

myriad voices that told us children
life was out there in the dark—
moths in the gardenia tree
singing in the hedges and the grass
and even now
when evening folds its wings upon the rise
I can feel in the great deep darkness outside
movements in the hedges and the grass
moths in the gardenia tree.

Although there is an obvious point of comparison between the poems as regards the poets' insight into the world of nature, the atmospheric associations cannot be further apart, as anyone who has spent only one night in a kampong house in Malaysia would agree. My Malaysian students could certainly appreciate the description of nocturnal noises as 'razor sounds that saw the night' (the cicada, for example, makes a noise not unlike a small razor-sharp saw) but could hardly be expected to identify with a situation at night when it is so quiet that one can hear a clock ticking 'from a distant floor'. Again, whereas the English poet identifies the creatures of the night that visit him within the confines of his own room, the Asian poet leaves the creatures outside for the most part unidentified, mysterious, noisy, even menacing, but still enchanting and objects of wonderment. This attitude would be easier to appreciate by a Malaysian student, than the attitude of the English poet. Notice also that Lee's poem, unlike Hardy's, does not contain any unusual forms of words and structures for the student to grapple with. So it is because, on the one hand, that the theme of the two poems is similar, but on the other hand, it is because the words and their connotations in Lee's poem are more accessible to Malaysian students, that it is suitable material to serve as a point of entry into the more culturally alien world of Hardy's poem.

Now let's turn to the Syllabus recommended by the Curriculum Development Committee of Hong Kong. We find that Literature students are expected to 'support their individual responses' to the works of English Literature 'with relevant reasons'. Of course, the teacher's task of trying to develop in students a personal response to poetry will be made easier if the ideas, situations, and experiences to found in the poems, used initially, are culturally appropriate. That is to say, the cultural context in which the ideas are expressed and the experiences described, should not, at first, be too far removed from their own. Littlewood (1986:175) reminds us that a student's interest in and appreciation of the experiences described in a literary text, will be enhanced if the content makes 'contact with their experience'. In addition, the full appreciation of a poem often depends on an awareness of the significance of a special event in which the poem is rooted. Hong Kong students, for example, could hardly be expected to identify with Wordsworth's delight at seeing 'a host of golden daffodils' if they have never seen daffodils growing in a park or a field.

Turning to the Hong Kong Syllabus again, we find these words; 'Candidates will be expected to show an ability to understand the thought and feeling in the poetry, and the ways in which these are conveyed'. Now in the case of some of the set poems that ability must be very difficult for students to acquire, let alone demonstrate. I say this because I believe that the door to the world of many an English poem is closed to Asian students of English Literature because the words do not have for them the connotations by means of which the poet paints a great deal of the picture and conveys much of the meaning. To illustrate what I mean I'd like to read two more poems to you. One is an English poem taken from the set book for Hong Kong students and the other is an Asian poem written in English.

MY GRANDMOTHER by *Elizabeth Jennings*

She kept an antique shop—or it kept her.
Among apostle spoons and Bristol glass,
The faded silks, the heavy furniture,
She watched her own reflection in the brass
Salvers and silver bowls, as if to prove
Polish was all, there was no need of love.
And I remember now I once refused
To go out with her, since I was afraid.
It was perhaps a wish not to be used
Like antique objects. Though she never said
That she was hurt, I still could feel the guilt
Of that refusal, guessing how she felt.
Later, too frail to keep a shop, she put
All her best things in one long, narrow room.
The place smelt old, of things too long kept shut,
The smell of absences where shadows come
That can't be polished. There was nothing then
To give her own reflection back again.
And when she died I felt no grief at all,
Only the guilt of what I once refused.
I walked into her room among the tall
Sideboard and cupboards—things she never used
But needed: and no finger-marks were there,
Only the new dust falling through the air.

FOR MY OLD AMAH by *Wong Phui Nam*

To most your dying seems distant,
outside the palings of our concern.
Only to you the fact was real
when the flame caught among the final brambles
of your pain. And lying there
in this cubicle, on your trestle
over the old newspapers and spittoon,

your face bears the waste of terror
at the crumbling of your body's walls.
The moth fluttering against the electric bulb,
and on the wall your old photographs,
do not know your going. I do not know
when it has wrenched open the old wounds.
When branches snapped in the dark
you would have had a god among the trees
make us a journey of your going.
Your palm crushed the child's tears from my face.
Now this room will become your going, brutal
in the discarded combs, the biscuit tins
and neat piles of your dresses.

I feel that the world of the Asian poem is more accessible to Asian students than the world of the English poem. They would, for example, be more comfortable with the connotation of words such as Amah, cubicle, trestle, spittoon, electric bulb, moth, old photographs, discarded combs, from the Asian poem, rather than the connotations of words like Apostle spoons, Bristol glass, and brass salvers from the English poem. Also, it is common for Asian children to be brought up by a woman other than their mother, at least for some period of time. Therefore, it should not be difficult for most Asian students to identify with the intensity of feeling felt by the poet on the death of a woman who might have become a second mother or grandmother, rather than the kind of analytical guilt complex of the English poet. Thirdly, it is much more common for old people in Asian countries to spend their last years close to their family rather than being left to grow old and eccentric alone and lonely, as is often the case in England. Here then is another poem in which the overseas student of Literature does not have to grapple with alien cultural connotations, and which could be used in the classroom as a 'way-in' to the English poem set for study.

It has often been pointed out that the comprehension of a written text depends to a large extent on a number of assumptions shared by both reader and author who, as Nuttall (1982:7) reminds us 'assumes that the reader shares attitudes, beliefs and values expected of people who have grown up in a similar society'. Indeed, as we know, reading is not a one-way process for there are clues in the written text left there by the author so that the reader can use them to construct meaning. If however, the reader does not see the clues, then part or all of the meaning may be lost. Moreover, with regard to poetry, the difficulty of understanding the poetic language will be compounded if the subject matter is unfamiliar or the characters described are not easily recognizable to the extent that the students cannot picture in their mind's eye, what is being described. So poems which deal with familiar themes and describe situations to which the students can relate are suitable material to use when attempting to develop the students' appreciation of poems as well as their awareness of the kind of language to be found in

them. To this end then, English Literature teachers in Hong Kong could perhaps include among their anthology of poems to be presented in class a few which have been written by poets from a similar cultural background as their pupils.

Here's another Asian poem written in English.

NEW YEAR'S MORNING

by *Lee Tzu Pheng*

The children come searching
among the scattered red of the road
for unexploded crackers,
turning over the charred heaps
and eagerly picking up one or two;
the night's dew has made them damp,
they no longer sound sharply,
but even a flash and a pop
is an artistic success,
an event of power.

Small faces smile
a celebration
as echoes rock the neighbourhood,
machine-gun into the new year.

The teacher could first anticipate the theme or topic of the poem by eliciting ideas connected to their own experience for it is better at first to focus on the topic and feelings of the poem rather than the vocabulary or structure. For example, before reading this poem to the students, the teacher could ask them to think about, in pairs, three things that people do during Chinese New Year. He could then pool all the ideas from the students trying to depict a similar scene to that which is going to be painted by the poet. Next, the students experience the text of the poem by listening to it being read aloud, in such a way that they at least hear the correct pronunciation and rhythm. This is important because, as P. D. Roberts (1986:15) reminds us, 'rhythm and sound are the heart of the poem'. Moreover, if E.S.L. students are simply allowed to read the poem for themselves the internal recording of the poem in their mind might be faulty and the opportunity to use this Asian poem to lead the students to awareness of how rhythm and sound contribute to meaning may be lost.

In conclusion, here are the additional three reasons why Literature teachers in Hong Kong could profitably exploit Asian poems written in English. First, their students may come to regard poems as accessible and therefore enjoyable. Second, they will be motivated to read other poems and even perhaps try their hand at writing some. Finally, I think they will be better prepared to read and respond to English poetry.

References

- Brumfit, C. J. (ed.) 1983. *Teaching Literature Overseas: Language-based approaches*. Pergamon and the British Council.
- Hewett, R. P. (ed.) 1968. *A Choice of Poets*. Thomas Nelson.
- Littlewood, W. T. 1986. *Literature in the Foreign-Language Course* in C. J. Brumfit and R. A. Carter (eds.) *Literature and Language Teaching* O.U.P.
- Moody, H. L. B. 1968. *Literary Appreciation*. Longman.
- Nuttall, C. 1982. *Teaching Reading Skills in a Foreign Language*. Heinemann.
- O'Malley R. & Thompson D. (eds.) 1957. *Rhyme and Reason*. Collins Educational.
- Roberts, P. D. 1986. *How Poetry Works*. Penguin.
- Thumboo, E. 1976. *The Second Tongue*. Heinemann Educational.
- Williams, R. 1986. 'Top ten Principles for Teaching Reading' in ELTJ 40/1 pp 42-45.

KEEPING STUDENTS IN THE MAINSTREAM: SCHOOL-BASED REMEDIES FOR SLOW READERS OF ENGLISH

Ora W. Y. Kwo

Department of Curriculum Studies

University of Hong Kong

Introduction

The standard of English in Hong Kong secondary schools has long been a public concern, and became more prominent with the extension of compulsory education after 1978. Whereas in the old system many low achievers dropped out at an early stage, now they must remain in school for a longer period of time.

In order to improve standards, in 1981 the Education Department established a post for remedial teaching of English in each government or aided secondary school. The move aimed to bring slower learners up to the general level within each school's ability range.

As a result of the government's intention to promote Chinese as the medium of instruction, schools are faced with a system of 'positive discrimination' in which they can obtain extra teacher resources in order to 'avert the consequential drop in the standard of English due to reduced exposure' (Hong Kong 1986). Schools have been encouraged to use the extra teachers to split classes and therefore to reduce class size.

However, it is questionable whether reduction of class size can by itself achieve the intended purpose. Research, both in Hong Kong and elsewhere, has failed to demonstrate clear correlations between class size and achievements (see e.g. Haddad 1978, Larkin & Keeves 1984, Ho 1985).

The situations of individual schools are too varied for problems to be tackled by a standard intervention policy. Accordingly, school-based identification of problems and solutions therefore becomes a more appropriate strategy.

This paper focuses specifically on student problems in the reading of English as a Second Language (ESL). It presents an alternative framework for teaching, and makes recommendations on school-based operation of the framework.

Problems and Remedies in ESL

One major source of difficulty for many ESL learners arises from the level of proficiency demanded by the system. The revolution brought by the great expansion of knowledge and the efficiency of the media in conveying it has required ordinary people to have much stronger literacy skills than used to be the case. This places new pressures on school pupils. Moving along different

levels of proficiency, they continually confront new problems and demands. It is arguable that attention to problems should not be restricted to the junior levels, and that remedial teaching is needed for slower learners in every form throughout the school.

Johnson' (1986) survey on remedial teaching of ESL in Hong Kong secondary schools highlighted two main issues. First, the survey exposed a lack of remedial English teaching in the higher forms. This matched the Education Department's view that remedial teaching was most needed in lower forms, though it did not match the reality of language demands facing all pupils. Johnson pointed out that in a situation of limited resources, it did not necessarily seem wise to concentrate attention on the students who are least able, least motivated and least likely to benefit.

Johnson's survey also addressed several curricular issues. It showed that very few schools had special remedial teaching programmes. In most schools the course books and examinations were identical for remedial and non-remedial students.

The lack of special course materials and examinations raises questions about the justification for setting up remedial classes in the first place. If the objective is to eliminate differences in proficiency levels and to reintegrate slow learners into ordinary classes, it is difficult to see how this can be achieved without separate programmes with attainable goals for the weaker students.

In addition, further problems arise from labelling. Relegation of slow learners to special classes may have a negative impact on their work and attitude to work due to the label attached to them (Rist 1978).

This perspective suggests that remedial programmes are often self-defeating. Without specially-tailored programmes, remedial learners may have little hope of reintegration to the mainstream. Indeed, rather to the contrary the remedial nature of such classes could even become more permanent, due to the effect of labelling. In this case, the justification for remedial teaching may be challenged. The alternative focus should be to keep students in the mainstream and help them to develop their reading skills, facing new problems and reaching full potential at all levels.

Textbooks and Achievement Levels

The composition of the mainstream varies according to individual school practice. Some schools stream students on the basis of their academic abilities, whereas others go for mixed-ability patterns across classes of particular levels. Yet whatever the definition of the mainstream, two basic facts generally remain consistent. First, students in the same form level vary in their English reading proficiency levels; and second, they all use the same set of textbooks. It should not be assumed that when students are kept in the mainstream it is possible to eliminate differences in student achievement levels. A varied pace of student progress is a natural phenomenon in learning.

One might then ask whether all students must have the same set of textbooks for the same form levels. Ideally, textbooks should match the students' standards, and there should be more than one set of textbooks within any form level. However it is administratively difficult to use different sets of textbooks even when classes are academically streamed. It would require fine placement tests for resetting students, and even when testing instruments are available schools encounter problems of timetabling.

When the same set of textbooks is used across different proficiency levels, a process-oriented pattern of teaching is needed. This technique relates to the issue of match/mismatch between textbooks and students, and emphasises the need to exploit textbook material to develop supplementary exercises to meet the different achievement levels.

Process-oriented Teaching

Recent discussions on the nature of reading have built on the notion that reading is only incidentally visual, and have emphasised that meaning is not fully present in texts waiting to be decoded. Smith (1983) and Carrell (1988), for example, have drawn attention to the ways in which meaning is recovered during the process of reading.

When individuals derive meanings from text, the researchers argue, they use a combination of 'bottom-up' and 'top-down' processes in interaction. Bottom-up processing refers to the way that readers derive linguistic inputs from the text. Top-down processing occurs when readers use prior knowledge to make predictions about the data they will find in a text. Successful reading, it is argued, requires both bottom-up and top-down processes. The 'interactive' approach to reading focuses on the interaction between the bottom-up and top-down processes in order to facilitate effective and efficient comprehension (Silberstein 1987, p. 31).

Process-oriented models of reading extend the range of possibilities for reading instruction in a second language. The ESL reader should be taught to utilize both knowledge-based and text-based processing skills. However if the text is too difficult for the ESL reader, the reading process may become a tedious decoding exercise in which learners cannot simultaneously use their prior knowledge.

Clarke (1980) has demonstrated that linguistic deficiencies may limit the transference of L1 reading skills to reading in second language. Some readers, he points out, may be inefficient bottom-up processors, decoding language only with great effort. This deficiency may lead to over-reliance on decoding, and thus a breakdown in interactive reading. This point draws attention to the importance of the match between the text and the reader, and demonstrates that there is a threshold to pass if ESL readers are to make progress through interactive processing of the textbooks prescribed for their levels. Help in developing decoding skills is vital to enable the learner to step over the threshold to interactive reading (Kwo 1989). Accordingly, to facilitate decoding less difficult material should be given.

This analysis suggests that ESL teachers are faced with two pairs of questions. The first pair is diagnostic:

- which students have reached the threshold level to benefit from the chosen English textbook?
- Which students will find the textbook reading material too difficult to handle through interactive processing?

The second pair of questions relates to teaching:

- For average students whose proficiency levels match the expectations of the textbooks, what reading tasks can be designed to encourage interactive processing of the material?
- For students below the threshold, what help should be given to improve their decoding skills so that interactive processing can take place later in the year?

Answers to these questions require not only the professional knowledge and skills of the ESL teachers, but also the support of the administration. Individual teachers working in isolation, even if they have both the skills and the enthusiasm, soon find that completion of these tasks is formidable and perhaps impossible. One solution lies in the development of a school-based team.

School-Based Development of Testing & Teaching Materials

The goal of school-based work should be the development of three types of package:

- * diagnostic and progress tests with standardized scores;
- * supplementary reading materials and exercises for learners below the threshold; and
- * supplementary exercises for faster learners above the threshold.

Development of these materials should not be a one-off operation. The materials should be put into classroom use, revised, and improved on the basis of practical experience.

Although keeping students in the mainstream is the central objective, the goal does not preclude grouping students within classes. In a mixed-ability class, students should be placed in different groups with different sets of materials and exercises of appropriate levels.

This approach will require students to spend considerable time in individual and group work, as opposed to working as an entire class at a uniform pace. A possible pattern of grouping and operating could be:

	<i>Group A</i>	<i>Group B</i>
Level:	Far below threshold	Just approaching threshold
Materials:	Self-accessible supplementary material	Material from the main textbook, with supplementary, easier reading exercises

	<i>Group C</i>	<i>Group D</i>
Level:	Above threshold, slower readers	Above threshold, faster readers
Materials:	Material from the main textbook with the attached reading exercises	Material from the main textbook, with the attached reading exercises plus supplementary materials.

Numerous other possible patterns of grouping could be devised, of course, depending on the size of the class and other factors. The chief arguments for grouping within a class, as opposed to streaming into different classes, are that mobility between groups is achieved more easily, and that students are kept on task as far as possible because of the better match between the texts and their ability.

It must be stressed that the teacher's role is certainly not that of a dispenser of standard solutions to reading problems. Because students differ in their skills and personal experiences, there are variations in the extent to which teacher input is necessary or appropriate. At no time should the teacher try to give equal attention to all students. Rather, the teacher should be concerned as much with the students' process of reading as with their product of maximum comprehension. With self-explanatory written instructions for the reading tasks and the availability of answers when exercises are completed, the teacher can move from group to group responding to queries, clarifying problematic areas and offering feedback.

Although the teacher should allow pupils to develop freely at their own rate, it is also essential to maintain sensitivity to pupils' progress or lack of progress so that appropriate reading materials and tasks can be selected. Progress tests will indicate desirable forms of re-grouping to encourage continued progress.

Implementation

Two factors strongly influence the ease with which the proposed approach can be implemented in any one school. The first concerns organisation of teacher resources. When extra teachers are allocated for split-class remedial teaching, the contact hours for individual teachers remain unchanged. It is thus unsurprising that recruitment of extra teachers has failed to encourage staff to undertake curriculum development in addition to their teaching duties. For optimum use of the limited extra teacher resources, it may be suggested that class sizes should be maintained, and extra teachers should be used to reduce average teaching hours. Such a reduction of the teaching load would permit more attention to development of curriculum materials.

A second determinant of success is the teachers' competence both in mixed-ability teaching and in collating and adapting teaching materials. The need for competence in these matters should be recognised by teacher trainers as well as by school administrators. It follows that teacher trainers

also need experience in school-based curriculum development. Teams could usefully be formed to encourage collaboration between university and school teaching staff.

Conclusions

Arising from a concern about student problems in the reading of English, the chief focus of this paper is on remedial teaching. Particular attention has been placed on the way that extra teacher resources are utilised. The paper has stressed the value of process-oriented teaching, and has highlighted the importance of teacher involvement in material development as well as teaching. This orientation, it is suggested, will help ensure an optimum match between reading materials and students' proficiency levels.

Although the paper is opposed to the practice of formal streaming, it recognises that students' abilities will always vary. To deal with this, the paper recommends grouping of students within their classes.

However, this recommendation still encounters thorny questions on class size. The paper has noted that both local and international research has failed to find firm correlations between class size and pupils' academic achievement. In the light of this, it has argued that additional teachers should not be used to split classes but instead to release existing teachers from class contact time and thereby to permit all staff to increase attention to development of materials.

References

- CARRELL, P. L. (1988): *'Interactive Text Processing: Implications for ESL/Second Language Reading Classrooms'*, in *Interactive Approaches to Second Language Reading*, ed. P. L. Carrell *et al*, Cambridge University Press, New York.
- CLARKE M. A. (1980): *'The Short-Circuit Hypothesis of ESL Reading-or When Language Competence Interferes with Reading Performance'*, *Modern Language Journal*, 64.
- HADDAD, Wadi (1978): *Educational Effects of Class Size*. Staff Working Paper No. 280, The World Bank, Washington.
- HO, J. C. (1985) *'Additional Teachers for Split-Class Teaching of English'*, Document LR/44, Educational Research Establishment, Education Department, Hong Kong.
- HONG KONG (1986): *'Medium of Instruction in Secondary Schools'* General Administration Circular No. 13/86, Education Department, Hong Kong.
- JOHNSON, Keith (1986): *'English Language Remedial Teaching in Secondary Schools in Hong Kong'*, *New Horizons*, No. 27.
- KWO, Ora W. Y. (1989): *'Approaching the Threshold to Interactive Reading in English: A Focus on Decoding'*, *New Horizons*, No. 30 (in press).

- LARKIN, A. I. & KEEVES, J. P. (1984): *The Class Size Question*, Australian Council for Educational Research, Hawthorne.
- RIST, Ray C. (1978): 'On Understanding the Processes of Schooling: The Contributions of Labeling Theory', in *Power and Ideology in Education* ed. Jerome Karabel & A. H. Halsey, Oxford University Press, New York.
- SEWELL, C. (ed.) (1980): *Mixed-Ability Teaching*, Nefferton, London.
- SILBERSTEIN, S. (1987): 'Let's Take Another Look at Reading: Twenty-Five Years of Reading Instruction', *English Teaching Forum*, XXV, 4.
- SMITH, Frank (ed.) (1973): *Psycholinguistics and Reading*, Holt, Rinehart and Winston, New York.

VOCABULARY LEARNING AND TEACHING: EVIDENCE FROM LEXICAL ERRORS IN THE SPONTANEOUS SPEECH OF ESL LEARNERS

Arthur McNeill

Department of English

City Polytechnic of Hong Kong

Introduction

The low priority which is often attached to vocabulary teaching has been commented on in the majority of papers on second language vocabulary during the past few years. Attention has been drawn to the dangers of leaving vocabulary learning to chance on the assumption that learners pick up most of their vocabulary incidentally. McCarthy (1984), in particular, has warned against the assumption that vocabulary acquisition operates as a 'cumulative by-product of the teaching of structures or the communicative functions of sentences.' One of the aims of this paper is to add its voice to the plea for vocabulary learning to be given a more central place in language teaching. However, its main purpose is to report the findings of a recent study of some of the lexical errors which Hong Kong students make when speaking English. The discussion will focus, in particular, on the errors which relate to the formal aspects of vocabulary learning and some related implications for vocabulary teaching.

The approach which is adopted in the study is deliberately general in character. There are several reasons for this, most of which concern the nature of recent research into the two aspects of vocabulary which this study addresses: second language vocabulary acquisition and speech error data, most of which have been limited to L1 studies. The main interest lies in two areas:—

- (a) the ways in which second language vocabulary is stored and retrieved
- (b) the ways in which learners *activate* the vocabulary (based, in this case on the words they use when they speak).

Assumptions about Vocabulary Learning

Before we look at the design and the findings of the study in detail, it might be helpful to identify some assumptions which are made about the nature of second language vocabulary learning. These will be useful in accounting for some of the peculiar features of the investigation.

1. *Vocabulary-learning as an on-going process*

Experiments in second/foreign language vocabulary learning have mainly been concerned with accelerating the initial learning of vocabulary and generally investigating the ways in which learners cope with a particular approach to teaching them new words. The present

study attempts to take into account that vocabulary learning is an on-going process, even in a second language, and concentrates on the ways in which students who have been learning English for a number of years use vocabulary in natural speech.

2. ***Direct and indirect learning***

The study also recognises that while vocabulary is learned both directly and indirectly, the bulk of vocabulary learning is indirect (Nation 1982). Consequently, the investigation is not concerned with evaluating any particular method of teaching/learning vocabulary. In direct vocabulary learning, a conscious effort is made to learn vocabulary either in context or isolation. In indirect vocabulary learning, new words are learned incidentally while reading, listening or viewing, usually as a result of the information provided by the context.

It is probably worth mentioning that there is some evidence to show that large quantities of vocabulary are learned indirectly. For example, Saragi et al (1978) found that after reading a novel, learners could recognize the meanings of 76% of the 90 new words tested. The learners in the study were not allowed to consult dictionaries and they did not expect to be tested on the vocabulary afterwards. It is assumed that the learners in the present Hong Kong study have learned vocabulary both as a result of direct approaches as well as by exposure to large amounts of reading and listening material.

3. ***Active and passive vocabulary***

Most second language vocabulary research has concentrated on learners' ability to recognise vocabulary. Typically, experiments measure learners' ability to distinguish between words and non-words with the use of tests designed to measure vocabulary size. Indeed, the concept of 'knowing' a word is often understood in the sense of the ability to recognise rather than produce the word. It is generally assumed that learners gain receptive control of new words before active control and that many vocabulary items never become part of the productive capacity and remain part of receptive competence. In focusing on data from spontaneous, spoken English, the study hopes to gain some insights into the extent to which a learner's second language vocabulary is actually used in spoken production.

4. ***Spoken Data***

Since the study is interested in the way vocabulary is organised and used, only spoken English serves as a source of data. It is obviously easier from a technical point of view to study lexical errors in text which is written rather than spoken, since the latter is more difficult to capture and describe. However, written text is processed in a more conscious manner than spontaneous speech and is edited and monitored in ways which make the product (the written text) less useful as data which can tell us something about vocabulary storage and retrieval since it is less immediate.

5. ***Errors and their relative frequency***

Another reason for the general nature of the data gathering phase is that it is intended to provide a broad picture of the types of errors which learners typically make. Rather than focus from the outset on particular error types, for example, by following on from types of error found in L1 research and designing experiments to find out whether similar patterns exist in L2, it was considered appropriate, since there are very few studies of speech errors in L2, first to build up a small corpus which consists of all kinds of speech errors, so that we could establish:—

- (a) the main error types (and make comparisons, if appropriate, with L1 studies)
- (b) the relative frequencies of the error types (in order to know whether a particular phenomenon was widespread and in order to see what the main implications for a language teaching programme might be.)

Related Research

1. ***L1 Vocabulary***

To date there have been few studies of speech errors in L2 and a key issue which presents itself is the extent to which the findings from studies of how the L1 mental lexicon is organised, based on L1 speech errors, are applicable to L2 research. To what extent does the second language learner's lexicon of a language resemble that of a monolingual native speaker of the language? The main research evidence so far has demonstrated that the word associations between the two groups are quite different (Meara 1983). For example, the relative stability of responses to word association stimuli found in monolinguals is not found in L2 learners. This might be regarded as evidence that second language words are stored and organised in a different way from the L1 vocabulary. On the other hand, it might provide evidence for the view that the second language mental lexicon is only different from the L1 lexicon because it has a far smaller number of words and does not (yet) need the sophisticated storage and retrieval faculty of the L1 mental lexicon. However, in the absence of clear research evidence either way, it makes sense for us to draw on L1 models when investigating L2 and to test their validity for L2 theory—in the meantime, at least.

Two pieces of lexical error research which influenced the way in which the HK errors are analysed are the Aitchison and Straf (1979) study of the differences between child and adult malapropisms in L1 and more recent research by Meara and Ingle (1988), which looks at the way in which English learners of French learn words. Aitchison and Straf observe that the errors made by the children in their study tend to be in the latter parts of words, whereas adults are more likely to produce the first and final syllables of target words correctly and to make an error in medial syllables. Aitchison and Straf argue from this observation that

children tend to store words in a linear manner, i.e. from left to right, whereas adults store and retrieve words according to certain 'salient features', such as the first and final syllables. Meara and Ingle make comparisons between their own findings based on L2 data and the conclusions reached by Aitchison and Straf and conclude that there are some clear differences between L1 and L2 lexical storage. The data in the present investigation have been arranged in such a way that comparisons with the above studies are possible.

2. ***Speech processing and production***

It will be useful for us to take a working hypothesis or model of how speech processing is believed to take place. We shall refer to the Fay and Cutler (1977) model, which Channell (1988) describes as a 'device for speech production . . . which consists of a grammatical frame for an intended meaning, marked with the syntactic categories of words, which are then found in the mental lexicon and placed in the grammatical frame for onward processing. Conversely, a comprehension device decodes sounds into word length segments and searches its mental lexicon for meanings to pair with them.'

Such a view implies that for both the L1 and L2 user of a language, the two distinct processes of production (speaking or writing) and comprehension (listening or reading) make differential use of the store of words in the mind. Part of the production process must consist of selection of appropriate words according to the meaning to be conveyed.

The word form is then converted into a phonological shape for onward processing into speech. Thus the direction of the mapping proposed is meaning > sound. In comprehension, the mapping is sound > meaning. We shall refer to this model of speech processing when we come to discuss some of the data.

Gathering of Data

In view of the broad interest of the project, it was necessary to try to collect as many items as possible produced in as natural a manner as possible. The method adopted was small group interviews, with four subjects taking part in informal conversations, which were recorded and subsequently transcribed for lexical error. A catalogue of topics was compiled and each of the group discussions was based on one topic only. It was hoped that by varying the topic from one group discussion to the next, it would not be possible for the subjects to prepare themselves for the discussions, for example, by rehearsing the vocabulary they expected to need. Since the main purpose was to get subjects to engage in some sort of natural, fluent exchange, topics were chosen which, it was hoped, would interest the students and yet not be so specialised or technical that some students would have any obvious advantage or disadvantage.

About 400 subjects, all of them students at the City Polytechnic of Hong Kong and drawn from different disciplines, took part in an informal 10-minute discussion in groups of four. The topics they were invited to talk about included their impressions of the new City Polytechnic campus, views on studying in Hong Kong as opposed to overseas, differences between secondary and tertiary education and aspects of their course. In order to make the discussions natural and relaxed (and to avoid any impression of an oral test), the students were told that they were taking part in a departmental research project designed to gather students' views on a number of issues and to find out how they expressed their ideas in English.

The target in this phase of the project was to build up a corpus of 500 errors. What could not really be predicted was the number of lexical errors it was reasonable to expect in a ten-minute conversation. In fact, the average was about 1 and 2 errors per student per conversation. This meant that a large number of discussion groups had to take place in order to gather sufficient data (about 40 small group discussions). Each group discussion was tape-recorded and analysed for lexical error. The errors were then coded and entered (together with the context in which they were made) into a D-Base software programme.

Error Defined

It is important to recognise that the notion of 'error' means different things to different people. It is particularly important for us to clarify what is meant by 'error' here, since we are drawing on the work of both applied linguists with an interest in SLA as well as on the work of psycholinguists, whose data are drawn, in the main, from L1. For the psycholinguists, a speech error tends to refer to unconscious slips of the tongue, resulting from wrong mental processing, while the applied linguists mean 'observable, systematic deviations from the standard norm of the target, from which we can analyse which bits of the target language the learner does not yet know, or half knows' (Channell 1988). L1 errors, then, are taken mainly to be evidence of what speakers know, whereas L2 errors are taken (mainly) to be evidence of what L2 speakers do not know. It will be appropriate to draw on both types of approach to error when we come to consider the data in the present study.

In gathering examples of lexical error, no attempt has been made to grade these in terms of error gravity, as perceived, for example, by native speakers. They are simply incorrect by any objective standard. Inevitably, some will be considered more unacceptable than others, but that does not concern us here, since we are primarily interested in identifying types of error and their frequency.

Categories of Error

The categories used are based partly on the L1 research and partly on the error analysis categories which are now widely used in second language teaching and derived from research in applied linguistics. Not surprisingly,

this results in a set of categories which looks somewhat hybrid. However, the categories adopted are intended to capture those features of L2 vocabulary error which allow meaningful comparison to be made with evidence from L1 studies, while at the same time identifying other types of error not relevant to L1 research, but potentially useful to L2 teaching.

Ten categories were adopted, as follows:—

1. *Initial*

The word produced resembles the target except for the first syllable.
Example: 'aspects' for 'prospects'

2. *Media*

The word produced resembles the target except for a medial syllable.
Example: 'extended' for 'exempted'

3. *Final*

The word produced resembles the target except for the final syllable.
Example: 'analysis' for 'analyst'

4. *Word Blend*

The word produced is a blend composed of two related words.
Example: 'Canchinese', a blend of 'Cantonese' and 'Chinese'

5. *Malapropism*

The word produced sounds similar to the target, but means something quite different. Example: 'scotch' for 'squash'

6. *Semantically-related*

The word produced is related in meaning to the target, but is not acceptable in the context. Example: 'lucky' for 'optimistic'

7. *Contraction*

The word produced is the result of a shortening of the target.
Example: 'sandwiches' for 'sandwich courses'

8. *L1 Interference*

The word produced has been influenced directly by the L1 equivalent. Example: 'grassland' (Cantonese) for 'lawn'

9. *Paraphrase*

The speaker is unable to produce the target word and expresses its meaning by paraphrase. Example: 'time between the terms' for 'term break'

10. *Register*

The word produced is correct in meaning, but not acceptable in the context. Example: 'disequilibrium' for 'imbalance'

In a number of cases, it proved impossible to assign an error to one category only. Where an error appears to belong in more than one category, it has been entered accordingly.

Distribution

The distribution of errors across the various categories is shown in Table 1.

Table 1: Distribution of errors across categories

TYPE	TITLE	OCCURRENCES	
		No.	%
1	INITIAL	16	3%
2	MEDIAL	20	3%
3	FINAL	92	18%
4	WORD BLEND	36	8%
5	MALAPROPISM	33	7%
6	SEMANTICALLY-RELATED	202	42%
7	CONTRACTION	17	4%
8	L1 INTERFERENCE	40	8%
9	PARAPHRASE	59	11%
10	REGISTER	12	2%

The ten categories fall into two groups. Categories 1 to 5 are concerned with the formal representation of words, while categories 6 to 10 relate to the ways in which the words are used, such as their semantic and collocational aspects. The 'formal' group (i.e. categories 1 to 5) consists of categories typically found in studies of lexical error in L1 speakers, while the 'semantic/collocational' group (i.e. categories 6 to 10) contains categories which are typically found in the L2 error analysis literature. Probably the most striking feature of the distribution of errors across the 10 categories is the fact that two categories show an unusually high rate of occurrence (categories 3 and 6) and that while the most frequent errors are 'semantically-related', the second most frequent are of a formal nature (category 3, final syllables). These results are interesting for a number of reasons. While the high incidence of semantically related errors (42%) may not be surprising, the second most frequent type of error, incorrect final syllables (18%), invites us to consider more seriously some of the formal aspects of the teaching and learning of new words. The discussion which follows examines only the errors in the formal group. The data in the semantic/collocational group (categories 6 to 10) are discussed elsewhere.

Discussion

1 *Stability of errors in partially correct words*

As Meara and Ingle (1988) point out, recent vocabulary acquisition research has tended to take the formal representation of words for granted and has concentrated more on collocational and semantic aspects. However, it has long been recognised that speech errors are not random. As far as L2 speakers are concerned, it is generally accepted that learners are likely to confuse words which are similar, just as L1 speakers do when they produce malapropisms and word blends. It is also clear that certain features of words are more prone to error than

others. For example, Brown and McNeill (1966), in their investigation of the tip-of-the-tongue phenomenon, demonstrated that the first syllable of a word is the most stable and the least prone to error. It appears that certain 'favoured' (Meara and Ingle 1988) or 'salient' (Aitchison and Straf 1979) features are fully specified for entries in the mental lexicon, while others are not. Examples of salient features used by L1 speakers of English to store and retrieve words from their mental lexicon are first and final consonants, syllabic structure and stress pattern.

This has led Meara and Ingle (1988) to hypothesise that if words in L1 are coded by certain features, then a similar system must exist for L2. What we shall hope to find in the data is evidence that certain types of error are systematically more likely than others. Two working hypotheses put forward by Meara and Ingle can be applied to the data here:

- (a) certain parts of words are resistant to error.
 - (b) the resistant parts will be the same for L1 and L2.
- While the first hypothesis is confirmed by the data, the second is not. Studies of L1 vocabulary have confirmed that initial and final consonants are relatively stable, while medial consonants are prone to error. An analysis of the errors in the present study shows that the final syllables are by far the most prone to error. The positions of error in the partially correct words (categories 1 to 3) are shown in Table 2. Tables 3 to 5 contain examples of the errors.

Table 2: *Position of error in partially-correct words*

SYLLABIC POSITION	OCCURRENCES	
	No.	%
INITIAL	16	13%
MEDIAL	20	14%
FINAL	92	73%

Table 3: *Examples of initial error in partially-correct words*

ERROR	TARGET	CONTEXT
ASPECTS	PROSPECTS	How good are the job <u>as</u> pects?
EXISTS	CONSISTS	The dinner <u>ex</u> ists of four courses.
EXPRESSIONS	IMPRESSIONS	My first <u>ex</u> pressions of the place.
RELATIVITIES	ACTIVITIES	We organise group <u>rel</u> ativities.

Table 4: Examples of medial error in partially-correct words

ERROR	TARGET	CONTEXT
TERRITORY	TERTIARY	Campus of a terr <u>it</u> ry institution
GENETIC	GENERIC	A more gen <u>e</u> tic view of criticism
HANDLING	HANGING	Stop han <u>d</u> ling around in the streets
EXTENDED	EXEMPTED	Hong Kong is ex <u>t</u> ended from natural disasters

Table 5: Examples of final error in partially-correct words

ERROR	TARGET	CONTEXT
ECONOMIC	ECONOMY	The econom <u>i</u> c will collapse
ANALYSIS	ANALYST	He's a systems anal <u>y</u> sis
POLITICAL	POLITICS	I'm sensitive about polit <u>i</u> cal
PRODUCE	PRODUCTS	You can buy computer produ <u>c</u> e in Hong Kong

It is not possible to make direct comparisons with the data used by Meara and Ingle because of differences in the material used. Meara and Ingle set up an experiment which measured subjects' ability to recall recently taught vocabulary. They noted the extent to which the versions of the words recalled resembled their target in terms of the correct recall of consonants. The words were selected according to a particular pattern: all were concrete nouns with the same phonological structure: CVCVC. Since the present study relies on spontaneously occurring errors, we have to rely on broader categories to analyse the formal aspects of the errors: initial, medial and final syllables. Nevertheless, there are some interesting similarities and indeed some differences between the results of the two studies. The Meara and Ingle claim that the beginnings of L2 words are relatively resistant to error are supported by the data here. Their claim that subsequent consonants are likely to be incorrectly retrieved is supported in the case of final syllables, but not necessarily by our data from medial syllables, where no significant difference emerged between initial and medial syllables. In the case of final syllables, the similarity of results is striking, particularly when considered in the context of the comparison which Meara and Ingle make with the claims arising from the Aitchison and Straf study of child and adult malapropisms in L1. Meara suggests that there are similarities between the ways in which children learn new words in L1 and the ways L2 learners learn words. The differences noted by Aitchison between the error patterns of children and adults in L1 vocabulary

appear to be relevant to the L2 learner's situation. The tendency of children to recall words from left to right, with the final syllables particularly prone to error, has some obvious parallels with L2 patterns. Aitchison concludes that lexical storage is a developing skill. As children learn more and more words, they have to develop more efficient ways of storing and recalling them and begin to rely more on the salient features typically used by adults and less on a simpler left-to-right approach. Meara's conclusion that no evidence exists to demonstrate that the learners in his study transfer typical adult L1 strategies to their L2 vocabulary can be applied to the learners in the Hong Kong study.

The high incidence of final syllable errors (73%) in the data invites further comment. The fact that the errors occurred in subjects' spontaneous speech seems to indicate that the final parts of words create greater problems for learners in spoken English than they do in writing, where the learner has more time to find the correct ending of words and to monitor his production. There is an obvious message here for the teaching of vocabulary. Learners appear to have difficulty in producing the correct final part of words and might require more guidance in this aspect of vocabulary work. In particular, the various morphological realisations of words probably require more conscious attention and practice, rather than assuming that a student who has learned the basic form of a word will be able to produce the correct form in spontaneous speech. This would mean, for example, that in addition to learning the word 'economy', a learner may benefit from being made aware of 'economist', 'economic', 'economical', etc., and indeed benefit from activities which require him to say these words.

2 *Word Blends*

This category contains errors which are produced as a result of a confusion of two other words. Typically, the errors are non-words, which consist of the first part of one word and the last part of a different word. Some examples are given in Table 6.

Table 6: Examples of final error in partially-correct words

ERROR	TARGET	ELEMENTS OF BLEND
ACCESSFUL	SUCCESSFUL	accessible/successful
CANCHINESE	CHINESE	Cantonese/Chinese
CONSTRUBUTE	CONTRIBUTE	construct/contribute
INCONFUSION	INCONVENIENCE	inconvenience/confusion
DULK	DARK	dull/dark
TRALK	WALK	travel/walk
RESISTANT	RESEARCH- ASSISTANT	resist/research/assistant

In the L1 literature on blend errors (e.g. Fay 1981), the range of possible interpretations of such errors is wide, ranging from minor phonological slips at one extreme to Freudian analyses of the students' unconscious meaning at the other. The above examples are probably easier to explain, at least in terms of the competing lexical items which contribute to the error. It is tempting to try to offer psychological explanations of the errors, for example, to speculate that the learner who produced 'resistant' (target: research assistant) was in a mental state of resistance when the research assistant interviewed him and that this led subconsciously to his producing the peculiar blend. However, in the interests of second language teaching, it will be of more practical value, though possibly less interesting, to focus on what the errors might indicate about the kinds of words which are likely to be confused.

In most cases, a phonological element, common to both of the blended words appears to provide the speaker with the opportunity to change direction and to switch into a new word. For example the 'cess' in both 'accessible' and 'successful' provides a kind of bridge which allows the speaker to slip into a new word and produce 'accessful'. Likewise, the 'con' in 'inconvenient' and 'confusion' makes it easy for the speaker to produce 'in confusion'. While in the L1 literature on word blends, the confusion of words containing similar sounds has provided evidence for the view that the mental lexicon is arranged phonologically, a more immediate conclusion from the L2 examples is that the errors occur because the confused words have been only partially learned, which means that the speaker is not totally in command of them and therefore likely to confuse them. The words may be partially learned in terms of their forms and their collocations. We might hypothesise from this that errors of this type will reduce as the learner's knowledge of the word improves, i.e., once he understands its various forms the ways in which it is typically used.

3 Malapropisms

In studies of malapropisms in L1, it is generally assumed that the errors are caused by ignorance. For example, when a speaker says (without trying to amuse) 'advocados' instead of 'avocados', in an utterance such as 'We had an advocado salad', it is assumed that he is ignorant of the correct form of the word. The distinction drawn by Fay and Cutler (1997) between 'malapropisms' and 'semantic errors' is appropriate here, since the source of confusion is phonological. However, in the errors which were identified as 'malapropisms' in the study, it is more likely that mispronunciation is the reason for the production of words which are similar in sound, but different in meaning to their target. Some examples are given in Table 7.

Table 7: Examples of malapropisms

ERROR	TARGET	CONTEXT
SCOTCH	SQUASH	My hobbies are badminton and <u>scotch</u> .
CHAINED	TRAINED	I was never <u>chained</u> as a teacher.
ESSENCE	ACCENT	The <u>essence</u> of his English is Australian.
PLEASURE	PRESSURE	We're under much <u>pleasure</u> at the moment.
SHIRT	CERT	I passed the school <u>shirt</u> .

When analysing L2 malapropisms, a key issue is to decide whether the error is caused by a breakdown in the process of speech production, for example, by a misselection from the mental lexicon, or whether the word produced has not yet been fully acquired in phonological terms. Generally speaking, the former type are made by L1 speakers, while most malapropisms produced by L2 speakers are of the latter type. In the case of the above examples, it is probably safe to assume that the speakers had simply not grasped the pronunciation of the items. The errors are nonetheless interesting from the L2 teaching point of view, since they demonstrate that such errors can lead to misunderstandings and even communication breakdown.

As in our discussion of word blends, we might hypothesise that an L2 learner is likely to produce more malapropisms of the 'ignorance' type as his store of words increases. With a large vocabulary at his disposal, the learner has greater opportunity of confusing words of similar sounds. While, examples of this type of error occur, they do not appear to be frequent.

Conclusion

In terms of identifying the relative frequency of lexical errors in spoken English, the study has demonstrated that problems with the forms of words are widespread. In particular, the final syllables of English words are unusually prone to error by L2 speakers. The similarities which we noted between the way in which children who are native speakers of English learn English words and the way foreign language learners learn words are further supported by the data from Cantonese learners of English.

Communicative language teaching has tended to encourage learners to develop their fluency in spoken English and to concentrate on getting their main message across without worrying unduly about grammatical accuracy. When learners make grammatical errors when they speak, it is usually assumed that they have temporarily forgotten or misapplied a grammatical rule. It is assumed that the rules have been learned at some time or other, since most language courses have a solid grammatical

base, whether they are designed along structural or functional lines. However, few courses contain a conscious lexical dimension, i.e. a focus on how words operate, including the various forms which a word can take. Unlike grammatical errors, formal lexical errors cannot simply be explained in terms of learners forgetting rules. It is more likely that the learners have never come to terms with the various forms of a word and that they have had too little opportunity to pronounce the words. In order to improve on this situation, vocabulary teaching must include morphological variation and the pronunciation of words, in addition to the semantic and collocational aspects.

References

- Aitchison, J. and Straf, M. 1979. *Malapropisms and child-adult differences in the mental lexicon*. Working Papers of the London Psycholinguistics Research Group, 1, 27–41.
- Brown, R. and McNeill, D. 1966. *The 'Tip of the Tongue' phenomenon*. Journal of Verbal learning and Verbal Behaviour, 5, 325–327.
- Channell, J. 1988. *Psychological approaches to L2 vocabulary acquisition*, in Carter, R. A. and McCarthy, M. (eds.) Vocabulary and Language Teaching. London: Longman.
- Fay, D. and Cutler, A. 1977. *Malapropisms and the structure of the mental lexicon*. Linguistic Inquiry 8, 505–520.
- Fay, D. 1981. *Substitutions and splices: a study of sentence blends*. Linguistics, 19, 7/8.
- McCarthy, M. 1984. *A new look at vocabulary in EFL*. Applied Linguistics, 5/1, 12–21.
- Meara, P. 1983. *Word associations in a foreign language*. Nottingham Linguistics Circular, 11/2, 29–38.
- Meara, P. and Ingle, S. 1986. *The formal representation of words in an L2 speaker's lexicon*. Second Language Research, 11/2, 160–171.
- Nation, I. S. P. 1982. *Beginning to learn foreign language vocabulary: a review of the research*. RELC Journal, 13/1, 14–36.
- Saragi, T., Nation, I. S. P. and Meister, G. F. 1978. *Vocabulary learning and reading*. System 6, 72–78.

ESSAY MARKING: A COMPARISON OF CRITERION-REFERENCED AND NORM-REFERENCED MARKING

David Coniam

Hong Kong Examinations Authority

Introduction

Whether criterion referencing or norm referencing is a more reliable method of marking composition scripts is an argument that has been going on for some time. With its new Form 7 Use of English examination, the Hong Kong Examinations Authority is about to move from criterion-referenced to norm-referenced marking. The small-scale study discussed in this paper examines the reliability of both types of marking.

Background

Since the 1960s, a number of research projects have focused on different systems of composition marking with a view to gauging the reliability of these systems. Researchers have looked at the efficacy of using an accuracy-count method whereby marks are deducted for grammatical errors; and the impression method where a piece of writing is awarded an overall grade rather than having marks deducted for errors. They have also investigated the influence on reliability which single or double marking (i.e. when a composition is marked by one or more markers) has when overlaid on both the accuracy and impression systems.

1. Britton, Martin & Rosen (1966) found that a system of using a team of three impression markers and one accuracy-count marker produced the most reliable results.
2. A study by the University of London in 1972 concluded that double impression marking was at least as reliable as a conventional accuracy-count marking scheme.
3. Wood & Quinn (1976) concluded that a change from single to double marking would produce an improvement in overall examination reliability.
4. King (1980) found that double-impression marking gave results which were at least as reliable as, and in most cases more reliable than, traditional accuracy-count marking.

In spite of the above evidence, the impression that the author has received from Form 6 and 7 teachers through seminars, informal contact and in the written comments which are submitted in a Markers' Report after markers have been involved in the actual marking process, is that teachers feel that the accuracy count method is fairer to students/candidates, as well as being a generally more reliable method of assessing students' writing. Typical

comments are that 'a certain number of scripts had to be given the top grade, but that very few scripts really merited it' or 'I had to demote scripts to the bottom grade when I felt that none were really that poor'.

The Hong Kong Situation

1. The three major English language examinations in Hong Kong are:—
 - (i) The Hong Kong Certificate of Education Examination—taken at the end of Form 5 (16–17 years); this is the principal secondary school certificate for most students, a grade C being equivalent to a London University GCSE. [This is the examination which 16-year-olds take in the U.K. after which the majority leave school. It is equivalent to Grade 11 in the American Education System.]
 - (ii) The Hong Kong Higher Level Examination—taken at the end of Form 6 (17–18 years); this is principally the English language entrance requirement of the Chinese University of Hong Kong.
 - (iii) The Hong Kong Advanced Level Use of English—taken at the end of Form 7 (18–19 years). This was principally created as the University of Hong Kong's English language entrance requirement. The revision of the examination is, however, expected to lead to the examination having much wider currency in Hong Kong, both in tertiary institutions and the community as a whole. [see 2.2 below] To be able to sit the Use of English examination, candidates must already have achieved a pass (grade E) or better in the Hong Kong Certificate of Education Examination. This obviously makes for a much greater homogeneity of ability at the Use of English level.
2. For a number of reasons, the Advanced Level Use of English examination has undergone serious revision: so much so in fact that the revised examination—to be first examined in 1989—is virtually a new examination, since of the four papers in both the new and the old examinations, the only paper that bears any resemblance whatsoever to the previous examination is Writing, the former Written Composition paper. Here, candidates have to write 500 words on a single topic in 75 minutes as opposed to 400 words in 60 minutes, as was the case previously. While the paper format has only changed slightly, the marking scheme is totally different from its predecessor, however, and follows the lines of the pattern-marking schemes for the Hong Kong Certificate of Education Examination and Hong Kong Higher Level written composition papers.

Marking Schemes

1. New Use of English Writing Examination (i.e. 1989 onwards)

The Use of English Examination, as from 1989, will now be impression-marked, like the Higher Level and the Certificate of Education examinations.

Written compositions have to be assessed impressionistically on a 9-point scale (1 being the weakest and 9 the best). A grade is awarded by the marker who assesses the script on its content and its organization, grammar and style.

The marker is asked to award the grade to a script bearing in mind an approximate 50–50 split between accuracy and content. Marking is strictly norm-referenced and the following distribution curve must be adhered to:—

Figure 1

No of exam. scripts	2%	8%	12%	16%	24%	16%	12%	8%	2%
	1	2	3	4	5	6	7	8	9
	Grade								

[See Appendix 1]

2. Former Use of English Marking Scheme

A total of 25 marks were allocated to Written Composition:—

Command of Language/Communication—20 marks

Content and Organization — 5 marks

[See Appendix 2]

The marker was not asked to conform to a pattern: he could award any script whatever mark he felt that script merited. The corollary of this method was that marks awarded to scripts were bunched around the lower end of the scale, with the mean typically around 35%.

The Study

The hypothesis was that in norm-referenced pattern marking inter-marker correlations would be higher, as would be the correlations between this method of marking the written composition and the overall examination result, i.e. (whole examination minus written composition).

Previous studies (King, 1981; Cheung 1983) have found strong evidence that double-impression marking greatly improves inter-marker reliability and gives a higher correlation between the paper being marked and other papers in the exam.

In 1988, the mean of the individual correlations of the Written Composition paper with the other three papers of the examination was .39 (the correlation was usually in the range .3 to .4). The correlation of the Written Composition paper with the other three papers together was .51. (This was again comparable with previous years; in contrast the range for the Certificate of Education examination is .7–.8).

In order to investigate the efficacy of changing the Use of English marking scheme from single criterion-referenced marking to norm-referenced double-impression marking, it was proposed to utilise the expertise and experience of seven experienced markers who had all had experience of marking using both marking schemes. They would mark (i.e. re-mark) a

number of candidates' actual 1987 examination scripts using both systems of marking. Markers would mark a number of scripts using one method and then, after a period of time, mark the same scripts again, but this time using the other marking scheme.

400 of the 1987 examination scripts on a single topic [See Appendix 3] were retrieved, and from these, 50 scripts were selected to represent a broad cross-section of candidate ability. These 50 scripts were then typed up to avoid handwriting/layout bias. [Previous studies (King 1980) have shown that scripts with good handwriting or good readability—e.g. where a candidate has written on alternate lines—do get awarded higher grades by examiners.]

Markers were first asked to mark the 50 scripts following the old Use of English marking scheme, i.e. by criterion referencing. After a gap of roughly two months, the markers were then asked to mark another 50 scripts, but to grade them according to the pattern laid down for the new Use of English marking scheme, as outlined 'The Hong Kong Situation' above.

Markers were informed that they would recognize some of the scripts, as a number of scripts from the first batch had been included for purposes of comparison. It was assumed that the markers would not recognize that they were re-marking the same 50 scripts they had marked two months previously. This assumption proved to be correct as none of the seven markers commented that the two sets were the same.

Results

Table 1

Marker	Criterion-referenced single marking						
	A	B	C	D	E	F	G
Mean correlation with other markers	.60	.63	.71	.69	.73	.69	.74
overall mean=.68							

Table 2

Marker	Pattern-marked single marking						
	A	B	C	D	E	F	G
Mean correlation with other markers	.78	.73	.76	.75	.76	.66	.77
overall mean=.74							

Comment

Table 1 and 2 show the inter-marker correlations for single marking, i.e. where each marker's marking is simply correlated with the other single markers' marking. The mean correlation between the six markers for the criterion-referenced marking was .68; the corresponding pattern-marking figure was slightly higher at .74.

Table 3

Marker	Criterion-referenced single marking						
	A	B	C	D	E	F	G
Mean correlation with other markers	.81	.81	.82	.82	.83	.82	.84
overall mean=.82							

Table 4

Marker	Pattern-marked single marking						
	A	B	C	D	E	F	G
Mean correlation with other markers	.86	.85	.86	.86	.83	.84	.86
overall mean=.86							

Comment

Tables 3 and 4 show paired correlations. Here, the *sum* of two markers' marks (e.g. markers A and B) are correlated with the *sum* of two other markers' marks (e.g. C and D). This is done on the grounds that two sets of marks correlated with two other sets give greater reliability than one set of marks correlated with one other set. Each marker then has 60 unique sets of paired correlations (AB with CD, AC with CE, AB with CF etc.); the mean correlation for each marker is therefore the mean of the 60 unique sets. As the tables show, the criterion-referenced mean correlation is .82, the pattern-marked figure again slightly higher at .86.

Table 5

Criterion-referenced marking: Single Marking
Mean marker correlation with 1987
(whole examination—written composition)=.70

Table 6

Pattern marking: Single Marking
Mean marker correlation with 1987
(whole examination—written composition)=.70

Comment

Tables 5 and 6 show identical correlation for single criterion-referenced and pattern-marked marking of .70 with the candidates' (whole examination minus written composition) scores. This may be attributed to the fact that the old use of English papers had an approximate .54 correlation with the other papers of the examination as opposed to closer to .6 for the new