

The test items show a wide spread of difficulty. The mean percentage of correct answers is 56 for the English test and 63 for the Chinese test. These means are similar to the original IEA study for population II (Thorndike 1973:25). However, for school type 4, the English test was clearly too difficult and over half of the items had a percentage of correct responses below what might have been expected by chance for four-choice multiple items.

In general, the items discriminate satisfactorily between good and poor readers (Tables 3 and 4). Over half of the items are above 0.3 for all the schools except school type 4. This drop in respect to the English test can be attributed to the fact that the test was too difficult for the type 4 students, and the drop as regards the Chinese test may be attributed to the fact that the passages were originally English and pretesting had not been done on the Chinese version. Comparing results with those obtained from the IEA survey, the discriminating power of schools type 1, 2 and 3 is similar to those found in English-speaking countries. This shows that the tests developed work equally well with English and Anglo-Chinese school students in Hong Kong and to a lesser degree with students studying in Chinese-medium schools.

Turning to the behaviour of individual items , item 1 of Section C was clearly too easy (the average facility level of this item is over 95%). It does not descriminate between students at all. This item is perhaps best considered as being strategically placed for motivation purposes.

With item 2 of Section C - the vocabulary item testing for the meaning of *auk* -- half of the group doing the Chinese test had 喔屋 which is a phonetic translation of the word *auk* and the other half had 鷽 which is a single-character word coined for the purpose[9] combining the radical 'bird' and the sound [ok] . The facility index for students in the first group is low (37 percent), whereas the facility level for students in the second group is much higher (63 percent), almost double that of the first group. Although the word 喔屋 is an actual word in Chinese denoting the sound produced by a cock, yet students find the 'nonsense' word 鷽 easier to decode. This indicates that Chinese readers do take semantic cues embedded in the character into account while working out the meaning of the word[10]. Turning to the behaviour of this item in the English test, the school type 4 students found the English item difficult; the facility index is 16 percent (much lower than the level of chance for a 4-choice multiple item). The facility index for the Anglo-Chinese group is about 50 percent and that from the English speaking group is 70 percent. As the meaning of the target word can only be obtained from context (it is not very likely that a 14-year-old would know this vocabulary item before reading the text), this result indicates that Anglo-Chinese students, and perhaps the English ones too, were more ready to make use of contextual clues than the Chinese students in arriving at the possible meaning of new words in the English test.

Among items which did not work and were found to be difficult were items number 5 and 26 in section C. Item 5 related to whether the selection of certain descriptive details successfully denoted certain

abstract qualities in the dogs described. However, what constitutes a 'brave' or a 'strong' or a 'savage' dog is difficult to delineate from language group to language group and indeed from person to person[11]. Results obtained suggest that the English-speaking students were able to abstract from the description and arrive at the correct answer, whereas all the Chinese students found the item difficult irrespective of the language medium of the test.

Item 26 in Section C required students to identify the attitude of the writer: whether he was trying to amuse, inform, persuade or to describe. Results show that Chinese-speaking students found the task difficult[12], whereas the English-speaking students had less difficulty.

This may be due to Chinese students being more involved with the specific details of the passage and hence overlooking its general purpose. English students seemed to be more capable of detaching themselves from the facts and taking an overall view. Perhaps this is an indication of the difference between the reading abilities between second-language speakers and native speakers. Second-language speakers are less secure in the language and have to spend more time on discrete points whereas native speakers can vary the distance of focus at will.

Question 15 in Section D was concerned with whether or not the writer successfully conveyed the impression that the character in the text felt intimidated, impressed, uninterested or astounded when confronted by the ancient city Fez. The Chinese-speaking students found this item difficult. For one thing the translated version of the words may have different semantic mapping in the two languages, on top of which the experience described in the passage may have different connotations to students in the two language groups. For school type 4 this item carried a negative discrimination index which means students who got this item correct were mostly the ones who did badly in the test[13]. This item, in its translated form, is therefore rather unreliable.

c) *Analysis of Reading Comprehension Mean Scores*

The mean scores, standard deviation and standard error are tabulated in Tables 7 and 8. The mean scores of the English test reflect a variety of reading achievements. T-tests done on all possible pairs (Tables 9 and 10) demonstrate that the results for each school type are significantly different from one another, except for school types 2 and 3, which are significantly different from types 1 and 4 but not from one another. In other words, the English test results put the schools back into their original categories: English schools (type 1), Anglo-Chinese schools (types 2 and 3) and Chinese schools (type 4). Using English, or English and Chinese as media of instruction does not significantly distinguish the ability of the students as regards reading English. When the actual scores of school types 2 and 3 are examined, we find type 3 students actually perform better than type 2 students

and have a smaller standard deviation for their scores than the type 2 scores (Table 7). This may suggest the incorporation of the native tongue as a teaching tool actually helps reading ability in English (v. Swain 1981 for a review of research which reaches similiar conclusions); however the level of significance for this to be true is only 0.9.

School Type	t Value	2-Tail Prob.
Type 1 with Type 2	6.29	0.001
Type 1 with Type 3	5.12	0.001
Type 1 with Type 4	16.73	0.001
Type 2 with Type 3	-1.64	0.106
Type 2 with Type 4	6.73	0.001
Type 3 with Type 4	10.05	0.001

Table 9 t-test values between the results of pairs of school types in reading comprehension tests -- English version

School Type	t Value	2-Tail Prob.
Type 2 with Type 3	0.93	0.354
Type 2 with Type 4	-0.21	0.834
Type 3 with Type 4	-0.26	0.213

Table 10 t-test values between the results of pairs of school types in reading comprehension tests -- Chinese version

The mean scores of the Chinese tests by different school types do not differ from one another significantly (refer to Tables 8 and 10). Perhaps by the age of 14, reading ability in the native tongue has already been stabilized and the effect of medium of instruction cannot be overtly demonstrated. After all, the advantage of exposure gained in the school environment can easily be offset by the vast amount of language usage both in the receptive and productive modes outside school.

Conclusion

The main objective of this project was to discover the effect of the medium of instruction on reading ability. Results from the English test show that there are significant differences in the reading scores obtained from the three different types of schools (English schools, Anglo-Chinese schools and Chinese Middle schools), but how far this is a direct reflection of the medium of instruction is difficult to ascertain. It is true that, along with the medium chosen, one particular school would be different from another in the choice of text-books, and in the set of values imposed upon the children. Besides the above attributes, which would be directly associated with the medium used, there are also individual characteristics which the school has developed for itself over the years, and all these factors help to make one school distinct from one another. To attempt to control all these variables for the examination of the effect of the medium of instruction on reading ability is impossible and at present we must, I think, be content with the observation that the main difference between Anglo-Chinese schools and Chinese Middle schools is one of medium of instruction and text-books used, and students studying in these two types of schools have significantly different abilities in reading English. However, the results obtained for reading Chinese are not significantly different.

Within the Anglo-Chinese school type, the sample was further divided into (a) schools using English as a medium of instruction all the time (school type 2) and (b) schools using English as a medium of instruction most of the time, with teachers being encouraged to use the native language when communication in English became difficult (school type 3). There was no difference in the reading ability of the students between these two types of schools if we take the 1 percent level of significance. But when one examines the actual scores, the latter type of school (school type 3) had in fact higher mean scores than the former type (school type 2) ($p = 0.1$). Working on the assumption that the more often one uses English to teach, the better the students will be with their English, then the type 2 schools should

be better than the type 3 schools and not the other way round. The fact that the mean scores were higher with the type 3 students may point to the importance of using the native tongue to teach at times when communication becomes difficult. It is true that the use of translation as a way of teaching has its drawbacks: it is usually too quick and takes away time that could have been used to expose the learners to English, and often there are not exact equivalents of English words in the mother tongue. However, as for the latter drawback, the same problem would apply even when we use English to convey meaning. Using and referring to experiences in the mother tongue may be one of the many useful techniques the teacher needs to employ when he wants to refer to the meaning of a word quickly, in order to get on with the main subject. Avoiding the use of the mother tongue when it seems preferable is often seen by the learners as an implicit criticism of the mother tongue, making it seem like a second-rate language (Nation 1978).

The other main finding is that there is evidence to suggest that Chinese and English speakers decoded new words and linguistic structures differently. The difference in response to the different graphemic translations of the word *auk* point to the fact that the Chinese did take the component radicals of the word into consideration when attempting to arrive at the meaning of the unknown word. Moreover, they were less ready to use environmental clues than native speakers of English.

NOTES

1. An earlier version of this paper was given, in Chinese, at the 'International Symposium on Psychological Aspects of the Chinese Language', organised by the University of Hong Kong Department of Psychology and the Hong Kong Psychological Society, at the University of Hong Kong, June 11-12, 1981.

2. As was recommended in the Hong Kong Government Education White Paper 1973, each school was required to make its own decision on whether it would use English or Chinese as the medium of instruction.

3. Achievement bands were bands awarded to students according to their level of performance in the Secondary School Entrance Examination.

4. As shown in Table 5, the point bi-serial correlation for over 25% of the items of school type 4 is below 0.09.

5. As can be seen in Tables 7 and 8, the mean score of the Chinese tests is actually double that of the English tests for this group of students.

6. The extent to which mental flexibility enters into performance is very difficult to specify. The above variable mentioned is only one variable in the midst of a great number of variables such as differences in the educational system, social habits of language use etc., which serve to separate the achievement of the different groups of students.

7. It is very difficult to specify the nature of the distance in this particular case. The distance may be one of habitual use of language in a specific language group or it may be a mode of thinking which belongs to the wider culture, or again it may be either/both the social, psychological or/and linguistic distance which separates the different languages.

8. For an item by item breakdown of item difficulty and discrimination or details of the tests, please write to the author, at the Language Centre, University of Hong Kong.

9. This word does not exist in Chinese but is a possible word conforming to all the rules of word formation of Chinese characters.

10. Both characters have a common radical 屋 which means 'house' when occurring in isolation and read as [ok] which is very near to the sound of the word *ɑuk* . In addition, the onomatopeic word

噬 had the radical 口 on its left which means 'mouth', to denote that this word stands for the sound of some animals and the newly coined word 雚 had the radical 雀 on its left, which means 'bird' in isolation.

11. For item 5 the facility levels of the English test for school types 1, 2, 3 and 4 are 47%, 10%, 11% and 26% and the facility levels of the Chinese test for school types 2, 3 and 4 are 12%, 11% and 3% respectively.

12. For item 26 the facility level of the English test for school types 1, 2, 3 and 4 are 72%, 23%, 23% and 29% and the facility level of the Chinese test for school types 2, 3 and 4 are 25%, 25% and 6%.

13. For item 15 the facility level of the English test for school types 1, 2, 3 and 4 are 82%, 45%, 57% and 18% and the facility level of the Chinese test for school types 2, 3 and 4 are 75%, 47% and 78%. The discrimination indices of the English test for school types 1, 2, 3 and 4 are 0.218, 0.276, 0.446 and -0.235 and the discrimination indices of the Chinese test for school types 2, 3 and 4 are 0.367, 0.363 and 0.003.

REFERENCES

Brumfit, C.J. 1977. 'The teaching of advanced reading skills in foreign languages, with particular reference to English as a foreign language,' *Language Teaching and Linguistics: Abstracts* 10(2): 73-84.

Foshay, A.W. (ed.) 1962. *Educational Achievements of Thirteen-Year-Olds in Twelve Countries.* Hamburg: UNESCO Institute for Education.

Inner London Education Authority. 1969. *Literacy Survey: Summary of Interim Results of the Study of Pupils' Reading Standards.* London: ILEA

Kuder, G.F. and Richardson, M.W. 1973. 'The theory of the estimation of test reliability,' *Psychometrika* 2 : 151-160.

Kumbaraci, T.E. 1966. *Translated Reading Tests as Culture-Fair Measures for Foreign Students.* Columbia University doctoral dissertation.

Massaro, D., (ed.) 1975. *Understanding Language: An Information-Processing Analysis of Speech Perception, Reading, and Psycholinguistics.* New York: Academic Press.

Modiano, N. 1968. 'National or mother language in beginning reading: a comparative study.' *Research in the Teaching of English* 2: 32-43.

Morris, J.M. 1966. *Standards and Progress in Reading.* Slough: National Foundation for Educational Research.

Murphy, B. 1980. 'Second language reading and testing in bilingual education,' *TESOL Quarterly* 14(2): 189-98.

Nation, I.S.P. 1978. 'Translation and the teaching of meaning: Some techniques, '*English Language Teaching Journal* 32(3): 171-4.

Oden, G. and Massaro, D. 1978. 'Integration of featural information in speech perception,' *Psychological Review* 85 (3): 172-191.

Osterberg, T. 1961. *Bilingualism and the First School Language -- An Educational Problem Illustrated by Results from a Swedish Dialect Area.* Umeå: Vasterbottens Tryckeri AB.

Swain, M. 1981. 'Bilingual education for majority and minority language children.' Plenary paper presented at the 6th AILA World Congress, Lund, Sweden, August 9-14, 1981 (to be published in *Studia Linguistica*.)

Thorndike, R.L. 1973. *Reading Comprehension Education in Fifteen Countries*. New York: John Wiley & Sons.

VALIDATING A COURSE IN READING FOR ACADEMIC PURPOSES

Lee Yick Pang
Language Centre
University of Hong Kong

Introduction

The Language Centre of the University of Hong Kong offers a number of supplementary English courses to first-year undergraduate students, particularly in the Arts Faculty. One major problem confronted by those teaching the courses is that students, while fairly well acquainted with English, having been taught it for over ten years at pre-university level, are in many cases handicapped in the actual use of the language in an English-medium university like the University of Hong Kong. Many of them can best be described as having, to use Allen and Widdowson's (1974) term, 'dormant competence'. However, the problem may become much more serious and psychologically loaded because some students believe, quite sincerely, that they have had enough instruction in the English language, and that no more help is either necessary or useful. Running supplementary English courses in such a context can result in serious motivational problems if students fail to see the relevance of the instruction given to what they perceive to be their actual language needs. This paper describes the course in Academic Reading Skills (ARES) currently being offered by the Centre to first-year Arts students. The rationale behind the course and its main design features are outlined. After this, the course evaluation procedure, involving both objective measurements and student (subjective) feedback, are discussed. The primary aim here is to illustrate a pragmatic, though not unreasoned, solution to a practical problem. Little claim is made in respect of general E.A.P. course design theory. However, it is hoped that this paper will throw some light on the domain of teaching advanced reading, which, as pointed out by Brumfit (1978), is still in many ways unexplored territory.

Rationale

It has become almost standard practice, at least in theoretical papers dealing with ESP/EAP course design, to begin with an empirically derived identification of student language needs (e.g. Cowie and Heaton 1977, Mackay and Mountford 1977, Holden 1978). However, identification of student needs in great detail can prove to be a very costly exercise, and, as pointed out in Lee (1982), students needs can be viewed not just as needs imposed by the language-use context, but just as validly as the needs that the students involved actually perceive as important. The two do not always coincide. Furthermore, the larger and more varied the target student population, the smaller would be the common denominator of needs and the more varied would be the list of specific needs. Since the Academic Reading Skills Course is aimed at first-year undergraduate Arts

students following various combinations of about a dozen academic subjects, adequate empirical profiling of student language needs would seem prohibitive. Therefore, a more abstract starting point has been adopted.

First of all, following Brumfit (1978), academic reading is taken to involve a set of study skills, which can be considered as discrete, but which are variably combined to perform specific tasks of academic reading. Certainly, reading tasks differ depending on the reader's purpose(s) (e.g. revising for an examination, or writing an academic essay). The structure of the ARES course is such that it focuses on each of the skills as the teaching points. However, the skills are continually recycled according to the various reading tasks the students have to perform.

The starting point of the design of the course is, therefore, a 'reading skill' rather than 'student needs' identification. This was done rather informally and intuitively for the first version of the course. No claim whatsoever is made about the list of skills chosen being exhaustive. However, it is hoped that, through careful monitoring of student feed-back on the course (see Section 4), the set of skills chosen will become more and more relevant and comprehensive with subsequent repetition of the course. The set of reading skills sampled in the course include:

1 Anticipation strategies.

2 Deducing the meaning of technical and unfamiliar vocabulary from discourse structure.

3 Organisation of reading notes from a graphic display e.g. a diagram.

4 Use of references to extend reading range.

5 Use of a subject index to identify relevant information for the purpose of constructing a coherent view of academic concepts (this being a realistic and academically relevant version of 'scanning').

6 Identification of the structure of arguments.

7 Identification of author bias/perspective.

8 Reading the same material from different perspectives.

My approach relies on the use of gross divisions into whole 'tasks' and is therefore distinctly different from the more common attempt by those designing reading courses to divide up the process of reading into a constellation of prime components. This latter approach may be useful for beginning and/or intermediate readers, but seems less likely to be useful for mature and advanced ones, since they will have probably integrated the prime components to a far greater degree than the beginners. In fact, a previous attempt by the author to devise a course using the 'fragmented' approach met with considerable student resentment.

Design Features

There are several general design features of the course:

a) Each unit of the course follows a general instructional sequence of presentation, group exercise, individual practice, and project. In the first (presentation) phase, the teaching point is illustrated to the students. In the second (group exercise) phase, the whole class is split into small homogeneous groups for further work on the task presented. Then in the third and final (individual practice) phase, students are given the opportunity to try the task out on their own.

b) Whole articles are used in all stages of the course. The reading materials are of course treated at different levels of delicacy and hopefully in a gradual progression from general to more detailed. The reason for using whole articles is two fold. Firstly, it preserves the authenticity of the text being read and the task being undertaken. Then, secondly, it gives students the opportunity to do (c) below.

c) Although the task is prescribed, students can choose their own reading materials, particularly in the individual practice phase. This, it is hoped, will be beneficial to student motivation, in that it may enable them to build up a collection of useful references during the course, which will be of use to them on their actual degree courses.

d) After the relevant skills have been mastered by the students, the last (Reading Projects) section begins. This is intended to provide students with an opportunity to have general and real life practice in academic reading and to merge the work of the reading course with their departmental work. Students have more or less complete freedom to choose both the task and the texts which they will use for their subjects, which can be either group or individual. The role of the tutor by this point is simply to assist and to evaluate the maturity of the project work.

Course Evaluation

In the course evaluation procedure an attempt was made to answer the following questions:

1 How can students genuinely needing the course be selected?

2 Have the design features of the course had the desired effect or do they need alteration? And, if they do need to be changed, how radically should the course be altered?

3 How can student improvement be effectively monitored?

4 To what extent does objectively measured improvement relate to students' subjective perception of the course being either interesting and/or useful?

To answer Questions 1 and 3, a test of academic reading proficiency was designed for inclusion in the English language proficiency battery administered annually by the Language Centre to all incoming first-year Arts students. The reading test is designed to be a reasonably direct test of the skills sampled in the ARES described above. The same test was then used as a post-course test. For this reason, the test was intentionally made fairly difficult, so that it could be sensitive to student differences both pre- and post-instruction. To answer Question 3, a feedback questionnaire was given to students at the end of the course. It requested a unit by unit evaluation of (a) the skills presented and (b) the various design features of the course. For each unit, the questionnaire tried to elicit students' subjective ratings on the two dimensions of 'interest' and 'usefulness'. A 5-point Likert-type scale was used (1: *not at all*, 2: *a little*, 3: *not sure*, 4: *quite*, 5: *very much*). Indices of 'interest' and 'usefulness' were derived for every unit by totalling the ratings.

A cut-off point denoting positive 'interest' or 'usefulness' was put at the second highest point of the scale (point 4) and from the frequency distributions of the various items it was found that 65.8 percent of the students (n = 60) thought that the skills sampled in the course were 'useful' and 45.5 percent found them 'interesting'. Furthermore, 52.3 percent found the various design features of the course 'useful' and 38.4 percent found them 'interesting'.

The various course efficiency indices were analysed using multiple regression. From this analysis, it was found that 'usefulness' was rated slightly higher than 'interest' in the overall evaluation of the course (usefulness: 55.5; interest 44.5).

To answer Question 4 in the course evaluation procedure (To what extent does objectively measured improvement relate to students' subjective perception of the course being either-interesting and/or useful?) an objective 'reading gain index' was derived by subtracting the pre-course reading test score from the post-course score (the two tests being identical). This objective reading gain score was then compared with the various items in the student feedback questionnaire.

However, before going into the details of that analysis, it is important to establish the validity of an apparent reading gain. For this purpose, t-tests were performed on all the questions and on the totals of both the pre- and postcourse tests. The results are given in Table 1 below:

	obs. t	sig.
Subtest 1	-2.83	0.008
Subtest 2	-2.76	0.009
Subtest 3	-5.39	0.0001
Subtest 4	0.42	0.676
Total	-7.71	0.0001

Table 1: T-tests of Pre- and Post-course Test Scores

It is clear that students made a significant gain in all but the last subtest as well as in the total score. It can therefore be stated with a reasonable degree of confidence that the reading gain index here does represent an improvement in the reading ability of the students.

The comparison of objective vs. subjective rating of improvement was done by a series of one-way analyses of variance with the reading gain index as criterion and the student feedback questionnaire items on overall course efficiency as variable factors. The results on two of the items are quite interesting. Student reading gain was significantly related to the questionnaire item on the 'overall relevance of the skills taught to students' departmental academic work' ($F = 3.176$, sig. 0.039). Reading gain was also related to the item on the overall usefulness of the instruction procedure adopted (presentation, group exercise, individual exercise) ($F = 2.9$, sig. = 0.08). Even though this last figure failed to reach the conventional 0.05 significance cut-off point, the 0.08 level is near enough to the 0.05 limit to indicate a trend. However, the detailed results of this last item will not be considered.

The actual breakdown of the reading gain index by the item on overall course relevance to departmental academic work is given below in Table 2:

Questionnaire Rating	Mean	SD
1 Not at all relevant	5.67	2.52
2 Not too relevant	1.33	3.20
3 Not sure	2.00	1.00
4 Quite relevant	2.62	2.06
5 Very relevant	6.00	1.00

Table 2: Mean and Standard Deviation Breakdown of the Student Feedback Questionnaire Item on Overall Course Relevance.

It is clear from Table 2 that there is, in fact, a progression in the mean reading gain from students rating the overall relevance of the course as *not too relevant* (2) to *very relevant* (5). From this progression it can be said that student progress in the reading skills taught was in all probability at least partly conditioned by their perception of the course being more or less relevant to their academic work. However, those students who rated the course as *not at all relevant* had the second highest mean reading gain index. The precise reason for this is a little difficult to determine. However, motivation could be a likely candidate.

It is indeed difficult to say whether the attempt to answer the last question in the course evaluation procedure was successful.

However, it was nevertheless an attempt to relate progress in a course of learning programme to student subjective perception of the relative efficiency of the course. Little has been done in terms of specific research so far in this area, and it is hoped that more consistent effort in this direction will eventually lead to more satisfying results.

Conclusion

The primary aim of the present paper is to illustrate one solution to the practical problem of designing a reading course for advanced second/foreign language users of English. However, it is also an attempt to highlight the importance, and perhaps the necessity, of providing objective and principled justification for any course of action taken in language course design. Language course design has developed to a stage when the uncritical following of any single presently existing theory, to the exclusion of all others, would lead to untenable and frustrating classroom experiences.

REFERENCES

Allen, J.P.B. and Widdowson, H.G. 1974. *English in Focus: English in Physical Science*. Oxford: Oxford University Press.

Brumfit, C.J. 1978. 'The teaching of advanced reading skills in foreign languages, with particular reference to English as a foreign language.' In *Language Teaching and Linguistics: Surveys*, edited by V. Kinsella. Cambridge: Cambridge University Press.

Cowie, A.P. and Heaton, J.B. (eds.) 1977. *English for Academic Purposes*. Reading: Association for Applied Linguistics.

Holden, S. (ed.) 1977. *English for Specific Purposes*. London: Modern English Publications.

Lee, Y.P. 1982. *An Analytical and Empirical Study of the Concept of Language Proficiency, and its Consequences for the Development of an English Language Proficiency Test Battery* M.Phil. thesis. University of Hong Kong (submitted).

Mackay, R. and Mountford, A.J. (eds.) 1977. *English for Specific Purposes*. London: Longman.

THE 'PROJECT ON ENGLISH LANGUAGE PROFICIENCY TESTING': AN OUTLINE.

Graham Low and Lee Yick Pang
Language Centre
University of Hong Kong

Introduction

The last decade has witnessed a growing interest in the design of language-use tests, as the effects of 'communicative' approaches to teaching second and foreign languages have carried over into language testing. On the one hand, this has led to a growing interest in developing links between formative testing and student self-assessment (see Rea 1981 and the summaries in Oskarsson 1978). On the other hand, it has led to a widespread interest in what Clark (1975) has called 'direct tests', as these offer at least a partial solution to the seemingly impossible problem inherent in the approach to testing advocated by Lado (1961): that of ever deriving a truly representative sample of 'items' for any given linguistic area or skill. Directness is a complex parameter, since the idea of a direct test is that it attempts to simulate all the relevant aspects of specific types of interaction or language-related activities. A direct test will therefore test how someone puts his or her linguistic knowledge and skills to use in realistic situations. For this reason, direct tests often attempt to encompass the creative and negotiative processes involved in interpreting and evaluating language contexts.

We have pointed out elsewhere (Low 1981) that, despite widespread interest in direct language-use tests, very little in the way of proper test development and evaluation procedures has been published. Carroll (1980) despairs of ever finding appropriate statistical techniques and Morrow (1979) appears to reject the use of statistics entirely (though he fails to discuss the point in anything approaching adequate detail). In Low (forthcoming) it is argued that an ad hoc approach to language proficiency testing is both undesirable and unnecessary. Only if test construction is principled is it possible to be explicit about what precisely is being tested and hence about whether, and to what degree, it is being tested in a direct way. In such a case, departures from complete directness can be introduced in a controlled way, and the term 'controlled reductions in directness' was introduced to describe them. Once a test designer can describe the shape of a test explicitly, he or she is in a position to produce hypotheses about the way the test scores should look. If a set of hypotheses about the results of a test can be produced, it becomes much more feasible to develop suitable validation techniques. Thus the development of appropriate methods of evaluating direct tests presupposes a principled approach to test design and construction.

Designing a direct language-use test is not a simple task, and indeed the job becomes harder the more we discover about them. One solution to the problem might be to find an indirect test which

predicted the desired language behaviour to much the same degree as a direct test. The loss in our knowledge about why such predictions were possible (since an indirect test might bear no clear resemblance to the task being tested) would be more than offset by the increased ease of construction, and possibly administration too. Thus, so-called 'integrative' tests like cloze and dictation might be preferable to extended, complex direct tests, if they could be shown to have equivalent predictive power in the relevant situations, assuming that detailed diagnostic information was not required.

At a more abstract level, it was suggested in Low (1981) that it would be of some use to be able to construct a language testing theory which placed highly direct tests in a central position, as the most desirable and effective in an ideal world. The test designer would then have to discover, for any given activity , what a high degree of directness actually involved, then decide where, why and how far he or she wanted a particular test to be indirect. An important part of presenting a test to possible testers (or even testees) would be a statement about the conclusions that were reached in this respect. However, before any such theory can become more than simply wishful thinking, a considerable amount of data is required about the predictive value of more direct versus less direct tests.

It was to examine just these two questions, particularly with regard to reading and writing, that the Project on English Proficiency Testing (henceforth PEPTEST) was set up.

The PEPTEST *database*

One problem which regularly afflicts language proficiency testing projects is that it is difficult, sometimes even impossible, to construct a sufficiently comprehensive database. This is a serious problem in the case of projects involving tests of differing degrees of directness since, ideally, hundreds of tests, each varying in one small aspect would be needed in order to establish that the variation was indeed one of directness and not something else. In practice, one is limited to what test results are available plus the results of any supplementary tests which the testees concerned can be persuaded to sit for. In both cases, there are clear limitations to what can be obtained. Firstly, language proficiency tests are unlikely to be taken seriously by testees unless the outcome is perceived as being significant for them personally. This means that for research purposes 'real' tests are likely to be preferable to tests administered for no apparent reason. Secondly, people in general do not appear to enjoy extended language testing, particularly where writing is involved. So, not only are testees likely to take 'supplementary' tests less seriously than 'official' tests, but they may well refuse to undertake any but the shortest of extra tests at all.

Bearing these two considerations in mind, we decided to base the project around a group of first-year Geography students at the University of Hong Kong. The project would follow their linguistic (and to a lesser degree academic) progress during their first year,

with the possibility of extending the monitoring to their second and final years, if the initial results were sufficiently encouraging.

There were two reasons for selecting Geography students. First, our direct test of writing was slightly biassed towards students of Human Geography. Since Human Geography (under the title *Man and the Environment*) regularly attracts the majority of first year Geography students, restricting the sample to students on the *Man and the Environment* course would give us both a well-defined and a reasonably large group of testees. Secondly, few students choosing Geography normally exercise their right to change courses during the first two weeks of Term 1, so the sample tested at entry would remain fairly stable. In fact we ended up with 81 students.

A large amount of information was already available. Results of the Advanced Level Use-of English test set by the University (passing the 'U.E.' is now a condition of entry) were available, with a question-by-question breakdown. The Language Centre standardly administers a battery of language-use tests (oral skills, listening I, listening II, reading, academic writing, cloze) to all new first-year undergraduates in September. In 1981 this so-called 'LAS' battery was to be accompanied by a questionnaire asking students for details of their socio-enconomic and linguistic background and for aspects of their present linguistic behaviour. It also asked them to self-rate their own proficiency in English and examined certain of their attitudes towards Chinese and English. The LAS tests are also readministered to a number of the same students in December. For the last two years, the subtests, and particularly the writing subtest, have been accompanied by validation questionnaires, so that face validity could to some extent be divided into two dimensions: predictiveness (in September) and sampling representativeness, in the light of having first-hand experience of the relevant task, (in December). Finally, discussions with the Department of Geography and Geology led to agreement that students' end-of-year exam, coursework and tutorial performance grades be released to us.

It was, however, clear that some supplementary tests would be needed. The short study by Lee and Low (1981) had explored the possibility of using clines of directness, where a series of tests was orderable in terms of increasing or decreasing directness, and we wished to make use of the same procedure here. Lee and Low (1981) used the LAS cloze as the least direct of the series, with the result that they had three compound [1] tests (ie. involving more than one subtest) and one simple test (ie. with only one subtest). This time it was thought desirable to have as many as possible of the tests which related to reading and writing as compound tests. To this end, a supplementary test involving two cloze passages was designed. One passage was more subject-related than the other, so that between-passage comparisons could also be made.

Next, a writing test that was less direct than the LAS tutorial paper simulation was needed (henceforth TPT). A second supplementary

test was therefore designed, this time involving a short test of editing as one question, and a test of writing about the development over time of some ideas on town planning as the other. The questions were on different topics so that there could be no line-of-development through the test, as there is with the TPT (details of the TPT can be found in Low, forthcoming).

Lastly, we wished to examine the effect of a different, more direct, format on Question 1 of the TPT, and so a third supplementary test was prepared. It was felt that such an exercise was feasible in the case of this particular question, since, as Question 1, it did not relate crucially to one's having answered the other questions. Questions 2 to 5, however, do require the testee to refer back to earlier parts of the test.

Since much of the main body of the project involves prediction studies, the choice of criterion or independent variables is of considerable importance. While it is fairly common to use end-of-year grades for this purpose - indeed we ourselves are engaged in a small pilot study which does just this, based on the sample used in Lee and Low (1981) (Lee and Low, forthcoming) - we felt that end-of-year grades might not, in fact, be the most appropriate figures to use. One reason for our misgivings was the fact that Geography Department grades are based on a student's apparent understanding of the material; the examiners proceed (quite rightly) as if a language handicap does not exist. A second reason is that the grades are in effect composite scores of unknown dimensions, which can make the interpretation of correlation or regression studies extremely difficult. In an attempt to overcome these problems, we designed, and asked the Geography staff teaching the course to evaluate, improve and hence validate, a diagnostic proforma to be filled in by them for each piece of written work completed by each student during the year. This means, for *Man and the Environment*, six written assignments between September and the summer exams. The proforma asks for information on both discourse and grammar features, plus degree of adherence to departmental norms as regards copying and quoting. In all, staff are asked to make 27 assessments beyond giving a final mark. The use of a series of diagnostic proformas has the further advantage that some sort of criterion-referenced developmental profile of each student's academic writing ability can be constructed.

As a result of the production of this original proforma, two developments occurred which were not originally planned, but which can easily be incorporated into the project. Firstly, the Department of Geography asked us to design an equivalent proforma to help assess students' tutorial performance. The resulting proforma asks for 22 decisions on a variety of activities and analytic levels, ranging from grammatical abilities, to logical expression, self-representation and aspects of small group dynamics (eg. use of helping strategies). As tutorial performance is assessed only once, at the end of the academic year, it will not be possible to build a longitudinal profile for oral skills in the same way as for writing skills.

The second development was that the Department decided to adopt a diagnostic proforma system throughout all three years. The PEPTEST proformas are being used on a trial basis for a year, and will be modified to suit departmental needs more closely over the summer vacation 1982. This means that by the time the PEPTEST sample enter their second and third years, there will be a readily available instrument yielding reliable diagnostic data waiting to be applied. Preliminary discussions are underway in an attempt to have this data released to us, so that the PEPTEST project can be extended (retaining our original sample of 81) across all three years.

This, then, constitutes the PEPTEST database. A summary version can be found in Table 1.

June 1981	*H.K.U. Advanced Level Use of English battery*				
	Reading Speed.	Reading Compreh.	Précis	Essay.	Listening. Total.
Sept	*H.K.U. Language Centre 'LAS' battery*				
	Reading acad. texts.	Listening to Lecture.	Listen to small gp. discussion.	Writing tutorial paper.	Oral tutorial perform-ance. Cloze.
	H.K.U. Supplementary material				
	Background + habit questionnaire	Validation questionnaire for TPT (students + for 1980, staff)	Outline validation questionnaires (other tests)		
Oct	*Supplementary tests*				
	ST1 Alternative TPT Quest. 1.	ST2 'No storyline' writing test.	ST3 Geog. cloze. Neutral cloze.		
Nov-June	Written assignment proformas (6 per student) (compatible with TPT).				
Dec	*H.K.U. Language Centre LAS retests* (on highest 30% according to initial LAS test). Writing test (TPT) accompanied by follow-up validation questionnaire.				
June 1982	First year exam marks.	Coursework Assessment.	Tutorial performance mark + proforma.		

Table 1: Summary of the PEPTEST database

Test score tailoring

Before the scores are used for research purposes, a number of validity and reliability measures will be made, in order to establish that no remarking or removal of subtests/items from the database is necessary. This outline of the project is not the place to describe them in detail, but several have been described in earlier work which preceded the PEPTEST project: for example the use of specfic variance estimation to examine the degree of overlap of subtests, (Low, forthcoming), the use of principle component analysis and marker score techniques to establish whether subtests may be validly grouped together as separate tests (Lee and Low, 1981), or the modified item analysis technique described in Lee (1982).

We intend at the same time to carry out two other operations, firstly indexing the cloze tests roughly for independent vs. grammatical meaning, as this might help to explain some of the score patterns obtained, and secondly, weighting and possibly recomputing some of the total scores (Lee 1981). This weighting operation will however be of marginal use, since we are primarily interested in the scores for individual questions.

Initial questions and hypotheses.

Hypothesis 1: Tests which are more direct are better predictors of desired linguistic behaviour than less direct ones.

We are in a position to approach this hypothesis from a number of directions. By this we mean that we can set up a number of clines of directness and examine the degree to which they predict desired linguistic behaviour. The main studies will be concerned with two clines (or rather sets of clines, since we can produce 'families' of clines in both cases); that of testing the ability to write a tutorial paper or essay, allowing for periods of reflection by the writer, and that of testing the ability to write shorter essay-type questions within a time limit, as when writing examination answers. In both cases, Supplementary Test 3 (ST3), which consists of the 2 cloze passages,can be considered as either one complex test or as two simple tests, differing a little in terms of topic relevance. Alternatively, we may be in a position to put together a new complex cloze test, by combining the scores of the LAS cloze test with those of the less relevant cloze on the ST3 test. At the heart of these studies is the expectation that the précis and short essay test on the Advanced Level Use of English Test will predict exam essay scores better than the TPT and that the TPT will predict tutorial paper grades better than the above U.E. tests.

On the question of criterion, as opposed to dependent, variables. the simplest solution would be to use end-of year grades. For the reasons explained above, however, these may be less than ideal. We shall therefore attempt to relate the test scores to the writing proformas. Since these represent, in effect, a developmental profile of the students concerned, predictions can be made for each of the 6

(cross-sectional) profiles. This should result in much more useful information being uncovered.

At a less detailed level, it should be possible to construct similar clines of directness for reading and oral performance in tutorials. The main problem with the reading cline is that there is no obvious way of obtaining a good criterion variable and the major problem with the oral skills cline is that there are not really enough tests. This is primarily because the project was not originally set up to examine oral skills.

Question 1: What can one most usefully reference language proficiency tests against?

The ability to compare the effects of referencing a variety of test types, all completed within a reasonably short time of each other, against a number of different criteria, should allow us to come up with some interesting answers to this perennial problem.

Question 2: What do standard essay tests measure?

The answers to the Question 1 should also permit us to make a few suggestions about what precisely standard essay tests actually measure. The conclusions will be tentative since we are only examining one such test (the short essay on the Use-of-English Test) and the Hong Kong Examinations Authority does not publish reliability statistics for it.

Question 3: Do the Language Centre diagnostic tests actually predict the language problems that first-year Geography students have?

The project will attempt to include some validation work on the LAS battery for internal use. We would like to be able to compare the results from the LAS subtests with the profiles on the two sets of proformas. Of particular interest will be the extent to which the 'goodness of fit' between test scores and profiles varies during the year. Indeed, the information from the proformas alone will provide much useful information which will help in the design of future diagnostic tests and even in the validation and design of English language teaching programmes designed specifically for Geography students.

Table 2. Possible clines of directness plus criteria

LEAST DIRECT ——— MOST DIRECT					CRITERION
1. *Writing a tutorial paper.*	LAS cloze	ST3 neutral, cloze	ST3 Geog. cloze	UE précis/ +/or UE essay	ST2 'no-story-line' test	TPT	1. Coursework grades 2. Written work proforma profiles.
2 *Writing an exam essay.*	varoius combinations of cloze passages as above.		TPT	UE précis	UE essay		1. Yr.1 exam grades.
3 *Reading textbooks*	cloze combinations		UE Reading Speed	UE Reading Comprehension	LAS academic textbook reading simulation test.		1. Reading sections of written work proformas. 2. Reading sections of tutorial assessment proformas.
4 *Performing in tutorials*	cloze combinations		LAS listening to small group discussion test.		LAS oral performance in simulated tutorial test.		1. Tutorial performance grades/ profiles.

Question 4: What test or cluster of tests best predicts students' language problems?

This is clearly related to Question 3, except that we hope to make use of clustering techniques to obtain the answers.

These then are the major objectives of the PEPTEST project. It will be noted that there is only one stated hypothesis so far. In practice Hypothesis 1 is a generalisation, since a number of less general hypotheses are contained within it, and will be made explicit when the results are written up.

Expanding the Project

As the Project was being set up, it became apparent that the data colected in the database could be used as the basis of a number of other lines of research, connected particularly with aspects of second language acquisition and approaches to second language teaching. As the Centre had been involved in looking at the way students monitor spoken (Morrison and Low, forthcoming) and written output, we selected this particular area as the direction in which the project would be initially expanded. The question we are interested in is the following:

Question 5: Do the oral monitoring and text editing strategies used by students correlate with each other and with characteristics relating to second language proficiency?

This is clearly a complicated area and any work here will inevitably be rather unsophisticated. However, we hope to try and develop ways of examining the question, which can be developed at some later date.

Initially, we hope to make use of a mini-computer, and to present texts and particular tasks on the screen. A record will be kept of sequence and time taken, where relevant. As a first step towards achieving this, we are designing a number of formats where text editing tasks can be performed but which involve fairly simple programming.

We anticipate that the results and techniques involved in this more exploratory section of the project will be presented separately from the main body.

Conclusion

The intentions of the PEPTEST project are therefore:

1) to make some contribution to the theory of direct testing, which will complement, particularly in the areas of writing and (to a lesser extent) reading, work currently being done on direct oral testing (eg. Clark 1978 and Hinofotis, Bailey and Stern 1981).

2) to examine how tests of English for Academic Purposes might usefully be referenced and validated.

3) to develop techniques by which text editing strategies may be examined, so that text editing may be related to oral monitoring strategies, and both may be studied in relation to test results.

4) to contribute to the language testing programme run internally by the Language Centre.

The project is scheduled to run until December 1982.

NOTES

1. There seems to be no generally accepted way of describing tests with one,as against more than one, subtest. The terms 'simple' and 'compound' have been coined here as they seem intuitively to express the right ideas. We prefer to talk of a 'compound test' rather than a 'complex test', since it it not necessarily complexity that is at issue, and 'compound'seems to refer more naturally to something composed of two or more clear elements. Our terminology thus parallels the distinction between complex and compound tones in intonation as described by Crystal (1969).

REFERENCES

Carroll, B.J. 1980. *Testing Communicative Performance*. Oxford: Pergamon.

Clark, J.L.D. 1975. 'Theoretical and technical considerations in oral proficiency testing.' In *Testing Language Proficiency*, edited by R.L. Jones and B. Spolsky. Arlington: Centre for Applied Linguistics.

————,(ed.) 1978. *Direct Testing of Speaking Proficiency: Theory and Practice*. Princeton: Educational Testing Service.

Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Hinofotis, F.B., Bailey, K.M. and Stern, S.L. 1981. 'Assessing the oral proficiency of prospective foreign teaching assistants: instrument development .' In *The Construct Validation of Tests of Communicative Competence*, edited by A. Palmer, and P.J.M. Groot and G.A. Trosper. Washington, D.C.: TESOL.

Lado, R. 1961. *Language Testing*. London: Longman

Lee, Y.P. 1981. 'Some notes on internal consistency reliability estimation for tests of language use.' University of Hong Kong, Language Centre, *Working Papers in Linguistics and Language Teaching* 4:19-27.

————.1982. *An Analytic and Empirical Study of the Concept of Language Proficiency, and its Consequences for the Development of an English Language Proficiency Test Battery*. University of Hong Kong: M. Phil. dissertation (submitted).

Lee Y.P. and Low, G.D. 1981. 'Classifying tests of language use.' Paper presented at the 6th AILA World Congress, Lund, Sweden, 9-14 August, 1981. To be published in *Journal of Applied Language Studies* under the title 'Towards a classification of tests of language use: The dimension of test directness.'

————————————.Forthcoming. 'Are direct communicative tests really worth the effort?' Paper to be presented at the TESOL International 16th Convention, Waikiki, Hawaii, May 1-6, 1982.

Low, G.D. 1981. ' Communicative testing as an optimistic activity.' University of Hong Kong, Language Centre, *Working Papers in Linguistics and Language Teaching* 4:1-18.

————.Forthcoming. 'The direct testing of academic writing in a second language.' *System*.

Morrison, D. and Low, G.D. Forthcoming. 'Monitoring and the second language learner.' In *Communicative Competence*, edited by J. Richards and R. Schmidt. London: Longman, Applied Linguistics Series.

Morrow, K.E. 1977. *Techniques of Evaluation for a Notional Syllabus.* Reading: Centre for Applied Language Studies, University of Reading (Study commissioned by the Royal Society of Arts).

Oskarsson, M. 1978. *Approaches to Self-Assessment in Foreign Language Learning.* Strasbourg: Council of Europe, Council for Cultural Co-operation.

Rea, P.M. In press. 'Formative assessment of student performance: the role of self-appraisal,' *Indian Journal of Applied Linguistics.*

COMMUNICATE IN WRITING. A FUNCTIONAL APPROACH TO WRITING THROUGH READING COMPREHENSION.

Keith Johnson. 125pp., plus Teacher's Book, 35pp.
London: Longman. 1981.

Reviewed by Evelyn Cheung.
Language Centre
University of Hong Kong

Introduction

Any textbook which claims to *teach* writing deserves careful consideration, as there is a dearth of good material which can fill the gap between writing courses which focus on the sentence (with heavy emphasis on grammatical points), and ones which concentrate on the end-products that writing often involves, like reports and letters. Materials are desperately needed for this middle ground, which to quote Candlin and Breen (1979: 183), would 'serve the *process* (italics mine) of teaching and learning rather than the *product*...' Teaching materials must therefore be able to develop the learner's underlying abilities and knowledge and act as a link between what he knows at the time of teaching and the target that he and the teacher think he should reach.

The aim of the present article is to examine whether *Communicate in Writing* (henceforth *Communicate*) could be considered as belonging to this middle ground, and secondly, whether it succeeds in being what it claims to be, "a course in advanced writing skills".

I shall be reviewing *Communicate* with these questions in mind:

(1) Does it actually teach advanced writing skills?

(2) What are the methods used to achieve these writing skills?

(3) Who are the target group?

In the first part of the review, I shall give a brief description of the aims in each of the Parts (following Johnson's division of the course into Parts I, II and III) as well as examine areas of interest related to the three questions mentioned above. In the conclusion, I shall look at the course as a whole and deal with points that affect the overall organization.

How COMMUNICATE *is organized*

Johnson has divided (rather crudely as he admits in the Teacher's Notes) the course into the following:

(1) Describing things and ideas

(2) Describing processes and events

(3) Developing an argument

Under the three main headings, the following functions are found:

I Describing things and ideas

(a) Referring back and giving new information (Unit 2)

(b) Defining (Unit 3)

(c) Classifying (Unit 4)

(d) Comparing and Contrasting (Unit 6)

(e) Giving examples (Unit 7)

(f) Making descriptive statements (Units 8 and 9)

II Describing processes and events

(a) Expressing purpose and means (Unit 11)

(b) Expressing prediction and expectancy (Unit 12)

(c) Expressing cause, effect, reason and result (Unit 13)

(d) Describing a sequence of events (Unit 14)

III Developing an argument

(a) Expressing degrees of certainty (Unit 16)

(b) Supporting an argument (Unit 17)

(c) Expressing reservations (Unit 18)

(d) Drawing conclusions (Unit 19)

Johnson stresses 'flexibility' as an important methodological point (Teacher's Notes p.4) in the course; he gives fairly precise notes on which different *exercises* which might be suitable for intermediate and advanced students, as well as indicating which would suit 'occasional' students. The notes also suggest how the exercises could be used on both long term and short term courses. This suggests then, that the exercises are 'modular' in nature and teachers can choose, to a great extent, the types of exercises that they wish to cover or leave out. It is of some relevance, therefore, to ask if this 'flexibility' extends to the *units* too. Johnson advises teachers not to think of *Communicate* as a book "to begin at the beginning and work through to the end". One is left wondering then, whether one has the choice of doing 'Giving examples' (Unit 7) and then working backwards to 'Defining' (Unit 3) for instance. Unfortunately, there is no explicit mention of this. In addition the cover of the book states that "the book takes the student through the main processes of learning how to write - description of things, description of events, development of an argument". Could one then assume that the order in which the headings appear, suggests that one should start with units within *Description of things and ideas* first, or that it doesn't really matter? One wishes the writer had made the organization of the units a little clearer.

The Units mentioned above are known as Main Units, and after three or four such units, there is a Consolidation Unit which revises skills and language items taught in the immediately preceding three or four units. Each Main Unit is again divided into three parts.

Part I

Each unit invariably starts with a reading passage which contains the language function(s) listed above. This seems to follow the usual conventional format of starting with a reading passage, except that instead of random comprehension questions, the reader is directed to look for 'main' points, sundry organizational pointers, and certain vocabulary items. Johnson has also identified what he calls 'techniques' such as note-taking, summarizing and recognizing main ideas, as some of the important components of the reading and writing process. These are not taught, so much as used as a framework for understanding the reading passage. *Communicate* however appears to have associated notetaking and summarizing too closely with the reading passage, so that the overall impression (to the reviewer at any rate) is that they are simply comprehension exercises or isolated writing tasks that have little to do with the main business of *learning* how to write. Towards the latter part of the course more challenging activities involving the student's skill in recognizing what to take down when he has to write a specific type of essay could usefully have been devised. This feeling was echoed by a few first year students, who pointed out to me that they did not see the purpose of a note-taking exercise (they had gone through Unit 20) which merely asked them to complete notes already half-formulated by the writer. Likewise with the summarizing exercises. The technique adopted by *Communicate* consists of filling in blanks or expanding notes obtained from a previous note-taking exercise. A more worthwhile objective might have been to have asked how and when one does in fact summarize and how the resulting types of summaries might differ, rather than set summarizing exercises seemingly for their own sake. If one reflects on the occasions when one needs to summarize, it can be seen that purpose determines to a great extent whether all the 'main' points of the passage/article are relevant, or only one 'main' point and a single illustration. Generally, students are told to look for 'main' points and to ignore details in a summary, but this may not necessarily be justifiable in every case, and of course 'main' is a subjective and therefore relative term.

What I like about *Communicate* is the way in which students are given practice in the organization of ideas, with questions about cohesion and coherence being introduced in a natural way. This is especially so in exercises called *Adding Information*. This type of exercise brings students to grips with the linking of ideas both within a paragraph and between paragraphs, and to consider alternative places where the sentences and ideas could occur. As a result, they learn to make decisions as to how the sentences could suitably link with preceding and following ideas. This sort of choice has the advantage that it forces them to think about overall effect and to decide which of several effects is preferable in a particular case. This seems to me to be a natural activity in the actual writing process.

To improve this aspect of the course even further, I think there should be more explicit teaching of various cohesive devices for linking ideas, other than simply testing the student's knowledge of the common items such as *but, however* or *while*. Perhaps this could be achieved in the manner of the related unit *Referring back and giving*

new information (p.6) or *Relative Clauses* (p.49) where fuller explanations are given and where the student can see for himself how such linguistic knowledge can assume a rhetorical function in writing. In this connection, some attention has also been given to ways of emphasizing a point or topic (in Unit 16 and 17), something which is often overlooked by textbooks which concentrate on teaching linguistic structures. More exercises of this nature would be of considerable help to the advanced student, who may need to highlight certain points or who may wish to indicate to the reader that there is a shift in emphasis. Quite often, such a student is asked to write on a controversial topic, which involves more than just presenting different views and results (as in Unit 20, p. 117). To this end, he often has to be taught how certain effects, like adopting an ironical or critical tone, can be achieved through organization of ideas as well as through the use of language structures and vocabulary. This ability is fairly important to a number of our own first-year students, as they have to write critiques on recommended articles, or critical appraisals of literary works and theoretical models.

The final feature regarding Part I concerns the way in which formality is taught. It is disturbing to note that formality in academic writing has been reduced to substitution exercises called 'Vocabulary Extension'. One questions the wisdom of such a method, as it unnecessarily confuses the student as to what constitutes a 'formal' or 'academic' tone. For instance, the student is told that *These distinguished themselves in that ....* is more formal and academic than *One important difference between these two* (Unit 4, p. 16, line 15). Secondly, students may be led falsely to the belief that if they sound pompous enough, their writing may become academic. Prize examples include *the field of artificial flight* instead of *the field of aviation; doomed to failure* instead of *would fail* and *aeroplanes would never have any utility* instead of *aeroplanes would never be useful.* After these random word and phrase substitutions in Part I, students are confronted in Part II with longer stretches of informal language, which they have to rewrite into formal and academic language. At the Language Centre of Hong Kong University, teachers of Writing find that they cannot hope to talk about formality without at least 4 hours of teaching and without explaining what a formal and academic approach actually means.

Part II

According to Johnson, Part II deals with language functions. These have been listed in the first part of the review. Language structures associated with the respective functions are taught at the sentence level, followed by controlled paragraph writing, which may or may not have model answers. The model answers have certainly been useful in giving the teacher some clues as to what the tasks demand. Although they are good examples which trained teachers can further exploit, nevertheless it might have been a good idea if model answers had been accompainied by explanations and teaching points. This would certainly have helped the Teacher's Book to live up to its name.

Basically, Johnson has selected useful functions which are likely to be important in a wide range of writing situations. I question, though, his decision to teach *Making Descriptive Statements* at the level that they are presented in Units 8 and 9. I am not sure why two units are given to this language function, while *Expressing cause, effect, reason and result* are dealt with in only one unit. I have always thought that describing objects or making factual statements is what *elementary* courses rather than 'advanced writing skills' courses are all about. If intermediate and advanced students have problems in describing, the problem is more likely to be the lack of a good range of vocabulary or organizational skills. Further development of *Expressing cause, effect, reason and result*, on the other hand, might bring in more practice with cohesive devices like discourse markers and introduce a wider selection than *thus, so* and *consequently* which is all that occurs in the unit. Intermediate and advanced students do not, in my experience, often need to revise language items like *because, this was caused by*, or *the result was*, so much as to be taught that they need to signal to the reader the relationships between sentences/ideas through a variety of discourse markers.

Part II ends with a mandatory exercise called *Writing about your subject*. The aim is to extend the language function(s) taught earlier to a topic of one's choice. This appears to be a concession to ESP students who, according to Johnson, will then have "an important opportunity to say what *they* want to say - to use, that is, the language for some communicative purpose". Pressure then is put on the student to select the input to the only free writing exercise in the section. At this point, perhaps, teachers should be advised that they need to help students who may have great difficulty in finding a topic. It is not easy for an advanced History student to write about "*how* things are done and *why* they are done" (*Purpose and Means* Unit 11, p.63) or for Literature students to make a table and "describe a process" (*Describing a sequence of events* Unit 14, p. 82).

Part III

Part III contains additional exercises which give further practice on the points taught in Parts I and II of the unit. One interesting feature is that some of the exercises can be slotted in at certain points in either Part I or II for further practice, or as extra material for the faster students. While one is happy that such "a bank of materials" (Teacher's Book, p.4) exists, the exercises are still fairly short writing tasks which rework the 'techniques' and language functions without developing them into more difficult tasks or aiming to expand the linguistic repertoire associated with the various language functions.

The unit ends with a section called *Further reading* which lists a selection of books which are related to the topic of the passage in Part I. The references are printed with the place of publication missing, which is unfortunate in a book that claims to be teaching academic writing.

The latter part of the review consists of a number of points which deal with the overall organization and a brief evaluation of the material.

Concluding remarks

1. One constraint in developing material meant for such a wide net of people concerns the question of length. The emphasis on writing one or two paragraphs in most of the writing tasks has prevented more interesting development of the existing material. In order to cover as much ground as possible to cater for general *and* academic needs, the overall impression is that the teaching points are presented piecemeal, partly I think, because of the decision to make the exercises modular within a single unit. What appears to hold the exercises together is the thematic link, but this link could be rather tenuous in many cases. By further development, I am thinking of tasks for advanced writing skills. The exercises in the course fall roughly under 3 headings: sentence-level drills, controlled paragraph writing and free writing. The controlled paragraph writing could be improved by helping students to handle not only the language function(s) of the particular unit, but to incorporate "rhetorical points" (Teacher's Book, p. 3) or other language functions in the course. Pedagogically it may sometimes be useful to look at language functions in isolation, but in real life advanced students have to think about a number of different things concurrently within the same paragraph. What I am arguing for is to make the controlled paragraph writing more difficult so that it becomes a genuine 'take-off' stage towards free writing tasks.

2. The reading passages seem to be for good intermediate students with perhaps Units like 2, 6, 9, 11 and 14 for students at the lower end. While retaining this close relationship between reading and writing in the course, one wonders if the passages could not have been better exploited by appearing as something other than straight factual accounts. The language functions could still be adequately introduced using contexts like articles, editorials, formal speeches, lectures or letters. The passage could also take the form of a pair of letters, as it does for instance in Unit 17 (*Supporting an argument*), where two writers propose different and conflicting ideas on the same theme. This brings in, in a natural way, points about highlighting an argument, and the different ways of doing so. The passage could also be broken up into shorter passages and located at different points of the unit, so that reading and writing occur rather more naturally than they do now.

3. My third point concerns what the term 'communicate' means in '*Communicate in Writing*'. The communicative feature is mainly found in groupwork (Teacher's Book p. 5) where a pair of students show each other their written work. The success of the task then depends on whether the students understand what their classmates have written. This calls into question the extent to which an

evaluation of a piece of writing should depend on a fellow student's understanding. Firstly, the student has to be confident that it is indeed the writing that stops him from understanding the text, and not his own mental abilities. Secondly, he has to be able to point out where misunderstanding has arisen in the text and thirdly he has to know how to go about repairing the language. In addition, for advanced students, the ability to communicate does not merely consist of presenting clear factual knowledge but involves knowing who one is writing for. This allows the writer to give appropriate background information, and to decide how much detail to give and the tone that he wishes to adopt. In this connection, I would like to mention that the passages used in *Communicate* tend to have a 'teacher/lecturer' tone which may not be an appropriate or a useful guide to the writing student whose work is to be read by his teachers. As an illustration of this 'lecturer tone' the following passage may be cited:

> "Are some people born clever, and others born stupid? Or is intelligence developed by our environment and our experiences? Strangely enough, the answer to both these questions is yes....
>
> Imagine now that we take two identical twins and put them in different environments. We might send one, for example, to a university and the other to a factory where the work is boring. We would soon find differences in intelligence...." (p. 114)

4. I am not sure how helpful the course will be to self-study students since the Teacher's Book is not as "invaluable" as it claims. Not a few exercises in Parts II and III are without model answers. In addition, I wonder how such students will ever know where they went wrong in the free writing exercises (*Writing about your subject*) and whether their alternative answers in Parts I and II are acceptable or not.

5. Does *Communicate* teach writing skills? I feel that it is definitely one of the better textbooks around which actually does teach some useful writing skills, but whether it can be considered an 'advanced writing course' is debatable. The bulk of the material falls between the intermediate level and the lower end of the advanced level. In order to deal adequately with the academic needs of the advanced student, perhaps a Book Two might be necessary.

REFERENCES

Candlin, C. and Breen, M.P. 1979. 'Evaluating and designing language teaching materials.' In (University of Lancaster, Institute for English Language Education) *Practical Papers in English Language Education* 2:172-216.

Johnson K. and Morrow K. 1981. *Communication in the Classroom.* London: Longman.

RESEARCH NOTES

The two short notes which follow were written for submission to Dr. M. Oskarsson, University of Göteborg, Sweden, as reports of work currently being done at the Language Centre, University of Hong Kong, into the uses of student self-assessment. Dr. Oskarsson is preparing a report on research into self-assessment for the Council of Europe. As these two short reports are unlikely to be quoted in full, they are published here as records of work in progress.

REPORT ON THE USE OF STUDENT SELF-ASSESSMENT IN THE TESTING PROGRAMME OF THE LANGUAGE CENTRE, UNIVERSITY OF HONG KONG

Lee Yick Pang
Language Centre
University of Hong Kong

The Language Centre of the University of Hong Kong administers annually a battery of English language proficiency tests to select those first year students in the Faculty of Arts who are felt to be in need of various supplementary English courses offered by the Centre. Due to the fact that these English courses are not academic credit courses, considerable pressure has been exerted on the Language Centre to identify those students *really* in need of supplementary help. Failure to achieve this would in all probability lead to serious motivational problems for students having to follow the courses.

Considerable effort was, therefore, spent examining and resolving test design problems, and, eventually, in the autumn of 1980, a battery of 5 English proficiency tests was designed, each being a resonably direct type of language-use test. The test content was derived from the various supplementary English courses in the Centre.

Given the fact that the battery was conceived as a possible solution to student motivational problems, a serious attempt was made to obtain student feed-back on the various testsin the battery. In fact, this was how we became interested in student self-assessment. Students were asked to fill in 2 self-assessment items on each test immediately after they had finished answering it. The questionnaire tried to find out how students rated their performance in the test and whether they would like to follow courses that taught them the skills which were being tested.

All the questions involved a three-point scale (1: *unsatisfactory*, 2: *satisfactory*, 3: *very good*, for the Test Performance question; and 1: *yes*, 2: *no*, 3: *not sure*, for the 'Do you want to follow a course of instruction?' question). A chi-square analysis was performed to find out whether student responses followed any biased pattern. The results showed that for all the items in question, student responses were significantly different ($p < .001$); and there were no carry-over effects. Since the questionnaire for the writing test was analysed independently by my colleague G.D. Low in great detail, I shall here deal simply with the results of the reading, the listening and the speaking tests.

The crucial questions were whether students could accurately judge their own performance in the various testsand how far student willingness to follow any particular course was related to test score. The answer to the first question would throw light on the accuracy of test; the answer to the second question would indicate possible trends in student motivation. Accuracy is here taken to mean agree-

ment between student self-assessemnt and test scores. It may be interesting to observe that of the three tests in question, reading and listening were objectively marked tests, while the speaking test was subjectively marked.

A one-way ANOVA was performed with the three test scores as dependent and each of the corresponding self-assessment items as independent variables. The results are summarized below in Tables 1 and 2.

Dependent Variable	Obs. F	Significance
Speaking	18.92	0.0001
Listening	11.08	0.0001
Reading	11.30	0.0001

Table 1: One-way ANOVA with self-rated test performance as independent variable.

Dependent Variable	Obs. F	Significance
Speaking	5.53	0.02
Listening	14.42	0.0001
Reading	5.35	0.02

Table 2: One-way ANOVA with self-rated desire to follow course as independent variable.

The results in *Table 1* show that student self-assessment of test performance and actual test scores agreed reasonably well. *Table 2* further shows that students' willingness to follow courses is related to their test performance.

It seems, therefore, that the English language proficiency test battery was able to identify those students in need of the courses offered and that students picked for the courses would want to follow them. Of particular interest is the agreement between student self-rating and test scores. Even though too early to propose any definite answers, it seems to us that direct tests would facilitate testees' rating of their own performance.

REPORT ON THE USE OF STUDENT FEEDBACK IN A TESTING PROGRAMME

Graham Low
Language Centre
University Of Hong Kong

For the past two years I have been interested in the use of student evaluation as an aid to validating a fairly direct test of writing academic English (more specifically, writing a tutorial paper), aimed at first year undergraduates at the University of Hong Kong. The philosophy behind the test, a description of it and some details of its validation can be found in Low (forthcoming).

Interest has centred on the possibility of fragmenting face validity into a number of indices and relationships, and on how the tester can establish whether these are genuinely independent judgements. To gather the data,questionnaires have been used and to establish reliability some check questions were built in. Further estimates of independence of judgement could be gained from statistical comparison between answers to different questions and the relationship of both to test scores.

The results showed firstly that students did feel they were able to make positive judgements about the task validity of the test. Of 365 students, only 11 percent used the *no opinion* box. Secondly, it seems that within certain limits, students were able to judge task validity independently of perception of their own performance on the test. Although a chi-square test gave a chi-sq = 35.97, df = 4, p = 0.003, an examination of the matrix showed that 91 percent of those who thought it was a *good test* felt they had performed less than *well*, and 96 percent of those who thought it was a *very good test* thought they had performed less than *very well*. This accounted for 71 percent of all respondents. Only at the bottom end did both of the students who felt it was a *very bad test* feel they had at the same time performed *poorly* or *very poorly*. Moreover, judgements of task validity appeared to be independent of perception of topic bias in the test (p = 0.98).

While on the subject of perceived bias, it is clearly important to be able to find out not just if bias was perceived, but also whether testees thought that it significantly affected their performance and whether there was any evidence of this effect on the scores themselves. To get at these questions, testees were asked to rate firstly how they thought they had performed on the test and secondly how they would have performed in the absence of topic bias (this having been previously established). The resulting scale, labelled 'significant bias' was unavoidably heavily weighted towards the bottom end of the perceived performance range, which is precisely where one might expect to find the greatest feeling of penalisation. The actual results in this case were that of the 161 students for whom the variable was computed, 75 percent felt that there was indeed a measure of significant bias. However, since testees for the most part thought the test was a 'good'

test, it seems that judgements of bias can be made relatively independently of judgements of task validity and that ultimately we perhaps need to think in terms of three not two categories of perceived bias: (1) Perceived topic bias, (2) Significant bias (as previously defined) and (3) Tolerable bias. A further measure of test validity can be obtained by comparing the perceived bias responses of the group as a whole with the actual test scores. If no significant relation is found, then the tester may use this as a gross indication of test validity. This was in fact true of the writing test (ANOVA Observed $F = 1.528$, $p = 0.211$).

Validation techniques need to be evolved which are appropriate to the purposes underlying the test. The writing test discussed here is intended as a direct test of tutorial paper writing for first-year Arts students and is used to identify students in need of help. It thus has a predictive side which needs to be validated separately from its representativeness (ie. the extent to which it genuinely reflects what testees will have to do at a later date). As an aid to establishing perceived validity as a predictive instrument, a crude 'hostility index' was produced. Testees were asked whether they wanted to undergo a writing course at the Language Centre after taking this exam and could answer *yes*, *not sure* or *no*. Negative answers could be taken as indicative of hostility. This however, even when refined by correlating the results with answers to other questions, remains a fairly crude indicator. A more useful approach is to look at ease of adaptation. One reason why many remedial second language placement tests are so unreliable is that they take no account of the fact that testees may adapt to the task at different and unpredictable rates. We do not wish a placement test to simply pick up students whose only problem is that they adapt slowly to the task. For this reason, testees were asked to rate, on a five-point scale, the amount of difficulty they had had in adapting to the test. This was then compared with the actual scores. In both cases (ease of adaptation index and hostility index) the results were non-significant. While there was some overlap between perceived ease of adaptation and both perceived performance (a slight tendency for those who found the test hard to feel they had performed *poorly*) and perceived topic bias (chi sq = 25.4, df = 4, $p = 0.25$), the students did seem able to split judgements of difficulty from judgements of task validity (chi sq = 23.85, df = 16, $p = 0.09$).

To assess the sampling representativeness of the test, a second questionnaire was given to students when they were retested, using the same test, 3½ mounths later. The results are being processed. The suggestion therefore is that a test-retest situation can be used to assess validity as well as reliability. In this case the test scores become unimportant; the retest serves primarily to refresh the testees' minds about the precise nature of the test so that they can compare it with their feelings about the actual task (as they now have first-hand experience of it).

The role of student evaluation in the design and administration of language-use tests is perhaps best envisaged as simply one of several tools by which to validate a test. Despite Palmer and Bachman's (1980) summary dismissal of 'face validity', the above techniques constitute the only method which allows conspicuous injustice to be picked up at th

same time. This fact alone would seem to argue strongly for its retention. It should be noted too that when using self-assessment as part of test validation we are not particularly bothered about the accuracy with which speakers can rate their own performance. This is important, since despite the encouraging results cited in Rea (in press), it seems generally impossible to control the criteria which learners use to make their judgements and thus there is often a serious comparability problem.

REFERENCES

Low, G.D. Forthcoming. 'The direct testing of academic writing in a second language,' *System*.

Rea, P.M. In press. 'Formative assessment of student performance: the role of self-appraisal,' *Indian Journal of Applied Linguistics*.

Palmer, A.S. and Bachman, L.F. 1980. 'Basic concerns in test validation'. Paper presented at the 15th Regional Seminar, RELC, Singapore, 21-5 April, 1980.

POSTSCRIPT TO WORKING PAPERS, NO. 4

I wish to acknowledge the invaluable assistance of Dr. Jimmy Chan and his colleagues in the Educational Research Establishment of the Hong Kong Department of Education for their assistance in the analysis of the data for 'The use of cloze procedure in the Form III English Language Secondary School Scaling Test' in Working Papers 4:28-47, June 1981.

R.K. Johnson
School of Education
University of Hong Kong.

RELC REGIONAL SEMINAR ON INTERLANGUAGE TRANSFER PROCESSES IN LANGUAGE LEARNING AND COMMUNICATION IN MULTILINGUAL SOCIETIES SINGAPORE, 19-23 APRIL 1982

The SEAMEO Regional Language Centre in Singapore will hold its 17th Regional Seminar from Monday 19 to Friday 23 April 1982 in the RELC Building, 30 Orange Grove Road, Singapore 1025, Republic of Singapore. The theme of the Seminar is "Interlanguage Transfer Processes in Language Learning and Communication in Multilingual Societies" and the objectives are the following:

To explore the complex conditions in which translation/ transfer processes occur in language learning situations and to examine their consequences for language learning;

To develop teaching and learning techniques and procedures using translation/transfer strategies as aids to increase the learner's communicative competence in his/her second language;

To review and investigate aspects and concepts of the scientific study of translation/transfer phenomena and their relevance for language teaching programmes;

To consider the cross-cultural implications of translation/ transfer phenomena in language learning situations.

名古屋大学語文研究所

語言理論教学研究

五