Editorial Advisory Board

Associate Professor Lawrence Chia, Anglo Chinese School
Professor Thomas J Shuell, State University of New York at Buffalo
Professor Philip G Altbach, State University of New York at Buffalo
Professor Zainal Ghani, University of Science, Malaysia
Professor J M Sinclair, University of Birmingham
Associate Professor N S Prabhu, National University of Singapore
Dr Gwee Yee Hean, Singapore
Professor Richard Pring, University of Oxford
Professor Barry J Fraser, Curtin University of Technology

Editorial Board

S Gopinathan, School of Education (Executive Editor)
Toh Kok Aun, School of Science
Tay-Koay Siew Luan, School of Education
Oliver Seet, School of Arts
Grace Loh, School of Arts
John Saunders, School of Physical Education

© National Institute of Education
Nanyang Technological University
469 Bukit Timah Road, Singapore 1025

The Singapore Journal of Education is published for the National Institute of Education by
Longman Singapore Publishers Pte. Ltd.
25 First Lok Yang Road
Singapore 2262

The principal aim of the Journal is to disseminate to the education community, both in Singapore and abroad, research conducted in education, and to provide a channel for discussion and research in education. This Journal is the sole responsibility of the National Institute of Education. The views expressed herein do not necessarily represent those of the Editorial Board, the National Institute of Education, or of the Publishers.

The Editorial Board welcomes contributions on educational developments, issues, pedagogy, research and related aspects. Articles, book reviews, research papers, etc., are welcome. All manuscripts should conform to the instructions given in the style sheet on the inside rear cover of this Journal, and should be submitted to the Executive Editor, Singapore Journal of Education, National Institute of Education, 469, Bukit Timah Road, Singapore 1025.

ISSN 01-29-4776
CONTENTS

1  Barry McGaw
   Measuring and Monitoring Educational Achievement

26  Kathy Sylva
    Work or Play in the Kindergarten?

35  Peter Bodycott
    Monitoring Protolinguistic Development in a Bilingual Context

50  Toh Kok Aun

61  Foong Pui Yee
    Development of a Framework for Analysing Mathematical Problem Solving
        Behaviours

76  Philip Wong
    Influence of Computer-Generated Visuals on Word-Problem Solving

RESEARCH REPORT

88  Soh Kay Cheng
    Measuring Motivation to Learn Chinese and English Through Self-Reported
        Feelings and Behaviours

DISSERTATION ABSTRACTS

95  Peter Lam
    Optimal Test Designs with Content Balancing and Variable Target
        Information Function as Constraints

96  Fong Ho Kheong
    Modelling Children's Thinking in Solving Ratio and Proportion Problems:
        An Information Processing Perspective
Notes on Contributors

1 Barry McGaw is a distinguished Australian academic and currently Director of the Australian Council for Educational Research. He has an impressive record of research and has served on numerous national committees.

2 Kathy Sylva is Professor of Child Development and Primary Education at the University of London Institute of Education. Professor Sylva has an impressive publishing record and is at the forefront of research in child development.

3 Peter Bodercott is a lecturer at the Learning Development Centre, University of Wollongong. He was with the National Institute of Education before he left for the University of Wollongong.

4 Toh Kok Aun is Senior Lecturer in the Division of Physics, School of Science at the National Institute of Education.

5 Foong Pui Yee is Lecturer, Division of Mathematics, School of Science. She obtained her doctorate from Monash University, Australia.

6 Philip Wong is Lecturer, Division of Instructional Science, School of Education. He obtained his doctorate from the University of Minnesota.

7 Soh Kay Cheng is Senior Lecturer in the Division of Psychological Studies, School of Education at the National Institute of Education. He is also on assignment to the Centre for Applied Research in Education.

In the 1960s, expectations of education system reform led many countries to believe that education reform would lead to improved educational attainment and individual development. In the USA, for example, the National Education Goals of 1966 defined a national goal to be achieved by the year 2000. The goals were designed to ensure that all students would achieve the highest possible level of education attainment.

In the 1970s, doubts about the effectiveness of education reform emerged. Critics argued that the emphasis on individual achievement had come at the expense of social and economic progress. They called for a more balanced approach to education reform.

In the 1980s, the emphasis shifted to the role of government in education reform. In Canada, for example, the federal government took responsibility for education reform in the provinces.

In the 1990s, the role of government in education reform continued to evolve. In the USA, for example, the role of the federal government in education reform has been questioned, with some arguing that too much power was being concentrated in the hands of the federal government. This paper was prepared for the 1993 Australian Conference on Education Reform.
Societies continue to have high expectations of education. There may be disappointment with the current achievements of education systems but commitments to change and reform bear testimony to an underlying faith that education has much to offer - to individuals and to societies as a whole.

In the 1960s and 1970s, disaffection with education led many countries to increase the level of funding for education. Developed countries that found their level of funding as a proportion of gross national product was below the average of other similar countries were inclined to use that average as a goal to be achieved - in the process raising the average of course. The preoccupation was with inputs.

In the USA, President Johnson's Elementary and Secondary Education Acts in the 1960s directed national government support to schooling. In Australia, a similar federation of States in which the national government has no formal responsibility for education, the national government also directed funds to schools and school systems. A committee of inquiry chaired by the economist, Professor Peter Karmel, in 1973 defined a series of resource targets to be achieved through increased funding for schools.

Doubts about whether the increased resources were yielding the desired effects emerged soon enough. In the US, Jesse Jackson developed his program PUSH/EXCEL because he saw that it was not enough for more resources to go to schools for black students. For him, the criterion of success had to be whether black students were achieving more. The strategy may be to improve inputs but the goal has to be to improve outcomes. In Australia, a decade after his first Committee's report, Professor Karmel chaired a new committee, called the Quality of Education Review Committee and charged to report on the effects of the substantial increases in education funding that had occurred over the previous decade. I was a member of that Committee and I would have to say that, while many people believed there had been great benefit, almost no one had any systematic evidence.

### Pessimism about Schools

At the time as investments in education were being increased in the expectation of improvement, research evidence was being accumulated that suggested a much more pessimistic outlook. In its Civil Rights Act of 1964, the US Congress directed the Office of Education to conduct a survey "concerning the lack of availability of equal educational opportunities for individuals by reason of race, colour, religion, or national origin in public educational institutions". The study *Equality of Educational Opportunity* (Coleman et al., 1966) commissioned to fulfil this requirement was expected to identify those characteristics of school resources and practices which made a difference to educational achievement in order to develop programs for school improvement.

This paper was presented at the Ruth Wong Memorial Lecture, Singapore, 27 March 1992.
The authors of the report concluded:

Taking all these results together, one
implication stands out above all: That schools
bring little influence to bear on a child's
achievement that is independent of his
background and general social context; and
that this very lack of an independent effect
means that the inequalities imposed on
children by their home, neighbourhood, and
peer environment are carried along to
become the inequalities with which they
confront life at the end of school (Coleman
et al., p 325).

This finding was, of course, highly
controversial since its message that schools could
do little to remove differences in educational
achievement and adult life chances due to
differences in social context flew in the face of
the strong prevailing commitment to educational
reform as a means of social reform. Numerous
re-analysis of the Coleman data were undertaken
and the results of other, smaller-scale studies
which pointed to stronger school influences were
appealed to in an effort to challenge the
conclusion that schools were impotent. The most
intensive review, conducted in a three-year
Harvard faculty seminar, in general
re-formulated the Coleman Report's major conclusions
(Mosteller and Moynihan, 1972).

With the challenges to Coleman's research
findings apparently unsuccessful, an alternative
interpretation of his findings was formulated.
According to this view, it was not that schools
were failing to provide equal opportunity as
they were intended or that they were doing so
in a way that social research could not detect. It
was argued that educational programs
successfully achieved quite different purposes,
those of reproducing and legitimating the existing
social order.

This radical view was well-expressed in the
US by Bowles and Gintis:

In promoting what John Dewey once called
the "social continuity of life," integrating new
generations into the social order, the schools
are constrained to justify and reproduce
inequality rather than correct it... Thus efforts
to realize egalitarian objectives are not simply
weak, they are also... in substantial conflict
with the integrative function of education.
The educational system legitimates economic
inequality by providing an open, objective,
and ostensibly meritocratic mechanism for
assigning individuals to unequal economic
positions (Bowles and Gintis, 1976, pp 102-
103).

This US critique had its parallels in Australian
writing (eg Connell, 1977; Connell et al., 1982).
Connell (1977) argued that:
The school system selects from all its clients
a smaller number to go on to "higher training"
or to enter programs which normally lead to
higher training... The main means of selection
are formal examinations and standardised
tests of abilities and skills... [and] there is
extensive evidence that children from upper-
status backgrounds do better in these paper
contests than children of lower-status
backgrounds (p 156).

These analyses suggest that, whatever the
good intentions of educational reformers seeking
to achieve social reform, they will inevitably fail because the whole educational process has
been constructed in a way that preserves the
status quo and so acts strongly and effectively
against reform.

A New Optimism

The optimism of the 1960s and early 1970s had
been dashed by the research claims that schools
had little impact. The social reproduction
explanations of writers such as Bowles and Gintis
and Connell in the late 70s and early 80s
generated a darker pessimism. A decade later
two important research developments have
created a new basis for optimism.

One has been the development of statistical
methods for analysing multi-level data. The data
typically obtained in studies of effects of
schooling are multi-level. Some are measures of
student characteristics, such as background
characteristics and school achievement. Others
are measures of classroom characteristics; others
of school characteristics. The older analytical
procedures essentially forced researchers to treat
their data as though they were all at one level.

The new multi-level techniques allow the
appropriate identification of stronger effects
than before.

The second is the reframing of the equation of
schools. The optimism of the 1960s was
achieved at the cost of ignoring related different
differences. Reformers seeking to address
inequality rather than correct it... Thus efforts
to realize egalitarian objectives are not simply
weak, they are also... in substantial conflict
with the integrative function of education. The
educational system legitimates economic
inequality by providing an open, objective,
and ostensibly meritocratic mechanism for
assigning individuals to unequal economic
positions (Bowles and Gintis, 1976, pp 102-
103).

This US critique had its parallels in Australian
writing (eg Connell, 1977; Connell et al., 1982).
Connell (1977) argued that:
The school system selects from all its clients
a smaller number to go on to "higher training"
or to enter programs which normally lead to
higher training... The main means of selection
are formal examinations and standardised
tests of abilities and skills... [and] there is
extensive evidence that children from upper-
status backgrounds do better in these paper
contests than children of lower-status
backgrounds (p 156).

These analyses suggest that, whatever the
good intentions of educational reformers seeking
to achieve social reform, they will inevitably fail because the whole educational process has
been constructed in a way that preserves the
status quo and so acts strongly and effectively
against reform.

A New Optimism

The optimism of the 1960s and early 1970s had
been dashed by the research claims that schools
had little impact. The social reproduction
explanations of writers such as Bowles and Gintis
and Connell in the late 70s and early 80s
generated a darker pessimism. A decade later
two important research developments have
created a new basis for optimism.

One has been the development of statistical
methods for analysing multi-level data. The data
typically obtained in studies of effects of
schooling are multi-level. Some are measures of
student characteristics, such as background
characteristics and school achievement. Others
are measures of classroom characteristics; others
of school characteristics. The older analytical
procedures essentially forced researchers to treat
their data as though they were all at one level.

The new multi-level techniques allow the
appropriate identification of stronger effects
than before.

The second is the reframing of the equation of
schools. The optimism of the 1960s was
achieved at the cost of ignoring related different
 Page 2 Vol.13, No.1
The new multi-level procedures treat the data at the appropriate level and generally reveal much stronger effects of schooling than had been found before.

The second new development has involved reframing the question about the effect of schools. The early studies measured student achievement on a single occasion and then related differences among students on this occasion to home differences and school differences. Recent studies, such as Mortimore et al. (1988) in the UK and Ainley and Sheret (1992) in Australia, have measured achievement on more than one occasion and then tried to account for growth. The conclusions so far are that schools have a more powerful influence on growth in educational achievement than do the background variables.

So we can conclude with some confidence, as we had wanted to in the 1960s and early 1970s but by the early 1980s feared we could not, that schools do make a difference. What is also clear from this new research is that schools differ in their impact. Some produce more growth than others. Or, to put that another way, some add more value than others.

### Monitoring Educational Achievement

The picture, however, is not a simple one because to judge a school it is necessary to separate the effects of school and background factors. Schools should be judged in terms of what they have added to what they started with. If we do not take this approach we fall into what Scriven, in his *Evaluation Thesaurus* calls the Harvard fallacy. He warns that the quality of Harvard graduates need not be due to Harvards’ teaching program. It may be due to its student recruitment program. Harvard’s success, Scriven suggests, may be due to its Admissions Office (which is not even on campus).

In Australia, there is a tendency to judge schools simply by the examination results they achieve. That can lead to the wrong conclusions. Ainley and Sheret’s study of upper secondary schools in New South Wales illustrates this point nicely. In their first analyses they have been investigating what factors cause the differences among schools in their abilities to hold students to the end of Year 12.

They found that the school in their study with the lowest retention rate was actually doing better than almost every other school given the nature of its student body. If these students had been in any other school they would have been less likely to stay at school.

Students like them in other schools were less likely to stay. At the other end of the scale, some schools with very high retention rates were doing worse than could have been expected given their students.

The same will be true when Ainley and Sheret complete the current analyses of HSC examination results (A-levels). Some schools with very good results will be shown to be doing worse than could have been expected given the kind of students they have enrolled. The public perception is that these are outstanding schools since the public has only the examination results to judge by. It takes a further analysis to assess what schools have added to what they started with.

There is a further problem with the use of examination results to judge schools or the education system as a whole. The results are norm-referenced, or normative. They are not expressed on any underlying scale which gives them a constant meaning. To make this point clearer I now want to talk about the nature of scales in educational measurement and to show you some examples of work we have been doing at the Australian Council for Educational Research in scale construction for monitoring education systems.

### Notion of Scales

#### Physical Scales as the Model

The origins of psychological and educational measurement lie in the development of scales in other domains, particularly in the development of scales for physical measurement. For physical characteristics such as height, weight, pitch of sound, and so on, it is possible
to calibrate a measurement scale independently of the process of measuring the properties of particular objects. The units on the scale can be arbitrarily defined, as they are in the definition of centimetres or inches, but they have a permanent meaning which transcends individual measuring devices and particular measurement applications. As we well know, the scales do not need to have a natural zero in order to have a clear basis for scale definition and for calibration. The everyday temperature scales developed by Celsius and Fahrenheit are perhaps the most obvious examples of scales on which zero clearly does not mean an absence of the property being measured.

The earliest attempts at psychological measurement were in the domain of psychophysics with the construction of scales of human judgement of properties for which there were external physical scales as well. The human judges were not asked to estimate the absolute value of the property but rather to make judgements of differences. They were for example, asked to judge whether either of two weights was greater and not what the weight of either of the weights were. The experimental approach, as many of you will remember from undergraduate laboratory classes if for no other reason, was to determine the smallest difference between pairs of objects which was just noticeable. From these 'just Noticeable Differences' scales of human judgement were constructed for a great variety of physical properties.

Normative Referents for Social Scales

The interests of psychologists, and educationists, soon enough moved to other scales of human performance for which there were no external physical scales to serve as referents. For that, they needed a new approach to scale definition and development. It is not coincidental that the development of techniques in psychological and educational assessment arose within the tradition of differential psychology. The view was that, without any external referent as a basis for scale definition, individuals' performances could be understood only in comparison with the performances of others. A whole series of techniques was developed for expressing the relative standing of individuals, including percentile scales, standard score scales, stanines, and so on - all of them intended to locate individuals in terms of the distribution of the performances of some population of individuals.

Use of Norm-Referenced Scales

This focus on individual difference both served and created a desire to identify and then capitalise on individual differences in various ways. For some, scale development became simply the pursuit of instruments on which individual differences could be revealed rather than the pursuit of a theoretical understanding of the nature of individual differences. In reference to some of the earliest work on measuring intelligence, Glaser (1991) reminds us that by 1905 Binet had spent about 15 years accumulating data on individual differences. His writings indicate that he had not been able to approximate a satisfactory definition of the nature of intelligence, ... [though} he remained convinced of the importance of developing an instrument to measure intelligence' (p 3).

Whatever Binet's theoretical interests, the purpose for which he was working was much less lofty than the pursuit of an understanding of human intelligence. His work had been commissioned by the Minister of Public Education in France who simply "wanted some means of ensuring that the benefits of instruction were not wasted on children who had difficulty learning" (Glaser, 1990, p 3). The focus was on individual differences and the capacity to measure them was harnessed in order to identify and exclude those who differed in a particular way from the norm.

The motives were not always to exclude, of course. In his book, The Measurement of Intelligence, a guide to the use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale, Terman (1919) pointed to the empirical evidence that teachers tend to overestimate the intelligence of retarded children and underestimate the intelligence of their more able children (pp 215-216). The motive, in this case, was to make sure that students of both types could be identified and placed in school (pp 217-218). Terman's research and his writings indicate that studies were conducted that determined what minimum 'intelligence' was required for passing success in early school.

Few believe that precise predictions of the future can be made, but this tradition does seem to have left a mark on educational assessment. Stevenson, a student of Piaget's at the University of Geneva, later moved to Taiwan, Japan, and Singapore, where he was asked to assess the performances of Japanese pupils and to identify the "correct thing. The US would do this, but so were less interested in doing things that count. Within education, as in broader sources of value, the dominating method of assessment was normative assessment. Thus, normative assessment was employed to identify changes in pupil performances; if they were moving in the right direction, a remained stable; and if they were moving in the wrong direction, a change was brought about. The reason, of course, if you are a normative psychologist, is that the strategy is advantageous. There were only 50 states in the US, the Watts Riots were below the notice of the US. The purpose of the six major IQ tests that were the available tests at the time was to find a mythical Lake Michigan. Selection of students has a name in the picture of The Education of Dr. Keilor in The Education of Dr. Keilor program as the "right sort of type". The education of all the children was left over.

More important to the present dissatisfaction with the School of Education by their critics is the existence of a series of institutions that were not invented by the School of Education but that played a central role in the restructuring of education. These institutions were created to serve the needs of the School of Education and were described by their critics as "the education system".
and underestimate the intelligence of superior children (pp 24-27). With the measures provided by the new intelligence scales, Terman argued that students could be more appropriately placed in school (pp 16-17). In addition, he speculated that the tests would soon be widely used to determine vocational fitness at least to the extent that studies would have determined "the minimum 'intelligence quotient' necessary for success in each leading occupation" (p 17).

Few believe any longer that this kind of precise prediction of peoples' futures is possible but this tradition of psychological testing has left a mark in the Western world at least. Stevenson, a developmental psychologist at the University of Michigan, compared parents in Taiwan, Japan and the USA. When parents were asked to account for how well their children were performing at school, the Taiwanese and Japanese parents said 'effort' was the important thing. The US parents said 'intelligence' - and so were less inclined to think anything could be done about it.

Within education, there were two other sources of dissatisfaction with normative assessment. One was the technical concern that normative assessment is sensitive only to relative changes in position. If the entire population were moving, for example, and all members remained stable with respect to one another, normative measures would declare no change had occurred. It only detects and counts as change, a change with respect to others. Of course, if you can play the game correctly, normative assessments can be used advantageously. Cannell (1988) "surveyed all 50 states in the USA and found that no state is below the norm at the elementary level on any of the six major nationally normed, commercially available tests" (p 5). The phenomenon now has a name - the Wobegon Effect, after the mythical Lake Wobegon presented by Garrison Keilor in The Prairie Home Companion radio program as the town where all the men are handsome, all the women are good looking and all the children above average.

More important, in education, was the dissatisfaction that normative assessments are, by their comparative nature, fundamentally competitive. Notions of personal growth or of 'personal bests' do not emerge since the yardstick is always the performance of others, not one's own prior performance.

Unease About Competitive Assessment

Criterion Referenced Assessment as an Idea

Disaffection with the competitive nature of normative assessment reached the point in the USA at least in the 1960s where, to achieve other assessment ends, a new technology of assessment was advanced. Glaser (1963) developed the idea of 'criterion-referenced assessment' in order to switch the gaze of the test developer from individual differences in performances to the underlying scale. In effect, this was an attempt to shift the test developer's focus from measurement to scale construction or from measurement to calibration.

The techniques were developed by others (see Popham, 1978) and various attempts were made to develop a more rigorous psychometric theory in terms of which to evaluate and improve criterion-referenced items and scales. The techniques never appeared, however, to have the power of the classical test theory upon which normative assessment had been constructed. Precision of assessment, for example, which had been expressed in terms of their reliability index for classical test theory never found a satisfactory equivalent expression in criterion-referenced test theory. Furthermore, the properties of the scales seemed too dependent upon the a priori definition which guided the item construction. There were no adequate procedures for judging the fit of items to the set of other items or to the notion of the scale upon which the selection depended (Harris et al., 1974).

Separation of Calibration and Measurement

The solution to the problem lay in another direction. Instead of the development of an alternative purpose built assessment strategy with
of the test (the items or tasks) can be represented in terms of their ability or achievement as shown in Figure 1. For attitude and other scales, the relevant analogues of the terms 'difficulty' and 'achievement' can be substituted.

### Differentiation by Purpose and Not Psychometric Theory

#### Numeracy and Literacy in Victoria

The differentiation of approaches to assessment by purpose rather than by psychometric theory is well illustrated in the strategy we adopted in surveying numeracy and literacy levels in Victoria (McGaw et al., 1989). We calibrated items for a Year 9 mathematics test onto a scale as shown in Figure 2. The items are arranged from the most difficult at the top to the easiest at the bottom. A sense of the scale can be obtained by looking at the arithmetic items that are shown. The easiest of the items shown in the figure is the simple addition 125 + 5; next is the subtraction 6.6 - 2.4; then the multiplication 38 X 9; the approximation of the division 1240 / 4, and finally the division 243 / 4.

The scale itself has interval properties so that pairs of items the same distance apart differ in difficulty by the same amount. In order to give the scale numerical values, we selected an arbitrary scale, just as Fahrenheit and Celsius selected arbitrary numerical scales for their representations of temperature. Our choice was approached negatively. We avoided a scale with zero or 100, since we did not wish to imply that we had somehow defined 'nothing' or 'everything'. Furthermore, because of the peculiar meaning with which '50' has been invested, implying that to know half of what someone thought to ask you is sufficient, we ensured that, though our scale range includes the value 50, it clearly cannot have its traditional meaning. Our scale values for the items ranged from about 20 to about 60. The overall scale and its meaning at different levels can be understood in terms of the items that define them.

We also developed a mathematics test for use with Year 5 students and, by having some items common, we could compare all items on a section of the scale. Figure 3 shows the right those two segments to look at the distribution of scores among Year 9 items and the Year 5 items.

With this scale, we also developed representative items for Year 6 and Year 9 and, where possible, we did so on the scale. The typical normal distribution for the Year 9 distribution is much wider than for the Year 8 distribution. There is a subjectivity in one sense to an extent that the Year 9 items have known items from previous years. Average scores on these scales we for each additional year of life?

What is my view, is that people, in terms of the kind of relative standard as an example, possess 'competence' and judge students in mathematics as we set our definition of Year 9 students as about 51% of the base case and of course, but not with dependence, we have different definitions of the scale at any age. Although the separation of these processes, a statistical separation in separation in difficulty of the items and the level of performance among students. The product of an important
items common to both tests, were able to locate all items on a common scale. Figure 3 shows a section of the scale with the items on the left being those from the Year 5 test and items on the right those from the Year 9 test. If you were to look at the whole scale, and not just the segment shown in Figure 3, you would see the Year 9 items extending further up the scale and the Year 5 items extending further down.

With this scale we were then able to measure representative samples of individuals at Year 5 and Year 9 and to estimate their performances on the scale. These we have represented in a typical normative display in Figure 4. The Year 9 distribution on the right of the scale is at a higher level than that for the Year 5 on the left. There is a substantial degree of overlap but only to an extent that is well known to those who have known and loved normative assessment for years. Average performances move up most of these scales by about half a standard deviation for each additional year of schooling - or is it year of life?

What is more important, from my point of view, is that performance can also be interpreted in terms of the scale itself and not just in terms of relative standing within populations. It is, for example, possible to define 'minimum adult competence' on such a scale and to use it to judge student performances. On this mathematics scale, after inspection of the items, we set our definition at around 35. About 94% of Year 9 students perform above this level and about 51% of Year 5. The definition is arguable, of course, but it is exposed because of its scale dependence. Others could develop a case for a different definition by appealing to the meaning of the scale above or below this point.

Though this presentation makes clear the separation of calibration and measurement processes, I should make clear that this is a statistical separation and need not be a separation in time. The data which reveal the difficulty of the items are the performances of individuals on the items. The data which reveal the level of student performances are the performances of the individuals on the various items. The processes are, however, separable in an important statistical sense. The estimation of the item parameters, in this case the item difficulties, is independent of which particular individuals' performances are used to calibrate the items. The measurement of the individuals' performances is independent of the particular items on which the measurement is made. In our case we measured Year 5 students with one set of items and with Year 9 with another. We could now take any mix of items from the two tests to construct a new test, still calibrated to the same scale, and use it to measure new students.

Monitoring Individual Student Growth

What I have displayed is a technique for representing population performances. Changes over time could be represented in the same way that the difference between year levels is represented here. Population movements across the whole distribution or in parts of it can be mapped and displayed.

In a similar way it is possible to map individual growth. Figure 5 shows the performance of an individual on a mathematics scale, with the items the student answered correctly coded to the left and items the student answered incorrectly coded to the right. The codes refer to components of the test, and thus of the curriculum, and give the teacher additional information about what the student knows and what the student does not know. The fact that there are virtually no items in the top left or in the bottom right indicates that the test works quite well for this student.

A year later the student's pattern had become as shown in Figure 6. There has been clear growth. The items which the student is now answering incorrectly are further up the scale, except for one which was probably a careless slip. Much which the student could not deal with a year before has now been mastered. What she is now in the process of mastering is fractions. What other students have been doing on the scale is irrelevant to an analysis of this student's growth, except to the extent to which one might wish to build expectations on the basis of the performance of others. The normative question can still be there. It is still legitimate, but it is not the central focus.

I am indebted to Dr Geoff Masters of ACER for this example.
For another student in the same class the pattern in the first year, shown in Figure 7, revealed a poorer fit with quite a number of relatively easy items answered incorrectly, particularly ones concerned with subtraction, despite some more difficult ones being answered correctly.

A year later, as shown in Figure 8, the problem is quite serious. There is little overall growth, despite the mastery of some more difficult skills during the year. The lack of overall growth is a consequence of a continued failure to deal appropriately with relatively easy problems of subtraction. The precise nature of the problems can be revealed by a more careful analysis of the particular tasks on which the student is now surprisingly, given the other things he can do, still scoring poorly.

The pattern and the detail to which it draws attention can be helpfully diagnostic.

More significantly, for the general point I am making, the measurement approach provides a map of an individual student's growth. Maps like these, sometimes called 'kidmaps' are being used as part of the Pupil Performance Profiling project being undertaken in Professor Sim’s Centre for Applied Research in Education in the National Institute of Education. Normative information need play no part in the understanding or in the representation of the student's growing mastery of a domain, but normative comparisons can be made where they are relevant.

### Reporting to Parents in New South Wales

For the recent work at the Australian Council for Educational Research (ACER) involving tests of whole year level populations in New South Wales in aspects of numeracy and literacy, the same approach to scale construction, with calibrations separated from measurement, is being used (see Masters et al., 1990). For this work, Lokan and her colleagues have developed some novel tests formats and some novel strategies for presenting information to parents. The physical form of the report as shown in Figure 9, presents a profile of performances on the five separate scales that are constructed, two for aspects of literacy and three for aspects of numeracy, indicating where on the scale the student's performances lies.

The information presented to parents also carries descriptions of the various parts of the scales in terms of four broad bands. The band descriptors for reading at Year 6 level, shown in Figure 10, show the manner in which the descriptors are written. They are generic but remain close enough to the actual items that fit on the scale with the band to avoid the blandness that comes with generality.

For the parents, the students and the teachers, the scale definition and the band descriptors, provide a criterion-referenced measurement of the student's performance. The descriptors indicate positively what the student can do and point to the skills that the student will next develop. By implication, the report carries normative information as well, of course, in the selection of the range of the scale that is displayed. It is obviously chosen as a range appropriate for the representation of the performances of students at that year level, so a parent can interpret a performance towards the top of the scale segment displayed as one among the best of those in the year.

### Building More Complex Scales

#### Partial Credit Models

As I indicated earlier, this approach to scale construction is not limited to tasks that can be scored right/wrong. The Americans have become so obsessed with multiple-choice items that some cynics say that a well-educated American is just someone who can recognise as true one statement in four.

The new approach can accommodate the kind of multi-point scale with which most teachers wish to give partial credit for partial knowledge. As an example, we calibrated scores on the 5-point scale used to mark short essays in the Victorian study of literacy and numeracy. With a right/wrong answer, there is only one point to calibrate - the point at which persons of that level of achievement have a 50:50 chance of answering correctly.

For marks on a 5-point scale, there are four locations to be calibrated. In Table 1, you can see that for the left hand column of the table, the levels of 28.5 and 57.8 (1 and 4) and 38.3 are at the top of the bands '5' and '2', and so on up to the next level between a '5' and '2'. An 'achievement' level is labelled 'report'.

<table>
<thead>
<tr>
<th>Mark</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>57.8</td>
</tr>
<tr>
<td>4</td>
<td>49.0</td>
</tr>
<tr>
<td>3</td>
<td>38.3</td>
</tr>
<tr>
<td>2</td>
<td>28.5</td>
</tr>
</tbody>
</table>

For the essay the 'Self', persons who score 53.0 had as mark '5'. For the essay the 'Self', persons above 62.4 would receive a '4'. The same rule applies to the calibrations on the next levels.

This is what can happen with essays - that it is easy to see that students on some than others, but this does not mean the same things as the level. If this approach has revealed, it is possible to do the calibration so that the same scale relates to the same scale range. The 'Self', students have written to the same university examinations papers, even though the level of this, though not necessarily the same level, is beginning to look quite different.

### Scaling to Bio-Dimensional Scales

Shifting the perspective, we see that from the explicit or implicit construction of a scale constructed...
see that for the writing task labelled 'report' (in the left hand column) persons with achievement levels of 28.5 are at the boundary between a '2' and a '1'; persons with an achievement level of 38.3 are at the boundary between a '3' and a '2', and so on up the scale. To be at the boundary between a '5' and a '4' a person needs to have an achievement level of 57.8 on the essay labelled 'report'.

Table 1: Difficulty Levels of Obtaining Additional Marks on Writing Tasks

<table>
<thead>
<tr>
<th>Description</th>
<th>Mark</th>
<th>Report</th>
<th>Self</th>
<th>Friend</th>
<th>Repair</th>
<th>Story</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>57.8</td>
<td>53.0</td>
<td>59.5</td>
<td>62.4</td>
<td>56.7</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>49.0</td>
<td>45.4</td>
<td>50.6</td>
<td>44.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>38.3</td>
<td>37.6</td>
<td>40.4</td>
<td>34.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>28.5</td>
<td>30.0</td>
<td>32.6</td>
<td>23.7</td>
<td></td>
</tr>
</tbody>
</table>

For the essay topic labelled 'Description of Self', persons with achievement levels as low as 53.0 had as much chance of scoring a '5' as a '4'. For the essay topic labelled 'Description of Repair', only those with achievement levels above 62.4 were more likely to obtain a '5' than a '4'. The same information is shown as calibrations on a scale in Figure 11.

This is what we have always known about essays - that it is harder to obtain a high score on some than on others. A '5' simply does not mean the same thing from topic to topic. With this approach to calibration not only is that fact revealed, it is taken into account in the calibration so the students are measured on the same scale regardless of the essay topic on which they have written. In public examinations, university examinations and in general school examinations in Australia no account is yet taken of this, though some interesting moves are beginning to take place.

**Scaling to Build Domain Understanding**

Shifting the primary focus in text development from the explication of individual differences to scale construction concentrates attention where it ought to be - on the nature and meaning of the scale. Delineating scales with which to explore and represent student growth can give expression to increasingly sophisticated understanding of the growth process. The broad band descriptors with which performances in the New South scales test are presented provide a coarse-grained analysis of five continuo along which students develop. This does not imply that the instructional strategies or, indeed, the instructional sequences, will fall automatically out of the performance sequences. It simply suggests that instructional decisions can be informed by this kind of analytic approach to measurement.

A forthcoming report (Adams, Doig and Rosier, 1991) gives a good example of this approach to scale development in a novel domain. In an investigation of science achievement in Victorian schools, Adams and his colleagues have included an investigation of children's scientific theories and constructed a scale with which to represent increasing complexity of the children's theoretical understanding. Science educators have explored the nature of naive theories for a number of years but have been restricted to a methodology that is intensively focused on individuals, using essentially Piagetian clinical interviews. Adams and Doig have developed a psychometric approach to this for the first time. They constructed a series of tasks such as the one shown in Figure 12 in which students are invited to provide an open-ended response to a cartoon presentation of the phenomenon for which they are to provide an account. In the case in Figure 12, the phenomenon is the appearance of water on the outside of a jug containing ice and water.

The set of categories shown in Figure 13 was developed a priori and then all responses were coded into one or other of the categories. Standard inter-judge analyses showed that a very high level of consistency was achieved in the classification of responses on this and each of the other items. The categories were presumed to be ordered in four levels as shown, the bottom two response types, 'uninterpretable responses' and 'thinking the liquid passed through the sides
of the jug' were combined into a single category scored '0'. Six percent of Year 5 students and 3% of Year 9 students said the water came through the sides of the jug. The other responses were scored '1' to '4' as shown. Responses in the category scored '4' indicated that condensation occurs when the air temperature is decreased and that the liquid on the outside of the jug came from the atmosphere. Of the Year 5 students, 2% gave answers of this type. At Year 6, 16% of students did.

The responses to this and to all the other items were then calibrated onto a scale in essentially the fashion we used with the essays in the earlier literacy study. For the condensation item, the scale values for the thresholds between categories are shown in Figure 14. On the numerical scale that Adams and his colleagues used, persons whose scientific understanding located them at 57.3 were equally likely to give answers in the category scored '0' (an uninterpretable response or saying the liquid came through the jug) or to give an answer in the category scored '1' (saying the liquid outside came from the liquid inside). The threshold between category 1 and category 2 was at 58.9 and so on. Those whose level of understanding placed them above 62.3 on the scale were most likely to have a full understanding of the condensation process and to provide an answer that revealed this.

There were nine items like the one on condensation, all built around the theme "The day we cooked pancakes". They all tapped students' scientific beliefs about the structure of matter so it was possible to calibrate them together to construct a 'structure of matter' scale. That scale is represented in Figure 15 where the descriptors are more generic than the particular category descriptions for any of the tasks.

Those persons who were measured at levels above 64 on this scale generally had a clear notion of chemical reactions. Those between 62 and 64 tend not to have the same level of understanding of chemical reactions but do hold a particular theory of matter (that is, believe that matter consists of small particles not visible to the eye). Those below 56 on this scale appeal to magic or give answers that are simply examples of the phenomenon for which they are asked for a description. The measurement of the population samples at Year 5 and Year 9 are shown in Figure 16. The differentiation between the two year levels is not as marked here as on numeracy (Figure 4) or literacy, and that is interesting in itself. The gender differences, displayed in the separate curves on each side of Figure 16, are not significant.

This kind of scale construction makes the analysis of students' scientific theories accessible to teachers and not just to those science educators whose intensive interviews with students built our first understandings of the importance of naive or everyday theories. These tests are not achievement tests nor are they probably best used in the assessment of outcomes. They may most helpfully be used by teachers prior to instruction in order to understand the beliefs and theoretical explanations their students already have developed. Teachers do not start with a tabula rasa, not even science teachers.

Conclusion

In this presentation I have sought to do two things. First, I have developed an argument that we need to focus on outcomes if we wish to monitor our educational efforts. In the process, I looked briefly at the early research that suggested schools do not make a difference and then at the more recent research which shows that, in general, schools do make a difference. They do add significantly to what students' backgrounds provide them with. And schools are not all the same. Some add more than others. Simplistic comparisons of schools on things such as examination results do not reveal these differences, however, because the examination results do not separate school and home effects.

Then I turned my attention from the monitoring task to the measurement task. In this stage, I wanted to show you how modern psychometric theory allows us to build and calibrate well-defined scales with which to measure student achievement.

The greatest benefit of the new approaches to scale development and measurement is that it gives us the capacity to separate tasks on a common scale from the other. With such a scale we can be forced to consider the specific nature of tasks and individual responses.

REFERENCE


Ainley, J. and Shute, V. (1979) A New Scales for Achievement Investigation, Oxford: NFER

Bowles and Gintis (1973) Schooling in Capitalist America, New York: Basic Books


it gives us the capacity to consider persons and tasks on a common scale and so to relate one to the other. With classical test theory, we were forced to consider tasks only in relation to other tasks and individuals only in relation to other individuals. Now we can express our theoretical views in the scales and use measurement to enhance our understanding of what an individual can do in terms of increasingly strong, generic descriptions of the underlying competences.

REFERENCES


Figure 1: Scale of person achievements and item difficulties
<table>
<thead>
<tr>
<th>Group</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.2</td>
<td>probability</td>
</tr>
<tr>
<td>47.3</td>
<td>symmetrical figure</td>
</tr>
<tr>
<td>46.0</td>
<td>24.56 + 0.4</td>
</tr>
<tr>
<td>44.7</td>
<td>bells ringing</td>
</tr>
<tr>
<td>42.8</td>
<td>fraction closest to 3/16</td>
</tr>
<tr>
<td>42.4</td>
<td>Joe's age</td>
</tr>
<tr>
<td>41.0</td>
<td>2 cubed by 3 squared</td>
</tr>
<tr>
<td>39.3</td>
<td>shaded area equals</td>
</tr>
<tr>
<td>38.0</td>
<td>1240 + 29 is approximately</td>
</tr>
<tr>
<td>36.7</td>
<td>Alison $10 more than Peter</td>
</tr>
<tr>
<td>35.6</td>
<td>number of teams of 7</td>
</tr>
<tr>
<td>34.3</td>
<td>38 x 9</td>
</tr>
<tr>
<td>32.6</td>
<td>which day calendar</td>
</tr>
<tr>
<td>31.2</td>
<td>difference in weights</td>
</tr>
<tr>
<td>27.9</td>
<td>6.6 - 2.4</td>
</tr>
<tr>
<td>24.9</td>
<td>125 * 5</td>
</tr>
</tbody>
</table>

Figure 2: Location of some items on Year 9 Mathematics test: Victorian Achievement Study 1988
Figure 3: Location of some items on Year 5 and 9 Mathematics test: Victorian Achievement Study 1988

Figure 4: Distribution of...
Figure 4: Distribution of student achievements on Mathematics scale: Victorian Achievement Study 1989
Figure 5: Jessica’s Year 3 Mathematics performance
Figure 6: Jessica’s Year 4 Mathematics performance
Figure 7: Tony’s Year 3 Mathematics performance

Figure 8: Tony’s
Figure 8: Tony's Year 4 Mathematics performance
Basic Skills Testing Programme 1990
REPORTS FOR PARENTS

Student: Aaron
Year 6
School:

Aspects of Literacy

<table>
<thead>
<tr>
<th>Band</th>
<th>Reading</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher</td>
<td>Literacy Scale</td>
<td>Higher</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>35</td>
<td>1</td>
</tr>
</tbody>
</table>

Aspects of Numeracy

<table>
<thead>
<tr>
<th>Band</th>
<th>Number</th>
<th>Measurement</th>
<th>Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher</td>
<td>Numeracy Scale</td>
<td>Higher</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>35</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

YOUR CHILD'S SKILL LEVELS (shown above as extract information from abstract or technical language:

- read an
- work on
- detect f

3
- understand
- understand
- detect f

2
- select a
- choose
- find an

1
- follow
- use TV
- find sin

Figure 9: Report to parents: NSW Basic Skills Testing Program 1990

Figure 10: Band Information about these scales is available from your child's school.

* Your child's skills also include Band 1 for skills up to this band. The full set of skill band descriptions is on the back of this page.
READING

Students in these bands are typically able to:

4 • read and sort out competing information;
• work out meanings of words from clues in the passage (a ‘float’ of $5 to $10)
• detect feelings not stated directly (humour)

3 • understanding less common meanings (protecting NOT defending from the wind)
• understanding meaning not stated directly (The mosquitoes gave us an enthusiastic reception);
• detect feelings in writing (tiredness, annoyance)

2 • select a correct piece of information from several given in the passage;
• choose words to sum up a piece of writing
• find and put together several pieces of information to reach a conclusion

1 • follow simple written instructions;
• use TV guide, newspaper index or weather forecast;
• find simple information in a short piece of writing

Figure 10: Band descriptors for Year 6 reading: NSW Basic Skills Testing Program 1990
Figure 11: Calibration of mark increments for essays: Victorian Achievements Study 1988
Figure 12: Condensation question: Victorian Science Achievements Study 1990

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>condensation, air temperature decreased, liquid comes from atmosphere</td>
</tr>
<tr>
<td>3</td>
<td>liquid comes from atmosphere, no mechanism or cause provided</td>
</tr>
<tr>
<td>2</td>
<td>condensation caused by coldness or ice, no indication it comes from atmosphere</td>
</tr>
<tr>
<td>1</td>
<td>liquid outside has come from liquid inside</td>
</tr>
<tr>
<td>0</td>
<td>liquid passed through sides of jug</td>
</tr>
<tr>
<td>0</td>
<td>uninterpretable responses</td>
</tr>
</tbody>
</table>

Figure 13: Categories for condensation responses: Victorian Science Achievement Study 1990
<table>
<thead>
<tr>
<th>SCALE</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>63</td>
<td>condensation, air temperature decreased, liquid comes from atmosphere</td>
</tr>
<tr>
<td>61</td>
<td>liquid comes from atmosphere, no mechanism or cause provided</td>
</tr>
<tr>
<td>60</td>
<td>condensation caused by coldness or ice, no indication it comes from atmosphere</td>
</tr>
<tr>
<td>58</td>
<td>liquid outside has come from liquid inside</td>
</tr>
<tr>
<td></td>
<td>liquid passed through sides of jug</td>
</tr>
<tr>
<td></td>
<td>uninterpretable responses</td>
</tr>
</tbody>
</table>

Figure 14: Category thresholds on condensation item: Victorian Science Achievement Study 1990

<table>
<thead>
<tr>
<th>SCALE</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>notion of chemical reactions</td>
</tr>
<tr>
<td>62</td>
<td>descriptions with particulate model of matter</td>
</tr>
<tr>
<td>60</td>
<td>aware of processes (dissolving, condensing, cooking)</td>
</tr>
<tr>
<td>58</td>
<td>involve changes not easily observed</td>
</tr>
<tr>
<td>56</td>
<td>identify components, but nothing beyond the observed</td>
</tr>
<tr>
<td></td>
<td>simple observations and definitions</td>
</tr>
<tr>
<td></td>
<td>examples or 'magic' as explanation</td>
</tr>
</tbody>
</table>

Figure 15: Structure of matter scale: Victorian Science Achievement Study 1990
Students have a notion of chemical reactions.

Students can describe simple processes such as evaporation, condensation, melting using the particulate model of matter.

Students are aware that during processes such as dissolving, condensing and cooking, changes occur that are not easily observable.

Students can identify key components but have little or no recognition of things or processes beyond the directly observable.

Students' responses rely upon simple observations and definitions.

Students' responses rely upon examples and 'magical' changes.

Students provide responses that are uninterpretable.

Figure 16: The structure of matter continuum: Victorian Science Achievement Study 1990
Work or Play in the Kindergarten?

Kathy Sylva

1 The Structure of Curriculum

All over the world there is searching discussion about how much structure is appropriate for children between the ages of 3 and 6. Many teachers still turn to the books of Susan Isaacs (1930, 1933) which give an exciting vision of how children's minds grow during play. Isaacs also describes the role of the nurturing adult who explores the world with children but never dominates or imposes her own view. Several decades later came the work of Marianne Parry (1975) whose contribution includes the distinction between two levels of play; 'that which educates', and 'that which fills time'.

The difference between the two levels of play is not easy to detect. Play can sometimes look good with the children actively involved, and yet lack the elements which contribute towards educational growth. In assessing the value of play situations, teachers should look for the security of the children within the groups and among known peers, for a continuity of care, interest and involvement from adults sharing the situation, for the appreciation of the importance of progression, extension and challenge for each child, for possibilities of a sense of achievement among the children and for ample time in which to explore a situation thoroughly.

At times, and especially in the United States, there have been calls for formal instruction in kindergartens. Advocates such as Bereiter and Engelmann (1966) claim that children, especially those from disadvantaged backgrounds, need the formal methods of an infant school. Bereiter stresses 'school readiness skills'. In programmes such as these teachers define the objectives for groups of children whose task is clearly to carry out adult-designed work. If the play curriculum is based on the theories of Jean Piaget, the training curriculum is based on the work of B. F. Skinner.

The conceptual framework usually employed when distinguishing curricular patterns in the kindergarten includes the notion of structure. This is often viewed as a continuum ranging from total adult dominance over children's activities to complete laisser-aller in which adults leave children to get on with it much on their own. This one-dimensional view of structure is very limited and it is suggested here that a description of structure in the pre-school requires two separate dimensions. David Weikart and his colleagues have put forward a two-axis model in which the amount of initiation of both child and adult is taken into account.

Figure 1 shows the two axes of initiation, horizontal for child and vertical for adult. The northeast quadrant (child responding, adult initiating) is rare in Britain where pre-school children are seldom instructed and trained in highly specific skills. The programmes in the southeast quadrant are not seen in Britain either because in this situation neither the child nor the adult takes the initiative. (An example might be immunisation programmes found in the developing countries where the work of the project is done through drugs and not by sustained interaction.)

This paper was presented as a Keynote Lecture at the “First Singapore Conference on Kindergarten”, 1990.
Figure 1: Pre-school education models

**TEACHER INITIATING**

**Progressive/Interactive**  
(sharing of initiation)  
*Eg: High/Scope Pre-school Curriculum*

**Behaviouristic**  
(children are trained)  
*Eg: Bereiter Engleman*

**CHILD INITIATING**  
**CHILD RESPONDING**

**Child-centred**  
(children thought to grow naturally)  
*Eg: traditional nursery school  
Susan Isaacs; Froebel*

**Custodial care with medical input**  
(nutritional or immunisation models)  
*Eg: some UNICEF programmes in the developing world  
Bereiter Engleman*

**TEACHER RESPONDING**

On the left hand side of Figure 1 are two different play curricula, one in which the adult responds to the child’s initiatives (southwest quadrant) and the other where adult and child share the initiative equally (northwest quadrant). Most likely a nursery in the Isaacs tradition would fall into the southwest corner and some more contemporary pre-school programmes (cf. High/Scope) fall in the northwest. In the northwest quadrant would also fall the curriculum discussed by Sylva (1984) as ‘guided play’.

Although diagrams such as this may make conceptually clear some issues in pre-school they cannot resolve questions concerning which curriculum is best. Which programmes will nurture the intellect? children’s autonomy? their self-esteem? These are questions which require empirical answers and some recent research will now be reported in hopes that it can shed light on the effects of nursery experience on children’s development. But first we must ask if pre-school attendance can have long-term effects.

2 Does pre-school attendance make a difference in children’s lives?

In the heady days of the early 70’s it was widely believed that pre-school attendance could raise children’s IQ scores and improve their chances for later attainment in reading and mathematics. Early research (see Mortimore and Blackstone, 1982) confirmed this optimism, but later results began to document the now familiar ‘wash-out’ phenomenon in which early gains disappeared by the age of eight or nine (Smith and James, 1975). These disappointing research studies led to a great deal of uncertainty at a time when the educational cutbacks began to bite. Committed nursery teachers remained convinced of the value of nursery provision but researchers and policy makers claimed that there was no hard evidence showing that early education made a positive contribution to later educational progress. If it did not make a difference, why use tax-payer’s money to support it?
In the early 1980's new reports came to Britain from America which documented some long-term effects of pre-school on children's educational success in late adolescence and employment in early adulthood. Whereas early studies had examined outcome in terms of IQ scores and attainment tests, the new breed of studies had examined outcome in terms of how children fared in real life rather than how they performed on tests. Why did the earlier studies show no clear-cut differences in favour of pre-school programmes? The reason is that earlier studies looked for short term effects on traditional tests; the more recent studies examined behavioural functioning in the school and the outside community - in other words how children fared in real life rather than how they performed on tests.

3 Do different pre-school curricula lead to different child outcomes?

Weikart is quick to point out that the studies which show the most startling effects are on "high quality" programmes. He means they had excellent, highly trained staff, frequent in-service and team planning, parental involvement, and a clear-cut curricular framework. Woodhead continues this theme (1984) when discussing the Lazar and Darlington review:

This does not mean that any pre-school programme is effective; far from it. This is not a license for laissez-faire. As we have noted, the Consortium projects, as research and demonstration projects, shared certain features in common, many of which were not subjected to experimental study but may well have been important to the effect. In brief these appear to be: careful planning and implementation; low child:staff ratios; high levels of professional support for project personnel; at least moderate levels of parental involvement in the programme; and cognitively focussed programme designed within a clear framework of educational methods and goals. Each of the various organisational patterns that make up existing pre-school provision in the UK falls short of the mark on one or a number of these features.

Another carefully designed study examined the effects on young people through age 15 of three different pre-school curricula: the High/Scope model using 'guided play', the traditional nursery-school model focussing on 'free play', and the Distar curriculum which is a formal, skills training programme. Schweinhart, Weikart, and Lerner (1986) studied 68 children from disadvantaged backgrounds in the United States. They were randomly assigned to the three programmes, attending them when they were 3 and 4 years old. At age 5 the IQ's of all children had increased significantly, but by age 8 only the Distar children retained the 'boost'. It looked for a time as though formal training programmes had the most lasting effects. At age 15, the children were visited once more for interviews and school assessment. These results were...
fascinating because, although there were no differences now in IQ scores amongst the three groups, it was found that children from the two play programmes were getting along better. For example, the children from the Distar programme engaged in twice as many delinquent acts as children from the two play programmes, including five times as many acts of violation of property. The Distar group also reported at interviews that they had relatively poor relations with their families. The samples are small, of course, but the pattern of the quantitative results is consistently in favour of children who attended the two play programmes.

Tables 2 to 4 show differences in school behaviours and attitudes of the three groups when they were 15 years of age. There are significant differences in favour of the play programmes in terms of parental perception of the child’s accomplishments, in sports and reading activities of the child, and also in student participation in school activities. Although it did not reach statistical significance, there is a clear trend showing that the Distar children had lower educational aspirations. Ten years after leaving the nursery the children allowed autonomy during the pre-school years were showing more social/educational engagement, positive attitudes towards schooling, and parental approval than children subjected to a more formal regime.

The findings with respect to delinquency are equally clear; they appear in Table 3. Because the sample is small many of the items do not reach statistical significance. Still it is clear that the play graduates committed fewer acts of property damage, drug abuse, and ‘status’ offences.

What is the message from this contrastive study of three curricula? It tells us that children who are encouraged to be independent, active learners while young will become more active and responsible citizens in adolescence. Some caution is needed in generalising these findings because they come from just one study in one American town and include fewer than 100 children. The study should be taken seriously, however, because the sample was carefully drawn, the outcome variables rigorously measured, and the quantitative analyses were scrupulously carried out.

4 ‘Free’ or ‘guided play’

The last study to be described is one conducted by Jowett and Sylva (1986) in Britain. They looked at two groups of children entering the reception class, one coming from local authority nursery classes and the other from playgroups. There were 45 children in each group and all came from working-class backgrounds. The children were matched on family structure, age, sex, and parental occupation. Did the nursery ‘graduates’ act different from the playgroup ‘graduates’? Before answering this question, it is important to ask which style of pre-school programmes the children attended before entering school. Although the precise pre-school experience of this particular group of 90 children was not studied, previous research in the same education authority (Sylva, Roy and Painter, 1980) showed differences between local playgroups and nursery classes. Whilst noting some overlap between the two kinds of provision (and both were assessed as being very good), the playgroups were seen to be more oriented to a ‘free play’ philosophy and the nursery classes to ‘guided play’. In this earlier observational study nurseries were more school-like in their expectations of children’s behaviour, although in no way did they carry out the training programmes based on Skinner’s theory.

Jowett and Sylva found that the children who had been to nursery engaged in more purposeful and creative play in the reception class than did the children from playgroup. During more formal sessions in the reception class, the nursery graduates spent more time completing workcards; in fact, one of their favoured activities was self-initiated writing. Furthermore, the nursery children spent less time watching than did children from playgroups, and seemed to settle more easily into the routine of the school.

Other observations were revealing as well. The nursery children were more likely than the playgroup children to initiate contacts with the teacher that were ‘learning oriented’ whereas the playgroup children tended to approach teachers for help or administrative queries. Again, the nursery children seemed more ready for independent functioning. One of the most
intriguing findings is shown in Tables 5. It summarises what children did during 60 hours of observation when they met an obstacle in work or play. If, for example, a child was unable to tie his shoelace, he might ask for help, give up, or persist and keep on trying. It can be seen that the nursery graduates were far more persistent and less likely to ask for assistance or give up.

Table 6 shows clearly that children who had attended the pre-school programme oriented to ‘guided play’ were more independent and ready for school than those who came from a ‘free play’ regime. They concentrated better, their play was richer, and they approached adults as resources for learning rather than sources of aid. Despite large differences in observed classroom behaviour, the performances of the two groups of children on the Boehm Test of Basic Concepts did not differ. (Recall that the children in the three curricula studies did not differ either on tests of academic achievement; rather, they differed on classroom behaviour and in ‘real life’ outcomes as young adults.)

5 How does early education affect children’s view of themselves and of their expectations to succeed?

I turn to the work of Carol Dweck (1988) on how children acquire beliefs and attitudes relating to their own talents and skills. Her views are supported by a large body of experimental work on children. Briefly, she demonstrates that children show either ‘Helpless’ or ‘Mastery’ patterns of behaviour when confronted with obstacles.

‘Helpless’ children avoid challenge and give up easily, whereas ‘Mastery’ oriented children persist in the face of obstacles; ‘Mastery’ oriented children seek new, challenging experiences. Furthermore, ‘Helpless’ children report negative feelings and views of themselves when they meet obstacles, whereas ‘Mastery’ children have positive views of their competence, even when confronting obstacles. This makes them task-oriented and resilient in the face of difficulties because they are confident and positive. The style of ‘Helpless’ or ‘Mastery’ oriented behaviour is not related to intelligence, rather it is a personality characteristic, a way of viewing oneself and one’s capacity to be effective in the world of things and people.

Helplessness is further related to differences in children’s goals. ‘Helpless’ children pursue goals of getting rewards from adults, whereas ‘Mastery’ oriented children pursue goals of learning and discovery on their own.

The two different kinds of goals may be dichotomised as ‘Performing well’ v ‘Learning something’; ‘Mastery’ oriented children would rather learn something new than perform an old skill well. Because they are oriented to learning and skill acquisition, ‘Mastery’ children do not feel like failures when things go wrong. Instead they concentrate on the task at hand. ‘Helpless’ children, by way of contrast, are erratic in strategy when difficulties appear and they engage in self-recrimination such as ‘It’s all my fault, I never do well’.

Because ‘Helpless’ children pursue goals of approval from others, failure or difficulties bring on negative moods when they anticipate disapproval from adults. ‘Mastery’ children care less about adult approval and so fear failure less.

Mastery/Helplessness is also related to children’s views of the relationships between effort and achievement. ‘Mastery’ children believe that effort pays off with the acquisition of new skill; ‘Helpless’ children avoid effort because they think it will get them nowhere. Children with goals for approval tend to believe extra effort is useless; they see success/failure as inherent in themselves. Thus an essential difference between the two kinds of children is the way they view intelligence as either fixed or incremental. The latter view of intelligence means that children believe that intelligence increases in small steps, it can increase as children persevere with flexibility. ‘Helpless’ children appear to inhabit a fixed world, where intelligence stays the same despite effort.

Kindergarten programmes must aim higher than instilling a few facts or rudimentary skills. They should be explicit about how they can instill in the child a positive belief in his own ability; encourage the child to be learning-oriented rather than performance-oriented, help the child acquire resources to recognise failure and for behaviours given but is, instead,

6 Summing up

We now know that lasting, positive effects on development. Rutter (1985):

“The long term benefits of early education are not only from what children learn through formal learning, on the basis of ‘task orientation’.

We know that the nursery programmes make a startling difference. Furthermore, we know that

REFERENCES


acquire resources for dealing with the stress of failure and for believing that intelligence is not given but is, instead, acquired through persistence.

6 Summing up

We now know that early education can have lasting, positive effects on children's development. Rutter summed it up in a review paper (1985):

_The long term educational benefits stem not from what children are specifically taught than from effects on children's attitudes to learning, on their self esteem, and on their task orientation._

We know that certain pre-school experiences make a startling difference in children's lives. Furthermore, we know that different programmes are associated with different 'outcomes' in terms of children's activities during the nursery years, in their behaviour on school entry, and even into the secondary school. It is suggested here that programmes which are founded on play rather than training may, in the long run, contribute more to children's sense of independence and esteem, and also to a more positive commitment to school and community activities. Lastly, there is some evidence that programmes of guided play, rather than free play, lead to greater independence in young children. Is this compatible with the work of early pioneers such as Susan Isaacs and more contemporary writers such as Marianne Parry? It seems that their early insights have been substantiated by a growing mound of empirical evidence; some adjustments and modifications to the original theories have been made, but in the main, the case of play appears to be proven.

REFERENCES


APPENDIX

Table 1: Major Findings at Age 19 from the High/Scope Study

<table>
<thead>
<tr>
<th>Category</th>
<th>Number Responding</th>
<th>Pre-school Group</th>
<th>No-Pre-school Group</th>
<th>p^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employed</td>
<td>121</td>
<td>59%</td>
<td>32%</td>
<td>.032</td>
</tr>
<tr>
<td>High school graduation (or its equivalent)</td>
<td>121</td>
<td>67%</td>
<td>49%</td>
<td>.034</td>
</tr>
<tr>
<td>College or vocational training</td>
<td>121</td>
<td>38%</td>
<td>21%</td>
<td>.029</td>
</tr>
<tr>
<td>Ever detained or arrested by police</td>
<td>121</td>
<td>31%</td>
<td>51%</td>
<td>.022</td>
</tr>
<tr>
<td>Females only: teen pregnancies, per 100</td>
<td>49</td>
<td>64</td>
<td>117</td>
<td>.084</td>
</tr>
<tr>
<td>Functional competence^c</td>
<td>109</td>
<td>24.6</td>
<td>21.8</td>
<td>.025</td>
</tr>
<tr>
<td>% of years in special education</td>
<td>112</td>
<td>16%</td>
<td>28%</td>
<td>.039</td>
</tr>
</tbody>
</table>

^aTotal n = 123
^bTwo-tailed p-values are presented if less than .10.
^cApplied and practical learning; maximum score of 40.

(From Berrueta-Clement, Schweinhart, Barnett, Epstein, and Weikart, 1984; the High/Scope curriculum is described in Hohmann, Banet and Weikart, 1979).
Table 2: Mean Number Delinquent Acts Reported at Age 15 by Children from Curriculum Groups

<table>
<thead>
<tr>
<th></th>
<th>Distar Curriculum (formal)</th>
<th>High/Scope (guided play)</th>
<th>Nursery (free play)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delinquency scale, 18 items (Total)</td>
<td>12.83</td>
<td>5.44</td>
<td>6.94</td>
<td>.04</td>
</tr>
<tr>
<td>Personal violence subscale</td>
<td>2.28</td>
<td>0.88</td>
<td>1.17</td>
<td>-</td>
</tr>
<tr>
<td>Property damage subscale</td>
<td>1.72</td>
<td>.28</td>
<td>.39</td>
<td>.04</td>
</tr>
<tr>
<td>Stealing subscale</td>
<td>3.06</td>
<td>1.72</td>
<td>2.22</td>
<td>-</td>
</tr>
<tr>
<td>Drug abuse subscale</td>
<td>3.17</td>
<td>1.06</td>
<td>1.89</td>
<td>.06</td>
</tr>
<tr>
<td>Status offences subscale</td>
<td>3.04</td>
<td>1.56</td>
<td>1.22</td>
<td>.04</td>
</tr>
</tbody>
</table>

From Schweinhart et al. (1986)

Table 3: Community Participation % Children Giving Each Answer

<table>
<thead>
<tr>
<th>Interview Questions</th>
<th>Distar Curriculum (formal)</th>
<th>High/Scope (play)</th>
<th>Nursery (play)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child participates in sports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Often</td>
<td>17%</td>
<td>50%</td>
<td>44%</td>
<td>.02</td>
</tr>
<tr>
<td>Sometimes</td>
<td>28%</td>
<td>44%</td>
<td>28%</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>56%</td>
<td>6%</td>
<td>28%</td>
<td></td>
</tr>
<tr>
<td>'In recent weeks have read':</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A book (N = 49)</td>
<td>31%</td>
<td>69%</td>
<td>59%</td>
<td>.09</td>
</tr>
<tr>
<td>A newspaper</td>
<td>67%</td>
<td>89%</td>
<td>72%</td>
<td></td>
</tr>
<tr>
<td>A magazine (N = 53)</td>
<td>44%</td>
<td>41%</td>
<td>72%</td>
<td></td>
</tr>
</tbody>
</table>

From Schweinhart et al. (1986)
### Table 4: Children's Self Perceptions (% Giving Each Answer)

<table>
<thead>
<tr>
<th>Interview Questions</th>
<th>Distar High/Nursery Curriculum Scope (formal)</th>
<th>High/Scope (play)</th>
<th>Nursery (play)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Getting along with family?’:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Great</td>
<td>33%</td>
<td>3%</td>
<td>28%</td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td>44%</td>
<td>67%</td>
<td>56%</td>
<td></td>
</tr>
<tr>
<td>Poorly</td>
<td>22%</td>
<td>0%</td>
<td>17%</td>
<td></td>
</tr>
<tr>
<td>‘Family feel about how you’re doing?’:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Great</td>
<td>0%</td>
<td>6%</td>
<td>6%</td>
<td>.03</td>
</tr>
<tr>
<td>All right</td>
<td>67%</td>
<td>94%</td>
<td>89%</td>
<td></td>
</tr>
<tr>
<td>Poorly</td>
<td>33%</td>
<td>0%</td>
<td>6%</td>
<td></td>
</tr>
</tbody>
</table>

From Schweinhart et al. (1986)

### Table 5: Children's Reactions To Difficulty (% of all children's responses)

<table>
<thead>
<tr>
<th></th>
<th>Asks for assistance</th>
<th>Gives up</th>
<th>Persists on own</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nursery centres (guided play)</td>
<td>17</td>
<td>1</td>
<td>82</td>
</tr>
<tr>
<td>Playgroup centres (free play)</td>
<td>36</td>
<td>22</td>
<td>42</td>
</tr>
</tbody>
</table>

All significant at p<.05

From Jowett and Sylva (1986)

### Table 6: Children's Language Initiations to Adults (% of all initiations)

<table>
<thead>
<tr>
<th></th>
<th>Requests for help or admin</th>
<th>Social positive</th>
<th>Learning oriented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nursery centres (guided play)</td>
<td>32</td>
<td>48</td>
<td>20</td>
</tr>
<tr>
<td>Playgroup (free play)</td>
<td>7</td>
<td>19</td>
<td>14</td>
</tr>
</tbody>
</table>

All significant at p<.05

From Jowett and Sylva (1986)
Monitoring Protolinguistic Development in a Bilingual Context

Peter Bodycott

Abstract

This study explores the protolinguistic development from birth of one child in a bilingual context. Protolinguistic data, non-verbal and socio-cultural influences were recorded using field notes, audio and video recordings. The data collection techniques were selected in order to develop upon and avoid criticisms levelled at similar, monolingual studies. Protolinguistic expressions were coded using the International Phonetic Alphabet. They were then analysed and classified according to linguistic function following models set by Halliday (1975) and Painter (1984). Central to the data analysis were the credibility checks undertaken. Each linguistic expression was subjected to “member checking” (Guba and Lincoln, 1985) by a team of English and Spanish speakers. Resultant data reflects a systematic development of linguistic options within six protolinguistic functions, strongly supporting the general sequential development as outlined by Halliday (1975) and Painter (1984). The development of lexical items from both the mother’s tongue (Spanish) and the father’s tongue (English) are identified and compared with these monolingual studies. Relationships were also identified between protolinguistic expressions and the language demonstrations provided by the parents.

Descriptors: Protolinguistic development, bilingual development

Introduction

One of the most remarkable accomplishments in a child’s development revolves around how, from birth the child is able to acquire, and make sense of the linguistic milieu in which she is immersed. This milieu of linguistic sound provides language demonstrations by the thousands, demonstrations of language in use. These demonstrations convey information not only about the sound system, but syntax and semantics. Yet, with apparent ease, the child within the first eighteen months of life is able to differentiate linguistic forms from this multitude of sounds, and begins to functionally communicate with outside members of the speech community. This is a remarkable achievement.

This ability to functionally communicate has developed from birth. Colwyn Trevarthen (1978) documented the beginnings of functional communication, showing how a newborn within the first few weeks of birth, engages in exchanges of attention. These exchanges are jointly constructed between the parents and child, and although they may contain no real content, in an adult sense, they do convey meaning, that is, they have a social function and serve a purpose. Exchanges of attention of the type described in Trevarthen are interpreted to be the very beginnings of language. Therefore an assumption that this study makes is that the function of language, from the beginning is to convey expressions of meaning. These meanings may be personal, for the child or shared constructions between the child and significant others.
Fascination with child language development is well documented in research and reviews, for example Butler (1985), and Hamers and Blanc (1990). In studying child language development in the early 1970s, Michael Halliday proposed a social-semantic perspective from which to view language development. Halliday (1975) traced the development in his son, Nigel’s language as it moved from a symbolic “protolinguistic” phase through its transition to a symbolic, “linguistic” phase.

Halliday developed upon the work of J.R. Firth (1968) specifically developing Firth’s interpretation of “category systems” into networks. Halliday (1970) claimed that a functional organization of language is reflected in the language, the meaning potential of language containing linguistic options that combine into “relatively independent networks” (Butler, 1985). These networks were seen to correspond with specific functions of early language.

Following Halliday’s work, comparable studies of children’s early language development in the home have been carried out by Painter (1984) and Oldenburg (1990). Qui Shijin (1985) also used a similar theoretical framework in observing Chinese children living in Shanghai; however different ages were covered within the group.

Protolanguage

Central to these studies and specifically relating to the present study is the Phase One, or “protolinguistic phase” of language development. This phase lasts from approximately nine to sixteen months. At its core, a set of seven functions which serve the interpretation of linguistic meanings of the very young child: the Instrumental or “I want” function; the Regulator, “Do as I tell you” and Interactional, “You and me” function; the very Personal, “Here I come” function; the inquisitive, Heuristic or “Tell me why” function; the Imaginative, “Let’s pretend” and the Informative, “I’ve something to tell you” functions.

In order to learn how to use language in its multitude of functions the young child creates a personal functional representation of the language, a symbolic set of linguistic expressions which serves various functions for the child. These protolinguistic functions, first suggested by Halliday (1975) guide identification of the content and meaning of what the child is learning to say. They represent a framework of potential meanings, which protolanguage serves. It is important to note that within each function there are ranges of linguistic options, meaning alternatives which the child uses (Painter, 1984). Likewise, as found in this study, expressions which constitute these options may also serve a range of functions, that is, a single expression may be used in a number of ways, serving different functions. Another important feature of language learning during this stage is the social nature of the learning environment. Linguistic expressions are shared and constructed through interactions with significant others. Interactions involve joint attention, action and communication, and from which the child learns the rules of social interaction and language within her culture (Bruner & Sherwood, 1981).

Significant others are those who can understand or “read” the child’s linguistic expressions or signs. For the child’s language system at this protolinguistic stage does not contain recognizable “vocabulary” or “grammar”, rather it is a system of signs. As the child draws upon and interacts with the oral language of her culture she is actively involved in constructing and reconstructing meanings. These meanings are shared through linguistic approximations of the language. The approximations are not at this stage recognizable as conventional language. Rather they are the child’s personal interpretation of the language, personal attempts at making sense of her world. The signs or linguistic symbols she uses are seen to be functional in different contexts. The child’s protolanguage provides a window through which we are able to observe how language develops.

The research on protolanguage has provided a great deal of information about how early linguistic meaning is organized and the subsequent transitional springboard toward conventional language. The studies however have focused on bilingual studies and unexplored an important question: how do children learn unexplored, and it is not clear how particular individuals develop language. This study described the protolinguistic stage of children’s language acquisition and provided a description of a theoretical framework in order to consider the stages of language development and expanded the understanding of bilingualism during language acquisition.

Two basic linguistic dichotomies are considered: monolingualism and bilingualism, and language is divided into two options—protolinguistic and conventional. The conventional language of the child is a system with rules such as syntactic or semantic alternation, the acquisition of a language, or from one language to another (Grosjean, 1982). Bilingual dichotomy is a system of language that is encountered, shared, and constructed by the child to particular points in time (Cummins, 1977). A dichotomy is a system of language that is felt, understood, and shared within the context and world of the child. The conventional languages can be either monolingual or bilingual. For the present study this was a monolingual English-Spanish-English bilingualism.

Another approach to the protolinguistic study and confusion among children’s context was found in this study. The approach to the study of two languages or the comparison of two languages. For example, if a child speaks both and consistently uses both languages, the child is a bilingual speaker. However, if the other language is used in a particular context, the foundation for bilingualism is not as well understood (Pavlov, 1913).
have focused on monolingual children. What happens when the child’s linguistic immersion contains two languages? Are similar functions developed, and in what order? It is these general questions which sparked the initial interest and organization for the present study.

**Bilingual Development**

Investigations into the “early” language acquisition of bilinguals found it to be a relatively unexplored area in the field of linguistics. Of particular interest however were reports which described the social context and the effects on bilingual acquisition. It was felt that any description of early language development must consider the strategies employed by the parents and extended family for they must influence the acquisition process.

Two basic strategies often adopted are **dichotomy** and **alternation**. In dichotomy, the language is divided and fixed on the basis of factors such as person, place, or time. In alternation, there is a spontaneous switching from one language to the other. Studies by Grosjean (1982) and Zierer (1977) report on bilingual dichotomy by age and problems encountered, such as the need to restrict access to particular playmates at certain ages (Zierer, 1977). A dichotic approach on age or time, it was felt, would provide an unnatural learning context and would require both parents to speak both languages. In the context of the present study this was impossible as the father was monolingual English and the mother bilingual Spanish-English.

Another approach used to avoid language confusion and provide a more natural learning context was first suggested by Grammont (1902). The approach involved a one person-one language or simultaneous immersion in two languages. For example, one parent constantly and consistently addresses the child in one language whilst the other does the same in the other language. This approach has provided the foundation for several major studies in the area of bilingualism, specifically the works of Ronjat (1913), Pavlovitch (1920) and Leopold (1939-1949). Ronjat (1913) reported on the development of French and German in his own son, “Louis”, finding the language of his son in the beginning to be “unilingual”. Pavlovitch (1920) recorded the development of French and Serbian in his son, “Douchan”. Leopold (1939-1949) reported on one of the first truly longitudinal “systematic” studies of bilingual acquisition. Studying the simultaneous acquisition of English and German in his daughter, “Hildegard”, from birth, her early language forms were characterized by the “free mixing” of both languages. Ronjat and Leopold both concluded that there was no obvious delay in bilingual development or evidence of language interference.

More recent studies which have used the one language-one person approach include Celce-Murcia (1978), Garcia (1983a,b), Huerta-Marcias (1983), Oksaar (1983), and Saunders (1980, 1983). See also the reviews of Genesee (1989), Hamers and Blanc (1990) Redlinger (1979) and Schinke-Llana (1989). Whilst this one person-one language approach may have disadvantages, Zierer (1977) expresses concern over the considerable mental exertion required by the child, and the affect simultaneous immersion may have on the family. In the present study, given the languages of the parents, a simultaneous approach was seen to be the natural choice.

The bilingual studies of Carrow (1971), Garcia (1983a,b), and Keller (1976) also relate to the present in that they specifically examine early Spanish and English bilingual development, however they differ in two areas:

(i) they deal specifically with bilingual development in older pre-school children, that is, they focus upon children from 13 months, and

(ii) they fail to explore relationships that exist between the child’s early speech, meaning and the child’s environment.

There was then seen to be a need for a closer examination of early bilingual language acquisition. This would provide insights into the developing functional structure of a bilingual child’s protolanguage as it moves toward language differentiation or language separation.
Secondly, by using an established monolingual classification model bilingual data can be compared.

Context of the Study

In summary, the following general guidelines were adopted in order to create and foster simultaneous bilingual development:

- The mother only spoke Spanish with the child, and the father English. This was to promote language differentiation by person, thus avoiding the possibility of language interference. In the presence of the child the parents would converse in English with one another, the mother switching to Spanish when addressing the child.
- Spanish and English speakers were encouraged, when initially addressing the child, to speak in their native language and to avoid switching. This provided a language identity and sought to avoid confusing the child and upsetting her established differentiation system (Grosjean, 1982; and Saunders, 1980).
- The parents sought to expose the child to totally Spanish settings. As the mother returned to work when the child was 10 months old, the English speaking grandmother babysat during the mother’s absence. Whenever possible the Spanish speaking grandparents were sought out to provide a totally Spanish context. In addition the child was frequently exposed to cultural events in Spanish.
- The child was provided with books, tapes and videos in both languages. Objects around the house were labelled in both Spanish and English, literacy development being a separate yet related part of the study.

Method

As parent participants researching language it was decided to collect data using a variety of techniques. Field notes in journal form were kept on a daily basis continually from birth while audio and video-recordings of the subject

were made at 4-6 weekly intervals. This enabled comparisons to be made with the monolingual studies of Halliday (1975) and Painter (1984) yet avoiding criticisms specifically levelled at Halliday. Dore (1977) questioned Halliday’s “virtuosity as a linguist” in the recording of his data by hand, and Butler (1985) suggested that video recording may have been useful in clearing up suggestions of subjectivity and transcriber error.

This multi-dimensional approach enabled the recording of authentic linguistic expressions from different communicative contexts and facilitated checks on the credibility of the data. To further ensure the encompassment of naturally occurring data the techniques were often used simultaneously. That is, as video or audio recordings were being made, field notes were also being kept of the same experience.

Fieldnotes

A major advantage of fieldnotes was that other participants could record while the parents were involved with the child. The fieldnotes contained transcripts of the linguistic expressions being used, notes on the environmental features, participants, objects involved and interpretative comments. They also provided an essential record of linguistic development that occurred between video and or audio taping sessions. All the linguistic samples of data collected in the fieldnotes were supplemented and cross-checked with audio and videotape recorded data.

Audio Recording

Following the model set by Painter (1984) audio-tape recordings were made during two 30-45 minute sessions at 4-6 week intervals. This interval was seen to allow enough time for significant changes in the child’s language development to occur whilst not placing the researcher at the mercy of random non-occurrences (Halliday, 1975). Transcriptions of the recordings followed the taping. Where recording involved the participation of other people, for example Zoe’s (the subject) English speaking grandmother, the person would sit with the researcher and assist in the transcription process.

Video Recording

Videotape recordings were made at 4-6 week intervals and lasted an average of 30-45 minutes. The audio unit was used to record linguistic data. Transcripts of the recordings were closely compared to and cross-verified with fieldnotes of the same events.

It should be noted that the linguistic interactions were non-consciously or consciously performed. The parents were informed of what they should say, not what they thought or demonstrated.

Data Analysis

Analysis of the linguistic data from audio-tracings and video recordings involved knowledge of protolinguistics as defined by Halliday (1975). Transcripts of the recordings were analyzed using the ‘functions’ framework and the inclusion rules provided by Dore (1977) as follows:

1 the protolinguistic functions produced
2 the expression of the function
3 meaning content of their function

The interpretation of protolanguage and functions outlined helped in the understanding of the linguistic expressions of the child.
Video Recording

Videotape recording began and continued at 4-6 week intervals from birth. Each taping session lasted an average of 50 minutes. A porta-pak video unit was used which enabled the recording of linguistic episodes outside the home. Transcripts of the videotaped sessions were closely compared with audiotapes and or the fieldnotes of the same session.

It should be noted that throughout the study, linguistic interactions with Zoe were never forced or consciously set up to test linguistic performance. As she was a first child, both parents were involved in providing intuitively what they thought to be appropriate language demonstrations.

Data Analysis

Analysis of the data occurred following the audio and video recording sessions whilst the intuitive knowledge of the language remained fresh. Protolinguistic expressions were coded according to the International Phonetic Alphabet (I.P.A.). A summary flow chart of the data collection and analysis procedure can be found in Figure 1.

As indicated in the flow chart, data were collected from a variety of communicative contexts. These ranged from early morning diaper changing sessions, free play with the video camera to birthday parties and barbecues that involved larger numbers of people. Analysis of the data transcripts as indicated occurred at 4-6 week intervals. Data were included and analyzed using the following three general inclusion rules:

1. the protolinguistic expressions were produced in different situations or communicative events, or
2. the expressions were repeated in similar communicative events, and/or
3. meaning could be interpreted in terms of their functional social context of use.

The interpreted social uses of the child's protolanguage were classified according to the functions outlined by Halliday (1975). The expressions were grouped into networks of linguistic options within each function. The structure and social function of each expression was determined largely through the use of audio and more demonstratively through the video data. Coupling the acquired data using the multidimensional approaches as outlined allowed the expression, and in most cases, cause and effects to be seen as well as heard. This was seen to be essential in any attempt in which functional analysis is being attempted. It is a tribute to Halliday and Painter that they could interpret and record linguistic function through pencil, paper and audio records alone.

Following the classification process, credibility checks were made of the data. Credibility checks were undertaken by a "member check team" (Guba and Lincoln, 1985). The member check team consisted of both parents, the English and Spanish grandmothers and a female relative. As participant observers the parents had the most intimate knowledge of the subject's language. Next to the parents, the grandmothers were possibly closest to the child. They provided an experienced neutrality having each raised several children to be English or Spanish speakers. The final member had herself raised two Spanish-English-German trilinguals.

These member check meetings served to check the author's initial interpretation and transcription of the subject's expressions. Where disagreements over interpretations existed, consensus was determined through a constant replaying of the recorded data. A technique that assisted in the analysis process involved the "fastidious oral reproduction" of the expression, that is, to code the expression (I.P.A.) each person would reproduce the expression, making efforts to match the intonation and sound.

In addition to member checking a "peer debriefing" session was undertaken which involved presenting the collated data for critical appraisal to knowledgeable but non-involved persons (Guba and Lincoln, 1985). The purpose of this peer debriefing was to ensure that the author maintained objectivity. This was achieved through probing and exploring the meanings that had been developed and which formed the basis for his interpretations. The session also provided an opportunity for the
Peer debriefing also forced the author to demonstrate an explicit data "audit trail" (Guba and Lincoln, 1985). That is, peers retraced back through the data checking the trustworthiness of the data analysis process. Credibility measures of this type have laid firm and credible theoretical foundations for naturalistic evaluation.

Results
Diagrammatic development was developed. Two general summaries of protolinguistic functions were developed. Figure 2.

<table>
<thead>
<tr>
<th>Instrument Options</th>
<th>1 linguistic options</th>
<th>2 expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reg. Options</td>
<td>2 linguistic options</td>
<td>4 expressions</td>
</tr>
<tr>
<td>Interactional Options</td>
<td>Personal Options</td>
<td>Imaginative Options</td>
</tr>
</tbody>
</table>
Results
Diagrammatic outlines of protolinguistic development from 8-16 1/2 months were developed. Two types of outline were produced: a general summary of development and specific protolinguistic networks at six weekly intervals beginning from birth. Figure 2 provides an example of protolinguistic development at 12 months of age. Summaries of this type provide a cumulative overview of protolinguistic development up to and including the specific stage being focused upon. Each included the identified linguistic functions and the total number of linguistic options and expressions used within each function.

Protolanguage: Birth to 1 Month
- Various random grunts and squeals conveyed needs and distress
- High pitched squeals received more urgent attention
- At 0.8 months certain sounds were favoured

Protolanguage: 1 to 8 Months
- Data obtained from daily routines
- Many demands were lengthy
- Continued interest in books and specific toys
- Objects inevitably ended up in mouth
- Expressions interpretable within three functions

Protolanguage: 10 1/2 months
- Function application highlighted by intonation
- Bilingual understanding in different communicative events
- Expressions interpretable within 3 functions

Protolanguage: 12 months
- Emergent regulatory function
- First words - [ma:] and [d d]
- Cross function use of expression
- Intonation determinant of function differentiation
- Expressions interpretable within 4 functions

Figure 2: Summary of Protolinguistic Development at 12 months
More specific information on the functions used, coded linguistic expressions and adult interpretations of their meaning were recorded in table form. See Table 1 for an example of Zoe’s protolanguage at 13 1/2 months. The tables included details of the linguistic options, coded expressions, tone and interpretations of meaning within each function.

Table 1: Protolanguage at 12 Months

<table>
<thead>
<tr>
<th>Function</th>
<th>Options</th>
<th>Expression</th>
<th>Tone</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrumental</td>
<td>General demand (1)</td>
<td>5: (one or both arms outstretched)</td>
<td>mid high</td>
<td>I want that</td>
</tr>
<tr>
<td></td>
<td>Specific demand (2)</td>
<td>action 2a bath 2b drink 2c</td>
<td>mid rise</td>
<td>I want help</td>
</tr>
<tr>
<td>Regulatory</td>
<td>General request (1)</td>
<td>initiate 1a goga; gs;</td>
<td>mid</td>
<td>Will you give me that?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>initiate 1b awa;</td>
<td>long mid</td>
<td>Stop that</td>
</tr>
<tr>
<td></td>
<td></td>
<td>initiated</td>
<td>mid high</td>
<td>No I don’t want that</td>
</tr>
<tr>
<td></td>
<td>Specific request (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>With objects (1)</td>
<td>general 1a running/blending sounds together</td>
<td>low rise/fall</td>
<td>Here you are...</td>
</tr>
<tr>
<td></td>
<td></td>
<td>initiate 1b</td>
<td></td>
<td>I’m telling you about this</td>
</tr>
<tr>
<td></td>
<td>Recognition (2)</td>
<td>personal greetings father</td>
<td>mid</td>
<td>daddy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mother</td>
<td>mid low</td>
<td>mummy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>grandma (E)</td>
<td>mid low/fall</td>
<td>nana</td>
</tr>
<tr>
<td></td>
<td></td>
<td>grandma (S)</td>
<td>mid</td>
<td>abuela</td>
</tr>
<tr>
<td>Interactional</td>
<td>Response to (3)</td>
<td>action 3a Authors</td>
<td>low</td>
<td>Thank you</td>
</tr>
<tr>
<td></td>
<td></td>
<td>conversation 3b various</td>
<td>blends of sounds</td>
<td>Let’s just talk</td>
</tr>
<tr>
<td></td>
<td>Feelings (1)</td>
<td>relaxing 1a</td>
<td>mid low/fall</td>
<td>I’m content</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pleasure 1b</td>
<td>mid low</td>
<td>I like that</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unhappy 1c</td>
<td>low rise</td>
<td>I’m not happy</td>
</tr>
<tr>
<td></td>
<td>Active involvement (2)</td>
<td>toys 2a</td>
<td>mid low</td>
<td>This is hard work</td>
</tr>
<tr>
<td></td>
<td></td>
<td>eating 2b m:</td>
<td>mid</td>
<td>Yum, I like the taste of that</td>
</tr>
<tr>
<td></td>
<td>Personal Interest (3)</td>
<td>general 3a</td>
<td>mid</td>
<td>This is interesting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>specific 3b</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8a8b8c b: papi guapo babá wáwá</td>
<td>mid low</td>
<td>dog</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mid low</td>
<td>cato</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mid low</td>
<td>baby, guagua</td>
</tr>
<tr>
<td>Imaginative</td>
<td>Pretend play(1)</td>
<td>clicking sound</td>
<td>mid</td>
<td>I’m riding a horse</td>
</tr>
<tr>
<td></td>
<td>Singing (2)</td>
<td>bebébabá</td>
<td>wide</td>
<td>I’m singing</td>
</tr>
</tbody>
</table>

Discussion

Protolanguage

Analysis revealed expressions mostly non-verbal, e.g., tongue-father to initiate play. The beginnings of symbolic participation occurred at 8 months and included the Instrumental functions. By contrast, Nigel and Paint began producing their first vocal sounds at 13 1/2 months. However Hal delayed and did not begin vocal expression until 15 months.

During this stage of development, a large amount of play was observed. This included sharing food and manipulating objects. As Halliday has outlined, this is a stage during which children are developing the essential symbolic systems. It is thus not surprising to find that the essential features of protolanguage are present.

Throughout the period, children were observed to use both verbal and non-verbal forms of expression and were able to interpret the meaning of their own and other children’s expressions. This indicates that the children were developing an understanding of the symbolic nature of language, which was evidenced in their ability to interpret and extend the meaning of protolanguage. This was evident in the children’s non-verbal interaction, where they were able to share and interpret the symbolic systems.

The observation of the children’s interaction and the use of protolanguage indicates that these children were well on the way to developing a symbolic understanding of language. The children’s development of protolanguage was evident in their ability to use the symbolic systems in a meaningful way, and their ability to interpret and extend the meaning of protolanguage. This indicates that the children were developing an understanding of the symbolic nature of language, which was evident in their ability to share and interpret the symbolic systems.
Discussion

Protolinguistic Sound System

Analysis revealed that random sounds and expressions moved toward recognisable mother tongue-father tongue language in the first year. The beginnings of more organized expressions occurred at 8 months and were grouped within the Instrumental, Interactional and Personal functions. By comparison, Halliday’s (1975) son, Nigel and Painter’s (1984) son, Hal, developed their first vocal expressions around this period. However Hal did not develop an Instrumental expression until around 9-10½ months.

During this stage noticeable changes in Zoe’s behavioural dependence were observed. She spent an increasing amount of time on her own playing with objects which interested her. She was developing the ability to support herself and manipulate objects such as spoonfuls of food. As Halliday (1990) suggests, the child has begun to develop the freedom of space-time. At the same time she is achieving the semiotic freedom of construing linguistic meanings into systems. It is this choice of meaning which is the essential feature or characteristic of protolanguage.

Throughout the study there were many expressions which, upon member checking, failed to fit exactly into the specific functions as outlined in Painter (1984). Painter’s functional descriptions acted only as a guide and all members of the team preferred to use their own or consensual interpretations and labels before consulting Painter. These differences however were often more a matter of semantic interpretation rather than alternative function development. For example, Zoe used expressions which were interpreted to mean “I’m telling you so listen”. These expressions were listed under the Regulatory and Interactional functions as outlined in Painter to maintain comparability.

There were noticeable differences in the development of a regulating function between Zoe and Hal. Zoe began Regulatory type behaviour and expressions in the period 10½-12 months whilst Hal only used non-verbal behaviour to regulate a person’s behaviour. Zoe in this period also began to walk. Along with this new-found independence she was observed spending much of her time attempting to regulate or interact with others.

The size of Zoe’s protolanguage, range of functions and meaning options increased steadily throughout the study. By comparison at 13½ months, Nigel had 32 different signs, Hal had 17 and Zoe 26. A feature of Zoe’s language from 13½-15 months was her use of phonemic expressions that served “different functions” and which were clear examples of Spanish and English words in the making. For example Zoe, with a book in hand, would walk towards her seated father turn around and back onto his lap. “bu” she would say, “bu”. If he did not respond she would headbutt his chest repeating “bu” until he began reading. The expression “bu” served to “regulate” her father’s behaviour. Similarly, Zoe would approach her grandparents or visitors. Holding out a book she would begin her interaction with a combination of unclear sounds and end with a firm “bu”. If unsuccessful in her attempts at gaining attention she would move onto another individual or small group and try again. Exchanges of this type exemplify the interactional or “Let’s read together” function of her protolanguage. The expression was also used while playing alone in her bedroom. Sorting through her toy box, as books were encountered she provided her personal “bu” label. A similar expression also used in this way, is “bu”, for example:

017 Z: “ba” (mid-tone)
018 F: “What is it Zoe?”
019 M: “You’re a silly billy Pete.”
   “Zoe, Queres un bano, cierto?”
   (You want a bath right?)
020 Z: “....(squeals).... ba” (looking at father and attempting to hurdle the tub)
021 Z: “ba” (low-low tone-looking at mother)
   “ba” (rpt)
022 M: “Eso es lo que queres.”
   (That’s what you want.)

The cross functional use of expressions illustrates the developing control Zoe had over her language at this stage. With the use of a single expression, as indicated, she was able to maintain contact with significant people in her
environment (interactional); and exercise control over her father’s behaviour (regulatory).

The phonological similarities between the labels “ba” for “bâno” (Spanish) and “bath” (English) also indicates the beginnings of interactional mixing. That is, not only was she developing the ability to functionally interact with her mother and regulate her father’s behaviour (regulatory), but Zoe could comprehend both languages in varying contexts, favouring a “receptive” mode as opposed to a “productive” mode of bilingual language differentiation. As she matured, she began to comprehend instructions from both parents demonstrating a more “receptive” knowledge of her two languages. At 12-13½ months she was moving toward the “production” of language specific to both languages. Although Zoe had not achieved the status of a “complete” bilingual, her language at 16½ months, as a whole, can be described as “incipient”, that is, developing toward a “complete” bilingual.

At 16½ months, transition from the protolinguistic phase was beginning. Similarities between the protolanguages of Zoe, Hal and Nigel at this stage are summarized below:

- Zoe and Nigel displayed clear distinctions between initiating and responding throughout their linguistic networks. They also used a “sharing” sign, which invited others to participate. Nigel’s was a “Shared regret”, Zoe’s was a “Shared hurt”
- Zoe and Hal both used non-verbal gestures when requesting names.
- Verbal request for names or which drew attention to objects within the environment: Nigel: “a::da” - you say what it is.
- Zoe: “Luk” - with non-verbal gestures that focussed the adult’s attention on an object, and who then proceeded to give the name.
- Zoe differed from both Hal and Nigel by using an instrumental demand expression for assistance “ak”. This was interpreted as a phonological approximation for the English word “stuck”.
- All three children used expressions which identified and maintained contact with specific adults. Zoe’s range of options of this type was considerably more than Hal and Nigel. This could in part be due to the changes in the socio-cultural context used by the parents in their effort to provide a balanced bilingual input. Zoe also developed phonological variants within these greetings and like Hal changed the intonation and stress according to whether the interaction was general or specific. Unlike Hal Zoe’s “general” greeting “bano” and “bano” occurred at 12-13½ months.

Personal signs were always accompanied with object labels: “bt.” Interactional recognition of Zoe’s first personal sign “bt.” came midway through her interest in books. Marjorie Painter (1984) interpreted the use of “bt.” as feeling general knowledge of books.

The first clear indication for Zoe occurred at 16½ months. Nigel’s first positive use of a personal sign occurred at 9-10½ months by using his expression “bt.” to label a specific book. Zoe’s first English approximation of “bt.”, which was a non-verbal expression applied to books, is interpreted as an indication that she was developing toward a “complete” bilingual, that is, developing toward a “productive” mode of bilingual usage. Zoe’s first personal sign appeared at 16½ months and was “gato” (cat).

This name (noun) reflects function and type of referential situation and the child’s experience of the expression developing in an almost simultaneous and verbal pointing and label for the object given. Zoe’s clapping developed from the use of “bt.” and “gato”.

10½ months, Zoe could comprehend both languages in varying contexts, favouring a “receptive” mode as opposed to a “productive” mode of bilingual language differentiation. As she matured, she began to comprehend instructions from both parents demonstrating a more “receptive” knowledge of her two languages. At 12-13½ months she was moving toward the “production” of language specific to both languages. Although Zoe had not achieved the status of a “complete” bilingual, her language at 16½ months, as a whole, can be described as “incipient”, that is, developing toward a “complete” bilingual.

At 16½ months, transition from the protolinguistic phase was beginning. Similarities between the protolanguages of Zoe, Hal and Nigel at this stage are summarized below:

- Zoe and Nigel displayed clear distinctions between initiating and responding throughout their linguistic networks. They also used a “sharing” sign, which invited others to participate. Nigel’s was a “Shared regret”, Zoe’s was a “Shared hurt”
- Zoe and Hal both used non-verbal gestures when requesting names.
- Verbal request for names or which drew attention to objects within the environment: Nigel: “a::da” - you say what it is.
- Zoe: “Luk” - with non-verbal gestures that focussed the adult’s attention on an object, and who then proceeded to give the name.
- Zoe differed from both Hal and Nigel by using an instrumental demand expression for assistance “ak”. This was interpreted as a phonological approximation for the English word “stuck”.
- All three children used expressions which identified and maintained contact with specific adults. Zoe’s range of options of this type was considerably more than Hal and Nigel. This could in part be due to the changes in the socio-cultural context used by the parents in their effort to provide a balanced bilingual input. Zoe also developed phonological variants within these greetings and like Hal changed the intonation and stress according to whether the interaction was general or specific. Unlike Hal Zoe’s “general” greeting “bano” and “bano” occurred at 12-13½ months.

Personal signs were always accompanied with object labels: “bt.” Interactional recognition of Zoe’s first personal sign “bt.” came midway through her interest in books. Marjorie Painter (1984) interpreted the use of “bt.” as feeling general knowledge of books.

The first clear indication for Zoe occurred at 16½ months. Nigel’s first positive use of a personal sign occurred at 9-10½ months by using his expression “bt.” to label a specific book. Zoe’s first English approximation of “bt.”, which was a non-verbal expression applied to books, is interpreted as an indication that she was developing toward a “complete” bilingual, that is, developing toward a “productive” mode of bilingual usage. Zoe’s first personal sign appeared at 16½ months and was “gato” (cat).

This name (noun) reflects function and type of referential situation and the child’s experience of the expression developing in an almost simultaneous and verbal pointing and label for the object given. Zoe’s clapping developed from the use of “bt.” and “gato”.
was general or more intimately specific. Unlike Hal and Nigel, Zoe dropped her "general" greeting expression in the period 12-13½ months, thereby no non verbal gestures when or if acknowledging unfamiliar people.

Personal sign expressions for Zoe, like Hal always accompanied first hand experience with objects in both the Personal and Interactional contexts. Particularly the recognition of objects in pictures which, for Hal came much later. Zoe's familiarity and interest in books, from 0-5 months resulted in the use of specific linguistic expressions relating to these first hand experiences with books. Many of these expressions were interpreted to be characteristic of Personal feeling generated through the daily sharing of books.

The first clear approximation of adult words for Zoe occurred in the period 10½ to 12 months. Nigel's first possible imitation of an adult word occurred at 9-10½ months while Hal's first word appeared during the 13½ to 15 month period. Zoe's first English word "mum" was seen as an approximation for the English word "mum." Painter (1984) distinguishes these types of expressions apart from "true names," as she believes they would appear to be linguistic signs that have a complex range of meanings. In Zoe's case, however, the closeness of the approximation to adult form, the contextual use and tonal differentiation suggests it to be a true name. This was supported by the recognition and interactional naming in pictorial form of the person which supports McShane's (1980) suggestion of an interactional route for understanding of names.

Zoe's first clearly identifiable Spanish word was "gato" (cat) in the 12-13½ month period. This name (noun label) occurred in the personal function and was repeated immediately in situations removed from the first hand experience, ie, with pictures of the object. The expression developed during an extended stay in an almost totally Spanish context. A non verbal pointing gesture led to the name being given. Zoe's classification of "gato" like objects developed from the very general to the very specific. Interestingly, Hal's first word, other than caregiver signs, also related to a "cat" label.

Zoe's development of lexical items such as "gato" support views of language development held by Bruner (1983) and Vygotsky (1962); that sociocultural experiences form the basis from which meaning and language develop. Similarly the use of language by others in the child's social environment plays an important part in stimulating actions and linguistic value placement (Vygotsky, 1962). Language used by an adult to "draw attention" to objects influences the child's perception of the world. This in turn "encourages alertness" and "stimulates curiosity and interest" (Tough, 1976). In "Thought and Language" (1962), Vygotsky illustrates how each time a particular word, ie, "gato" is used, the child's attention is drawn to another instance of the concept. In time, the word comes to represent a general idea that has developed from many of these experiences.

In summary Zoe used the expression "gato," which at first was applied to all short hairy animals. This generalization developed as she gradually began to focus upon the specific features of "gatos," distinguishing them from other "developing" concepts or objects, for example dogs and koalas.

Cross analysis of networks also revealed the natural phonological development of specific lexical items. For example, the word "stuck":

(æ) - 10 months
(ə) - 12 months
(A) - 13 months, 15 months
(ak) - 16½ months

Other expressions however were seen to begin as a direct result of repeated attempts at teaching by the English grandmother.

(aen) - 13½ months - nana
(aen) - 16½ months - nana

Zoe appeared to display a language preference in the development of specific lexical items. Whether this linguistic preference relates to Slobin's (1973) notions that syntactic devices used in the production of an expression are simpler in one language than the other or Celce-Murcia's (1978) suggestion that bilingual children may avoid phonologically difficult words is difficult to ascertain from the
preliminary analyses. There is generally a tendency to favour vowels, nasal consonants and stops to the exclusion of fricatives and affricatives and /r/ and /l/.

Further examination of the linguistic networks needs to be undertaken to determine specific trends in the phonological development as compared to other monolingual and bilingual studies, specifically the studies of Bai (1984), Bizzarri (1984), Celce-Murcia (1978), Meisel (1986), and Schinke-Llano (1986).

Evidence from the transcripts suggests a difference in the way language was “demonstrated” by the parents and grandparents respectively. Linguistic demonstrations provide the learner with a clear example of how to use language. Direct and indirect demonstrations of language provided Zoe, with a literal immersion in functional language. The father and both sets of grandparents generally responded to Zoe’s linguistic approximations by asking “What do you want Zoe?” After a short pause they then provided the word, for example “banana” or rephrased the question “Do you want a banana?” Both parties initiated conversation in much the same way. Zoe’s mother on the other hand was referring to.

In 15 months there was a noted increase in the linguistic options and a phonological closeness to adult words, particularly in Spanish. This is attributed to the change in social context which saw her Spanish grandparents baby-sitting continuously for extended periods of time. The type of communicative exchange as discussed, became a feature of the communicative events of that period. The linguistic developments which resulted serve to highlight the effect environmental factors such as input and interactional mode can have on a child’s bilingual development. In the context of the present study these factors need to be examined in more detail to determine more fully the parents’ “theory of linguistic accommodation” (Evans, 1987), the “discourse strategies” they employ (Dopke, 1986), and the “interactional patterns” (Garcia, 1988) they engage the child in.

Conclusion

The functional organization of protolinguistic networks provides a window through which to view the development of language. Examination of the linguistic expressions, their meanings and the context from which they develop serves chiefly to highlight the uniqueness of the child’s language and the acquisition process.

The results of this study are unique in that the protolanguage described draws specifically upon one child and her life’s experiences. Zoe’s language developed as a consequence of her involvement with the people, the objects, the events, and the actions that characterized her world. Therefore the findings of this study cannot, as a result of language “uniqueness,” be generalized or applied to the majority. In many ways however she does share with other children, at a general level, characteristics of language acquisition. Particularly other monolingual studies of protolinguistic development (Halliday, 1975; Painter, 1984) who used similar data collection and analysis techniques.

A feature of this study was the irregular development of language. Some words gradually appearing from linguistic approximations over time, while others appeared suddenly in the context of personal, interactive interest. It was noted that as the protolanguage took on characteristics of both the mother’s tongue and father’s tongue, a mixing of expressions occurred. Several expressions were phonemically similar in both languages. These phonemic similarities raise issues regarding general phonological development and phonemic tolerance. At this stage of analysis these issues remain areas to be more fully explored.

The study highlighted the importance of using a range of data collection and credibility checks when attempting to ascertain linguistic meaning and function. The shared responsibility of checking and evaluating the data helped to attain a more multidimensional, clearer picture of what was “appropriate” by this child, or the society in which the individual lived. Specifically, the function of particular expressions, their meanings and their development through time. More detailed examination of these factors reveals the importance of linguistic accommodation and options. Viewed in this way, the importance of the study is attributed to analysing the significance of language, meaning.

At 16½ months Zoe was again a two-level system that was developing. Her meaning (morphosyntactic) was not involved, and the development of language had not yet been seen to be an input but rather a produced construct. Zoe’s meanings were.

REFERENCE

checking and cross checking interpretations helped to attain trustworthy data, eliminating the possibility of single researcher bias. The multidimensional approach provided a much clearer picture of the nature and functional use, by this child, of specific linguistic expressions. Specifically, the cross functional use of expressions, and related options largely determined through the visually recorded video data. More detailed analysis of the protolinguistic stages reveals in many cases an overlapping of expressions in respect to the assigned function and options. Video recorded data highlighted the importance of contextual support in respect to analysing (proto)language function and meaning.

At 16 1/2 months, language was identifiable as a two-level system, one level of functional meaning (morphemic and relational), the other involving expressions (phonological). The development of the child's linguistic system was seen to be arbitrary. The oral expressions produced conveyed meaning however these meanings were not initially represented in the expressions but rather were attained over time. Characteristics of linguistic conventionality and language differentiation became more evident at 10 months of age.

Zoe's transition from protolinguistic stage displays greater development within the pragmatic functional options, particularly with her father. That is her demands for objects, for action, and her interactions in which others construe her intended meaning, seemed to develop ahead of the mathetic or confirmation options. This suggests a possible difference in the roles played by the parents; something which needs to be explored further in future analysis.

Finally, through a multidimensional approach to data collection and the credibility checks undertaken, a range of multidisciplinary perspectives were generated from which to view the data. These perspectives are yet to be fully explored; however preliminary analysis suggests these avenues of inquiry may lead toward a much more informed understanding of this early phase of childhood bilingualism.

REFERENCES


Toh Kok Aun

Abstract

Research should play a central role in decision making. Practitioners as well as policy makers can ill-afford to ignore the importance of research, particularly those involving the local context, in informing practice. Contextual differences are often incurred when overseas research findings are imported. A very strong case can therefore be made for more local research studies for use within the local context. There is, however, a dearth of local studies in science education over the past two decades. This paper attempts to briefly survey the field to account for the impoverished state. It will then review some thirty-six research studies carried out in Singapore spanning the period from 1971 to 1990—these representing the studies where documentation were available—with the view of pooling together the findings of the studies and eliciting trends for the benefit of future research efforts as well as enlightening practice and policy.

Descriptors: Science education, research environment, science teaching.

Introduction

The day-to-day practice of teaching and learning can benefit much from relevant research. Sadly, however, there is a dearth of research emanating from the local scene. It is possible that the culture for research is yet to be ingrained into the bulk of practitioners in the field. Most practitioners do not see themselves as researchers, and go about their daily routines by virtue of habit. Two reasons may be advanced for this. Firstly, practitioners are not only not generally research-oriented but are relatively unused to initiating innovations as a way of life. This results from practitioners’ preoccupation with their teaching role, leaving research to a small handful who venture into often sporadic and ad hoc investigations on their own. The research work carried out are thus more of an ad hoc nature. It can thus be characterised as being uncoordinated and lacking in any particular agenda.

Secondly, the usefulness of research in science education is often questioned, for even where findings were not equivocal, contextual differences could always be demonstrated to limit or undermine their generalizability. Being uncoordinated, and often lacking in resources to carry out studies involving large sample sizes, it is not unexpected that many local studies result in limited generalizability.

Notwithstanding the above, there has been a gradual expansion, albeit rather slowly, of the volume of research in science education. This is motivated by the need for research that was the change. Research is primarily based on, and included recruitment, training, promotion and the like. There is a particularly a move to training non-scientists in research. As the number of researches to attain criteria is relatively small than many have been carrying out.

Secondly, there is also encouragement given for research, and a commitment to $2000 to $3000 to support some or another the researcher. It is also highly likely to attract the individual researcher to the school; or opposed to it.

The third was the launch of the Master of Educational Studies course in 1982 at the university. The programme was a research stream of the programme of a dissertation by students in particular. The amount of research trained to carry out research. The fourth was the growth of research in Science Education in Singapore. It has also been termed as a catalyst for research, under one concept annual conference, sharing the research in becoming a movement attracting a large number outside Singapore. There is more research and more research in...
motivated by at least four forces. One of them was the change in emphasis from a tradition primarily based on teaching to one which included research as a criterion for faculty promotion at the tertiary level. This change particularly affected those who had neither the training nor the inclination to do serious research. As faculties of tertiary institutions strove to attain credibility, the “publish or perish” syndrome inevitably set in. But it came sooner than many had expected.

Secondly, at the school level, teachers were also encouraged to carry out some form of action research, and could even apply for a grant of up to $2000 to carry out the study. For one reason or another this inducement did not attract many takers. One can venture to guess that teachers were either too busy doing what they are doing, or felt inadequate to embark on research work. It may also have been due to a lack of reward to attract the classroom teacher to carry out individual research, or a lack of support from the school; the school may also have been opposed to it being used as a research laboratory.

The third force came from the introduction of the Masters degree programme in education in 1982 at the then Institute of Education. The programme required the candidate to complete a research study culminating in the submission of a dissertation. This brought in postgraduate students in general, and science education in particular. This caused an increase in both the amount of research and the number of persons trained to conduct it.

The fourth push came in 1987 when a group of researchers felt the time was ripe to set up an Educational Research Association (ERA) in Singapore. Its inception was timely and served as a catalyst for more researchers to be brought under one common umbrella. The ERA organises annual conferences to provide the forum for sharing the findings of research, and is fast becoming accepted as a regional platform, attracting a sizeable number of presenters from outside Singapore. In its attempt to encourage more research in schools, it will also be organising an award for the best school-based research in 1993.

**Rationale for the Review**

Information on significant findings of research done in the recent past is necessary for the purpose of indicating the extent and quality of research carried out. This review summarises the main strands in science education research in Singapore over the past two decades, with attention to both methodology and findings. It facilitates the process of reflection and analysis, thus serving as a critical mode of evaluation as to how wisely researchers have invested their time and resources. It will also help ascertain the existing trends to aid in generating guidelines for the resolution of problems as well as for the improvement of science programmes or instructional strategies. The needs and directions for future research can also be established. This review will therefore:

(a) examine the trends that are discernable from the studies from the perspective of investigational methods and subject variables;
(b) present a critical analysis of the studies under each of the subject variables;
(c) discuss the significant findings and identify the crucial factors that will be meaningful and effective in the teaching and learning of science.

**Data Sources**

Any review should start with a consideration of those aspects to be included and those to be excluded. This review took into account studies over the past two decades, spanning from 1971 to 1990. For this review it was necessary to consider only those for which documentation was available based on a library search of documented listing of research reports and studies from various institutions, government agencies and ministries. The documents that finally contributed to this paper were mainly from institutions of higher learning. Many of the reports obtained were, in fact, Master or Doctoral dissertations.
Research Trends

Time and space limitations do not permit a detailed elaboration of each piece of work reviewed. Among the diversity of topics, methods and interests, the 36 studies reviewed are analysed under the following headings:

- Investigational Methods: the data gathering procedures;
- Subject Variables: the areas of study investigated;
- Target Level of Studies: the level from which the study sample was taken;
- Source of Study: the document source.

Within this framework all studies have been analysed for their basic features (see Table 1).

While many studies cover a large number of factors, only the primary intentions of the authors have been included. At the same time as much contributing data as possible have been assembled.

Investigational Methods Used

The much maligned and, currently, heavily criticised agricultural-botany model of research still seems to be popular, with investigators continuing to recognise it as being useful. Continuing faith in its efficacy is apparent with investigators in Singapore (e.g., Chia et al. 1987, Toh 1990a). It is of course in the UK where the strongest criticisms have been voiced and this

The protoc

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental/Quasi Experimental</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Surveys</td>
<td>-</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Protocols</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Case Studies</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Descriptive Studies</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

Subject Variables

<table>
<thead>
<tr>
<th>Subject Variables</th>
<th>1971-80</th>
<th>1981-85</th>
<th>1986-90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching and Learning</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>Assessment and Examinations</td>
<td>-</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Teacher and Teacher Education</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Development Projects</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Curriculum Development/Implementation/Evaluation</td>
<td>1</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

Target Level of Studies

<table>
<thead>
<tr>
<th>Target Level of Studies</th>
<th>1971-80</th>
<th>1981-85</th>
<th>1986-90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>-</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Lower Secondary</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Upper Secondary</td>
<td>-</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Junior College/Pre-University</td>
<td>-</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>All levels (except tertiary)</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Tertiary</td>
<td>-</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

Source of Study

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ph D</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Masters</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Research Reports</td>
<td>-</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Chapter/Book</td>
<td>-</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Conference Paper</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Working paper</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
might have contributed to the relatively higher frequency of the alternative technique of surveys in Singapore. The excellent transport and communications infrastructure of Singapore and the relative proximity of all schools as compared to a large country, where personal contact is difficult because a nearby school could be miles away, may be the reason for the in-roads made by the survey method. It is also quite possible that many researchers resort to surveys to avoid having to set up experimental-control group situations with appropriate considerations given to sampling of who should go into each group.

The protocol and case study methods which are popular in the UK and the USA are slow to catch on in Singapore. Does the research question posed influence the choice of technique to be employed? Is it a question of hermeneutics? Is there a particular belief in empirical studies as being more superior? Is it the small sample size of protocol and case study methods that is being questioned? Certainly, a glance at Table 1 indicates the surge in the use of empirical techniques as compared with the others.

One notable point is the marked increase in the overall number of studies carried out in the 1980s as compared to the 1970s, in particular the late eighties. Undoubtedly this is the result of strong emphasis given to research, the embarking on the M.Ed. programme in earnest in 1982 by the then Institute of Education, and the inception of the Educational Research Association.

Subject Variables

The 36 studies reviewed fall into five major strands of inquiry (see Table 1). The strands of assessment and curriculum seem to hold sway among researchers. The author does not think that this is because there are more problematic areas under these subject variables than in other areas. Rather, the popularity of the area of assessment is understandable in the context of a meritocratic society and the need of researchers to probe into the efficacy of different techniques of assessment. The surge in the area of curriculum development/implementation/evaluation in the early part of the 1980s (Singham et al. 1983, Soh et al. 1983, Tan et al. 1983) was no doubt the result of the introduction of new curricular materials by the then newly established Curriculum Development Institute of Singapore. The low level of activity in this area presently is indicative of the current shift in research emphasis from curriculum development towards other areas.

Also notable in the latter half of the 1980s was the move into investigations concerning the pre- and in-service teacher education domain. This is a sign not only of maturity but also of the researchers' awareness of the importance of the teacher variable in the overall equation for the education of the school-going population.

Despite the importance of teaching and learning, this area has not attracted as many researchers as one would have assumed. Perhaps this is the result of our unique geo-political environment which inhibits researchers from probing into sensitive learning areas arising from racial and ethnic considerations. One would have, however, expected researchers to attempt to investigate the improvement of teaching through the problem-solving strategy, for example, to compare the lecture with the problem-solving approach. The lecture method is now the dominant form of delivery in junior colleges, as a precursor to university-style teaching. The coupling of small tutorial groups with large lecture groups provide a useful transition between school and university. Another possible area of research could be the optimum load that a teacher could undertake. Is the present teaching load too heavy, thus preventing the teacher from giving off his best? Current attrition rates in the ranks of the teaching force may warrant taking another look at teacher burn-out.

Research into development projects has taken on a low key despite emphasis given to non-formal education. Singapore boasts of having a Science Centre of international repute, and one of the best Zoological Gardens in the region. There is also active governmental sponsorship of students for participation in International Science Olympiads. In spite of all this, research into this aspect has not really taken off.
Critical Analysis of Studies

Teaching and Learning of Science

Although the primary intentions, as expressed by the titles of the six studies (see Table 1), appear to be somewhat different, they all lead to the common concern of bringing about optimal conditions for meaningful learning. The study by Lim (1986) and Toh (1990b) similarly identified affective factors as the most important variable in improving performance.

Gender effects in favour of males reported by a number of studies (Yeoh and Tan 1986, Lim 1986, Woo 1989, Ng 1989, Goh and Chia 1990b) are in line with many other documented studies (Comber and Keeves 1973, NAEP 1978, Hobbs et al. 1979). The study of Toh (1990a) for hands-on practical investigations, however, does not demonstrate this. It reported that girls who had similar ability levels as boys, performed just as well as boys and even outperformed the boys when provided with content familiarity. One would have expected the hands-on practical investigation to be a male-dominated domain with boys showing superiority over girls. This reversal is being reported in growing abundance in the literature (e.g. Klainin and Fensham 1987, Johnson 1987) and this study adds to the growing number of studies that refute the claim of male superiority.

Since Singapore is expanding its pool of R&D personnel to carry out innovative research and product design, it is of great importance that the female pool be motivated and challenged to take up science careers as well. The findings of female students lagging behind their male counterparts in science achievement are disconcerting. A "missing half" is a severe strain on any work force, and even more so for a small work force. The study by Toh (1990a), which concurs with those from other countries (see, for example, Klainin and Fensham (1987) in Thailand, and Johnson (1987) in U.K.), does offer some hope. More in-depth studies are still needed to probe into the possible causes of this phenomenon, so that steps could be taken to enable more female students to excel in the field of science.

The finding of student cognitive preference for recall of fundamental principles, and particularly of factual information as modes of presenting information (Chan 1984) seems to indicate the use of surface structures (Larkin and Reif 1979) rather than meaningful learning (Ausubel et al. 1978). This would indicate that low-level learning is predominantly taking place. In confirmation of this, Ong (1987) found severe shortcomings in students' understanding of concepts.

These studies point to the need for a constructivist pedagogy in science teachers' approach to teaching rather than merely regurgitating the scientific view as presented in the students' text. This is further reinforced by the finding that what teachers perceive as difficult is not always what students perceive as difficult (Goh and Chia 1990b). The need therefore for teachers' perception of difficulties to match those of their students' suggests the importance of a constructivist approach to teaching and learning (Cheung and Toh 1990).

Assessment and Examinations

The findings that students in the lower academic stream lagged behind those in the higher academic stream in science achievement under the aegis of the IEA's Second International Science Study (Yeoh and Tan 1986) seem to justify the provision made for the less able students to learn at a slower pace. However, there is a need to look into other ways of narrowing the gap in performance between the less and more able students, as it is the policy of Singapore's education system to "allow a child every opportunity to go as far as he can" (Goh et al. 1979).

Two studies reported a significant relationship between science achievement and general ability (Yeoh and Tan 1986, Woo 1989), while one reported a significant relationship between science achievement and past performance (Ng 1989). These are in accord with our intuitive understanding of student performance. However, these are unalterable variables, as a teacher is not able to alter the general ability of a student or to go back and alter his past performance.

In view of the important part played by assessment in Singapore's educational system, there is a need to review the assessment procedures teachers in the U.K. are using. Of particular concern are the assessment techniques teachers use to identify pupils which cannot be done by conventional pencil methods. A video technique, for example, a video technique which has gained wide acceptance in the U.K. as well as in Singapore, seems to indicate the need for a viable alternative to the traditional assessment task. In particular, the video technique can provide an opportunity to assess the practical skills of students.

One must note, however, that for a competency-based assessment of practical skills is synergistic with the General Certificate of Education Ordinary Level examination system. The integration of science with scientific understanding, problem-solving skills, and establishment of the assessment of practical skills is an important component of Singapore's educational system. The practical examination (General Certificate of Education Ordinary Level) skills of science (general abilities and skills) cannot be ignored.

Locally developed and administered Primary School Science (PSS), Secondary School Science (SSS) and Science Awareness (SA) assessment are designed to problem-solving and to foster development skills. These tests are aimed at implementing the science curriculum goals. The findings of this research, however, that the Lowe Learning Styles Questionnaire (LLSQ) is not particularly useful in identifying and orienting the needs of students, points to the need for further research on the development of practical skills. These findings are in accord with other research (Ng 1989). The findings of the survey also indicate the need for instruction and training on the part of teachers collectively, so that the teachers can aid students' success.
there is a need for more active research into assessment techniques that will assist classroom teachers in the assessment of pupils' attainment. Of particular importance is the development of techniques in the assessment of practical skills which cannot be assessed with the paper-and-pencil method. Two studies took up this challenge and examined the possibility of using a video technique as an alternative mode of assessing certain practical skills (Toh 1986, Toh et al. 1987). The authors made good use of zooming techniques to capture fine details to test students' observation skills. Unfortunately, the video technique is only good for testing observation skills and cannot be considered as a viable alternative of assessing the entire range of practical skills.

One must not assume that acquisition of, or competency in performing the different practical skills is synonymous with carrying out practical work. The integrated practical skills associated with scientific inquiry, such as hypothesis generation, planning and designing experiments, establishing the variables to control before proceeding to test out the hypotheses, are far more important. However, these integrated practical skills are still not assessed in the Singapore-Cambridge GCE O-level and A-level practical examinations. The importance of these skills of scientific inquiry (or problem-solving skills) cannot be over-emphasised.

Locally developed curriculum materials (the Primary Science Project and the Lower Secondary Science) reflect science educators' awareness of the importance of developing problem-solving skills as part of the overall development of the learner. However, what is implemented falls far short of the hoped-for goals. The findings of Kam (1985) informed us that the Lower Secondary Science, which is an inquiry-oriented course, only enhances the lower order skills. The higher order, or more integrated, skills of planning and hypotheses generation are still lacking. Specific coaching in these skills is necessary as they are not easily mastered (Tamir 1989). The findings of Toh (1990a) indicated that instruction and practice, both individually and collectively, contributed significantly towards the students' success in laboratory investigations.

The unequal emphasis given to different process skills by teachers teaching the Primary Science Project (PSP) was also noted in the study by Singham (1987). The skills of observation were highly emphasised, while those of hypothesising and planning investigations were grossly neglected. Could this be the outcome of discrepancy between the intended curriculum and the assessment for the course? Or could this be the outcome of inadequate teacher preparation for mounting an inquiry-based programme?

**Pre-Service and In-Service Training of School Science Teachers**

The study on expert chemistry teaching (Wong and Koh 1990) confirmed what research on expert teaching has found, and is, at best, exploratory. There is a need to identify more of such experts for further inquiry before any profile on expert chemistry teaching can emerge or generalisations made. An understanding of expert teachers' thinking and their teaching behaviours is useful for the preparation of science teachers.

Goh and Chia's (1990a) study showed that student teachers can be trained in good teaching practices (eg asking operational questions) if their pre-service training exposed them to such opportunities. Their study also showed that student teachers' attitudes to science teaching can be enhanced through carefully structured and worthwhile learning experiences. Their study offers much scope for improvement in science teacher preparation programmes.

Cheung and Toh's (1990) study is important for probing the pre-service teachers' conception of the nature of science and science teaching. Their findings indicated that many would-be science teachers were not conversant with the nature of science and the ways science should be taught for conceptual understanding. The teachers who show continued preference for viewing scientific facts with certainty (or as infallible) present an epistemological obstacle towards a constructivist pedagogy when they teach. This has important implications for the pre-service training of science teachers.
Development Projects in Science

The widespread availability of personal computers at home and in all schools in Singapore has not been fully exploited for educational purposes. The findings of the study by Foong et al. (1982) clearly indicated that well-designed programmes, which include graphic and sound effects, are more useful than printed matter in promoting learning. Science programmes based on the existing syllabuses can be prepared as part of the curriculum materials to which students can return for repeated reference. They are more user-friendly than textbooks because they can be made interactive. The programmes can also be made to cater for the different academic streams.

The finding that students’ participation in enrichment activities relevant to school science would improve science achievement (Chang 1974, Lam-Kan 1985), is noteworthy to school principals and educational leaders. Informal science education has a role to play in science teaching and should not be underrated. Science teachers should be encouraged to organise field trips to science resource centres such as the Science Centre, Bird Park, Zoological Gardens and Botanic Gardens. Many of these organisations employ education officers who are more than willing to work out a programme with teachers who wish to bring their pupils there for a particular purpose. One should also not overlook students’ participation in science fairs, projects, and olympiads.

Science Curriculum Development, Implementation and Evaluation

In view of the generally known fact that teachers tend to be greatly dependent on textbooks, and in fact, tend to equate textbooks with the prescribed syllabuses, it is perhaps a little surprising that except for the summative evaluation in 1983 of the Primary 3 (Normal) and Primary 4 (Extended) curricular materials, there has been no comprehensive evaluation carried out by separate, autonomous and impartial researchers on the science curricular developed for the other grade levels. In fact, in view of the cruciality of science curriculum development, implementation and evaluation in the educational enterprise, the need for systematic and comprehensive evaluation of the entire curriculum process, as well as the products of the curriculum planning, development and implementation cycle by autonomous and impartial researchers, cannot be over emphasised.

With regard to the summative evaluation carried out by researchers from the former Institute of Education (IE) on the Primary 3 (Normal) and Primary 4 (Extended) curricular materials, it is noted that the IE researchers (Tan et al. 1983) recognised that “there are many aspects of summative evaluation that can be fruitfully investigated”, but due to the constraints of time, manpower and other commitments, they were restricted to looking at only three aspects, namely,

- Intrinsic evaluation of the materials developed and the methodology advocated;
- Classroom observation of the materials in actual use;
- Teacher opinion survey regarding teachers’ feelings on the materials and methodology.

Thus it can be seen that while materials themselves are subjected to close scrutiny and teachers’ opinion polled, no study has been made of how pupils view the Project, or of whether their attitudes towards science and science learning have changed as a result of the implementation of the Project.
Another aspect of evaluation that could be looked into is that of pupils' science achievement. This could involve investigation at the trialling stage of the materials, when not all the pupils in each cohort are involved in the Project, leaving some classes intact to serve as control groups; the science achievement scores of pupils who are exposed to the Project could then be compared with those of the control classes. A positive gain in science achievement scores would then be an indication of the effectiveness of the curricular materials.

Another indication of the effectiveness of the science curriculum projects that have been developed and implemented could be based on a longitudinal study, involving the comparison of science achievement scores of Singapore pupils of different age groups under the aegis of the International Association for the Evaluation of Educational Achievement (IEA) Science Studies vis-a-vis those of similar age groups in other countries.

It is indeed unfortunate that not only was summative evaluation of the "home-grown" science curriculum projects at other grade levels not carried out, longitudinal studies involving the same schools, same teachers and same classes, were also not carried out. It is expected that with time, and perhaps greater familiarity with the Project, teachers' opinions and attitudes, as well as the classroom interaction processes could change. Such longitudinal studies, carried out at least over a few years, could afford more accurate data and feedback on the curriculum development process.

Summary and Recommendations

One cannot review the science education studies of the last twenty years without wondering why they have not led to greater impact on the way science is taught and on what is learned in the classroom. Explanations for this disappointing degree of influence may stem from the gulf between scholarly writing and the writing that practitioners are used to. This section attempts to bridge this gulf by presenting the summary and recommendations of the main findings of the studies reviewed, with minimum technical jargon. Each of the findings identified below is the outcome of more than one study, and is thought important enough to be pooled together as a summary.

1. The low level of learning occurring in schools, one based on fundamental principles and strong reliance on recall for presenting information, is a cause of concern for all those concerned with outcomes in the educational enterprise. Teachers continue to fuel this practice in the course of obtaining examination performance results. This could be a case of examinations dictating practice.

2. With examinations reinforcing the tendency to regurgitate factual information, many students up to the eve of examinations still exhibit severe shortcomings in their understanding of concepts. This could reflect a mismatch in perceptions of learning difficulties between the teacher and the student. It could also indicate that teachers do not subscribe to a constructivist pedagogy.

3. Most educators are aware of the important part played by a good attitude to school or to the subject in influencing learning and achievement. Studies done in Singapore now confirm the importance of these affective variables. "Get the students' attitude right, and the rest follows." This advice is given at the risk of over-simplification, of course.

4. Denying the less able students of opportunities to do practical work is a sure way to turn them off from their science lessons. By doing this, the teacher does not promote good attitudes in students, and this perhaps reflects upon the teacher's attitude to science.

5. Measures of general ability and past performance continue to be good indicators of future performance. These could continue to serve the teacher, but they are unalterable variables.

6. Only low order skills are being practised during laboratory sessions. This is a discrepancy between the implemented
curriculum and that intended by the curriculum developers. There is a need for teachers to allow their students to confront problems requiring both hypothesising, planning and designing of hypotheses-testing experiments as intended by the curriculum developers. This discrepancy is an indication that it is not just the curriculum development aspect that is important, but rather, the implementation aspect that is just as crucial.

(7) Teachers who do not have a strong academic background in science are unlikely to possess the desired attitudes to science and science teaching, and the beliefs about science to help them shape the attitudes and beliefs of their students.

(8) The vast potential of micro-computers as an interactive teaching tool has yet to be exploited. Lessons can be canned on disk for individualised instruction. They are more user-friendly than textbooks. They can also be varied to suit the needs of different academic streams. The lower academic stream students specifically need special ways to motivate them to learn. The search for more effective methods of teaching the lower academic stream students should be the concern of researchers interested in improving their performance level.

Conclusion

This review aims to show the diversity and richness of indigenous research. Though they were uncoordinated, the author has attempted to show that they fall into neat subject variables. It is hoped that this review will provide some guidance for future research efforts. Recommendations and implications of findings were highlighted in the hope that Asian practitioners can benefit from them, since contextual differences of studies done in the West may have limited generalizability in the Asian context.

The author wishes to acknowledge the contributions made by Miss Yeo Tiew Kin, Dr Diong Cheong Hoong, Dr Chia Lian Sai and Miss Boo Hong Kwen, in the review of some of the 36 studies used in the preparation of this review.

REFERENCES


Development of a Framework for Analysing Mathematical Problem Solving Behaviours

Foong Pui Yee

Abstract

The taxonomy described in this paper was developed to investigate the process of mathematical problem solving in terms of definable behaviours. It was also used as an instrument to classify and encode behaviours in their sequence of observed occurrence during the process of mathematical problem solving. It is a behavioural analysis framework formulated to examine the “thinking-aloud” protocols of individuals for comprehensive information about the problem solving process itself, the individual differences in the behaviours of subjects and the strategies applied by each in dealing with non-routine mathematical problems.

Descriptors: Mathematics problem solving, thinking aloud, protocol analysis.

Introduction

uch of the recent research on mathematical problem solving attempts to describe and characterise the problem solving process, derived from verbal reports. It has become increasingly important for research to develop valid and reliable instruments which reflect actual behaviours in the problem solving process. Kilpatrick (1967), Lucas et al. (1980), Rowe (1980), Schoenfeld (1983), Putt and Pountney (1989) and many others have devised protocol analysis frameworks to record and analyse behaviours in a sequence of observed occurrence.

This paper describes a framework for analysing the “thinking aloud” data from problem solving performance on certain non-routine mathematical problems. The methods of data collection and protocol analysis were derived from research within the framework of information processing influenced by Newell and Simon’s (1972) theory of problem solving.

The present study was able to develop a taxonomy comprising cognitive, metacognitive and affective behaviours manifested in individuals during problem solving activities. The taxonomy identified 28 behaviour categories which could be coded reliably. These behaviours could be considered as low inference measures that were objective, observable and most importantly could be defined in operational terms. Inter-coder agreement was used to establish reliability and to refine concepts which had to be reconciled with the actual data under the conditions set up by the methodology in the present study.

Design

The Sample

The nine volunteer subjects in this study were adults enrolled as pre-service trainee teachers at the Singapore Institute of Education. A range
of mathematical abilities for this adult level was chosen so that all possible behaviours could be expected to occur. The nine subjects were from three levels of mathematical background. Two had General Certificate of Education (GCE) ‘O’ (Ordinary) level secondary school mathematics with low credits, two GCE ‘A’ (Advanced) level post-secondary mathematics with average credits and five had university degrees in mathematics with two having upper class honours. They were recommended by their respective mathematics education course lecturers who considered that the subjects’ mathematical backgrounds were suited to the investigator’s purpose. These volunteers were briefed by the investigator on the aim of the project and given training in “thinking aloud”.

All the participants were proficient in the use of the English Language. This proficiency was both a prerequisite for their selection into the teacher training programme and had been the language of instruction throughout their mathematical education. This was an important consideration as this investigation utilised verbal reports as data and the subjects, who were not native speakers of the language, needed to be confident in the use of the language in order to participate in the “thinking aloud” procedure.

The “Thinking Aloud” Technique

As the processes during cognitive performance tend not to be directly observable, the first task was to find a methodology which would make it possible to identify and monitor valid elements, in terms of operationally defined behaviours. A promising method was the “thinking aloud” (TA) technique which generates data from the concurrent verbal reports of subjects while engaged in problem solving. The main concern was that the TA instructions should be given to the subjects so that they could vocalise all their problem solving activities with minimal interruption while performing the tasks.

Previous research, using a similar methodology, commonly gave training to subjects before the study began so that during problem solving they could “think aloud” in an ongoing manner and without time delay. It was found that “thinking aloud” was relatively easy to learn given sufficient practice and as long as the subjects were motivated to cooperate. A standard procedure was essential for all subjects and tasks. The experimenter should not say more than absolutely necessary; neither direct the subject’s approach, ask leading questions or reinforce while the subjects were “thinking aloud”.

The TA instructions were adapted from previous studies reported in Ericsson and Simon (1984). The subjects were asked to vocalise their thoughts as though they were talking to themselves. They were told to constantly vocalise whatever came to mind and that the experimenter was not primarily only interested in their final solution but also in their thinking processes. To prevent subjects from explaining their solutions aloud, the experimenter followed Krutetskii’s (1976) warning to subjects at the beginning against confusing the instruction to think aloud with that of explaining the solution: “Do not try to explain anything to anyone else. Pretend there is no one here but yourself. Do not tell about the solution but solve it.” (Krutetskii, 1976, p 93)

Procedures for Data Collection

The subjects were arranged in individual recording booths in a language laboratory to take part in the experiment. The experimenter started each subject off with a problem presented on a typed sheet of paper. The subject was asked to begin work on each task by reading the problem statement aloud. Once this had begun, the experimenter moved on to the next subject to start on the same problem. The subjects were told to put up their hands when they had finished with a problem and the experimenter would come to present them with the next problem. Each problem was presented on a separate sheet of paper and subjects were also provided with pencil and paper for their scratch work. Five problems (Appendix A) were presented, one at a time in the same order, for every subject. No time limit was imposed and the subjects were reminded at the beginning of the session of the TA instructions.

All the subjects’ comments were recorded on tape. Each subject’s protocols were transcribed and protocols written for the nine subjects. Each protocol was comprised of a subject’s audio tape and the experimenter’s transcriptions of the subject’s solution processes. The protocol was then divided into solution behaviours. Each protocol was independent of all other protocols and consisted of the sequence of cognitive problem solving processes. Each protocol consisted of the sequence of cognitive problem solving processes. Each protocol was divided into a series of short recording episodes (Appendix A). These were then coded and classified as follows:

Protocol Analysis

Protocol analysis was carried out using an event-by-event coding scheme on five separate protocols for each subject as a source of data for coding. The taxonomy of the framework that was used was:

(a) a taxonomy of the subject’s behaviours plus the occurrence of a code

(b) a working list of possible protocols

The Taxonomy

The term taxonomy refers to a predetermined set of specific and behavioural variables. The study presented five distinct categories of codes for characterising the cognitive problem solving behaviour. The five categories are:

Encoding decision

(1) PROBLEM STRATEGIES
All the subjects attempted to solve the same five problems. Each problem solving session was recorded in its entirety and the audiotape protocols were transcribed verbatim. Forty-five protocols were collected, each protocol consisted of one complete transcription of a subject’s audiotaped recording of his or her solution process to a problem. Each transcribed protocol was divided into segments of behaviours. Each segment was then classified and encoded according to a predetermined taxonomy of behaviours and a process-coding scheme. For each subject and task, the data consisted of the types, the frequencies and the sequence of coded behaviours used during the problem solving process.

Protocol Analysis

Protocol analysis refers to the use of a systematic event-by-event record of an individual’s behaviours while engaged on a cognitive task, as a source of data. The analysis of the “think aloud” protocols in this study were based on a framework that consisted of two components: (a) a taxonomy with a process-coding scheme to code all observable problem solving behaviours in actual sequence of their occurrence. (b) a working model for episode-analysis of the protocols to identify global patterns of metacognitive processes.

The Taxonomy

The term taxonomy used in this research refers to a predetermined set of problem solving behavioural variables that are classifiable into five distinct categories. It serves as a dictionary for characterising the mathematical problem solving behaviours in the subjects’ protocols. Encoding decisions made by a coder were based on this dictionary of coding categories. A review of the problem solving literature provided the investigator with a list of frequently recurring behaviours that could be classified into five major categories:

1. PROBLEM-ORIENTATION HEURISTICS: strategies through which a problem solver attempts to analyse and understand the problem situation.
2. PROBLEM-SOLUTION HEURISTICS: rule-of-thumb strategies through which a problem solver moves towards a solution.
3. DOMAIN-SPECIFIC KNOWLEDGE: the inventory of mathematical facts, procedures and skills that a problem solver is able to use in the solution process.
4. METACOGNITION: a problem-solver’s awareness and monitoring progress of his or her own thinking during the task.
5. AFFECTIVE BEHAVIOURS: self-related expressions and emotional responses that are aroused in the problem solver.

A preliminary list consisting of all the possible behaviours in mathematical problem solving was prepared from the study of the literature. Altogether 40 behaviours were identified as relevant to the tasks used for this research. They formed the initial taxonomy which was used as a preliminary guide to identify observable behaviours in the protocols of the subjects in this study.

After a series of modifications through tests of reliability a final taxonomy (Appendix B), with a coding system was derived. Although it was initially based on prior assumptions about problem solving processes from previous research, the taxonomy was modified empirically, through the consensus of different coders after they had applied it to actual problem solving protocols of different subjects across different problems. In order for the coding judgments of any coder to be as objective as possible, the definitions of each specific behaviour in the taxonomy were stated in operational terms and were supported by examples from the protocols (see Appendix C for some definitions and examples).

Process-Sequence Encoding Procedures

Encoding was carried out on the transcripts of the subjects’ audiotaped protocols, together with their written work, to provide a more comprehensive analysis of the solution path. The analysis of the protocols was based on the taxonomy. The overall strategy was to divide a
protocol into segments of behaviour. Each segment was matched against the behaviours in the taxonomy and assigned a code. A whole protocol was eventually recorded as a sequence of encodings in a horizontal string corresponding to the order of the actual problem solution of an individual. Conversely, a string of encodings could provide a relatively clear description of what had actually happened during the problem solving process. (A sample of the protocol analysis is shown in Appendix D.)

The coding of individual segments of the protocols was intended to facilitate later comparison of encodings by different coders to establish the reliability of the taxonomy. Each segment of a protocol corresponded to a statement. According to Ericsson and Simon (1984), if the verbalisations were completely grammatical, a statement would essentially be a clause or a sentence, but in the thinking aloud protocols which were close to normal speech, statements were often abbreviated to phrases or even single words. It was also found in the protocols that some subjects often verbalised a single behaviour or operation ungrammatically as an aggregate of phrases. Each segment or an aggregated segment of behaviour was then encoded from the information contained in it.

A Working Model for the Episode-Analysis of the Protocols

The flow chart in Figure 1 represents a working model of a global problem solving process where “chunks” of consistent behaviours identified in the taxonomy could be parsed into six types of episodes which were modified from Schoenfeld (1983) for the purpose of this study as:

1. SCANNING
2. ANALYSIS
3. EXPLORATION
4. IMPLEMENTATION
5. REVIEW
6. VERIFICATION

An episode defined for this research is a period of activity during which the problem solver is engaged in a set of consistent behaviours that can be described as a certain event, such as analysis of the problem or exploration of strategies. The episodes can occur in any sequence or they can be recurrent. For example, some problem solvers may, after scanning the problem, have gone straight to implementation without analysis or exploration and in review became “stuck” or “does not work” and might go back to scanning or analysis and work through the process again, or they may “give-up”, as illustrated by the variety of possible paths in the flow chart. The review episodes are junctures where evidence of metacognitive activities are most apparent.

Each episode consists of specific behaviours that are based on the taxonomy. Table 1 shows the list of predominant behaviours in the six episodes which can be classified into three main components of the mathematical process: “orientation to the problem”, “execute the solution” and “evaluation”. Each protocol sequence was parsed by examining the behaviours and dividing them into episodes that were enclosed and labelled in boxes. An example is shown in Figure 2. The episodes were parsed in sequential order of occurrence and were presented in the form of a flow-chart. The behaviours within each episode were presented in code symbols and the interpretations of these codes are shown in the key.

Establishing the Reliability of the Taxonomy

Issues of reliability and replicability in the coding scheme based on the taxonomy are concerned with the questions: “Would others see the same behaviours in the protocol?” and “Would they ascribe to them the same coding categories?”. To address these issues tests of reliability based on agreement between different persons encoding independently of each other were used. The correlation between the categorisations made by different coders in relation to the same task provides an index of the objectivity of the taxonomy for categorisation of problem solving behaviours.
Figure 1: A working model for episode-analysis of the mathematical problem solving process
Table 1: Predominant behaviours in the six episodes of the three main problem solving components

<table>
<thead>
<tr>
<th>I</th>
<th>ORIENTATION TO THE PROBLEM</th>
<th>2</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Scanning</td>
<td></td>
<td>P4 - examine dimensions</td>
</tr>
<tr>
<td></td>
<td>P2 - re-reading of problem</td>
<td></td>
<td>P5 - construct representation</td>
</tr>
<tr>
<td></td>
<td>P3 - paraphrase statement</td>
<td></td>
<td>H2 - draw diagrams</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II</th>
<th>EXECUTING THE SOLUTION</th>
<th>4</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Exploration</td>
<td></td>
<td>H7 - generalise</td>
</tr>
<tr>
<td></td>
<td>H4 - look at cases</td>
<td></td>
<td>H8 - make a deduction</td>
</tr>
<tr>
<td></td>
<td>H5 - guess and check</td>
<td></td>
<td>K3 - use relevant mathematical procedures</td>
</tr>
<tr>
<td></td>
<td>H6 - search for patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>K1 - computation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>K2 - state a fact or rule</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III</th>
<th>EVALUATION</th>
<th>6</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Review</td>
<td></td>
<td>H9 - check computation and final answer; some metacognition and P, H and K behaviours</td>
</tr>
<tr>
<td>M1</td>
<td>suggest a plan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>assess difficulty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>review progress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>recognise error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>new development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>self-question</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Verification</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key:
P1 - first
P3 - rephrase
P4 - analyse
P5 - construct representation
P6 - compute
K1 - compute
K3 - use relevant mathematical procedures
M1 - suggest a plan
M2 - assess difficulty
M3 - review progress
M4 - recognise error
M5 - new development
Q1 - self-question

Figure 2: Ar
FIGURE 2: An example of the Episode-Analysis of a process-sequence - A protocol for the palindromes-problem (Time Taken: 7 min)

EXCERPTS OF THE VERBAL PROTOCOL

"...that means in general a 4-digit palindrome can be written as ABBA..."

"...2 digit number, xy multiply by 11, something xy plus something xy..."

"...no it cannot be...let's find counter examples of palindromes."

"How about 3223...divide by 11...uhmm no...3443 by 11 ....313...uhmm..."

"...wait...there must be some properties of 11 that I can make use of...let's go back to this 4...ABBA."

"...so we need a 2 digit number greater or equal to 91...so when multiply by 11 you get a 4 digit number..."

"...what I'm doing...no...not necessarily..."

"let's try again...I must start with 1001..."

"Oh!...I see...ABBA, I can do it this way ..."

"ABBA is equal to 1000A + 100B + 10B = 1A, alright...then I group them together..."

"100A + 1A + 100B + 10B = A(100) + 10B (10 + 1)"

"...I can take out a common factor ...11"
The protocols collected in this study were divided into individual segments in order that direct comparisons between different encodings of the same segments by different coders were made relatively easy. A percentage of agreement between coders was calculated to give an aggregate measure of reliability. However, this reliability mainly reflects the very frequent coding categories, while the encoding of infrequent categories may be more or less reliable than the average.

To further test the replicability of each of the individual behavioural variables in the taxonomy by different coders, a coefficient of agreement calculation modified by Putt and Pountney (1989) from Lucas et al. (1980) was used. Inter-coder agreement was assessed systematically in two stages. In the first stage, tests of reliability were performed for the five categorical constructs of the taxonomy. This led to a refinement of the preliminary taxonomy by either eliminating, combining, or redefining behaviours which were redundant or vague. The second stage involved tests of reliability for every of the 28 specific behavioural construct defined in the final taxonomy.

On the whole, the majority of the twenty eight behaviours had relatively high coefficients of agreement. Of the total 943 encodings across all problems, inter-coder agreement was relatively high at 83.4%. The most frequent behaviour “H4-Looking at particular case(s)” had a coefficient of 0.86. The mean coefficient values of all the behaviours in each of the five categories were: P-category 0.89, K-category 0.77, H-category 0.76, M-category 0.74 and A-category 0.93. The higher values of the P-category and A-category indicated that these categories of behaviours by problem solvers to analyse the problem situation and behaviours that expressed emotions and self were easier to identify reliably.

Limitations

A number of limitations for this study have resulted directly from the chosen method of the “think aloud” data collection and protocol analysis based on a predetermined taxonomy for encoding behaviours. The contents of the subjects’ problem solving protocols were restricted to what subjects were able to verbalise during the process. The possibility of observer bias in making inferences from the encodings of the subjects’ responses cannot be denied. It is a well known fact that no matter how carefully a research strategy has been designed to be objective, the method of behavioural analysis through the use of verbal data and encoding categories has to contend with some distortions and incompleteness on the part of the subjects’ verbalisation and the inconsistency in inferences on the part of the coder. The taxonomy that was developed is just one of the possibilities which could have been developed in such a study, depending on the position adopted by the investigator on the phenomena under discussion. This would eventually determine what the investigator observed and to some degree influence the regularities and patterns of behaviours that might be identified.

The task of coding was time consuming and required great concentration on the part of the coder. It was not feasible to engage an inexperienced coder for the task. A good level of mathematical background and teaching experience to detect the nuances and complexity of thinking in the students’ problem solving processes for a semantically rich domain like mathematics was found to be essential. The coding system developed for this study required coders to be able to identify a wide range of behaviours, including those that deviated from the ideal. Coders needed to be able to detect any errors in mathematical structure, strategies that could lead the problem solver astray or mathematical insights which an untrained person might not have.

Except for some behaviours which were relatively difficult to code or which had few occurrences, the majority of the 28 behaviours in the taxonomy yielded consistent judgement between the different coders. The inter-coder agreements resulted from the two stages suggest that the problem solving behaviours investigated in the present study can be regarded as having been assessed reliably. The final taxonomy that evolved was intended to be used as a tool for investigating a variety of not only educational research programs, but a variety of programs in the cognitive sciences.

Conclusions

The notion that problem solving can be seen by Schoenfeld (1985) and others as a cognitive science that involves problem solving. In general, the construction of the particular problem solving problem solving process of the interaction between the student and the environment is a central issue in this approach. The present study involved the number of students was relatively small. This study was intended to help the present study.

For too long, educational and psychological researchers have evaluated students’ ability to solve the problems they have been presented with. In this study, the described research strategy has been designed to be a valuable tool for investigating a variety of research programs, including educational, psychological, and cognitive science.

REFERENCES

investigating problem solving processes in a variety of non-routine mathematical tasks for a variety of problem solvers in the main part of a research project.

Conclusion

The notion of "cognitive process analysis" is seen by Schoenfeld (1987) as central to the cognitive science approach to explore problem solving. In general, the approach involves the construction of a "process model" specifying the particular knowledge accessible to the problem solver, the thinking strategies the problem solver appears to have and the nature of the interaction between the two. Studies using this approach are carried out in great depth and the number of subjects involved is usually quite small. This same research strategy was used in the present study.

For too long mathematics teachers and educational researchers have traditionally evaluated students' performance with results obtained from product-oriented testing. Very often a series of multiple choice tests have been designed and used to measure some dimensions of mathematical understanding or ability. In such tests the only aspect of the student's work that received attention is whether or not the correct answer was selected or whether a certain step or rule had been used correctly. Such assessment of students' performance usually has very little information to offer to the teachers who want to make their students better problem solvers.

However, if mathematics teachers could be trained to listen to students solving problems aloud and to then analyse the processes that were used, they would gain greater insight into their students' thinking. If teachers used some of the classifications of the problem solving behaviours, similar to the taxonomy developed in this study, they might also be able to diagnose the nature of the difficulties encountered by the problem solver along the solution path. With this facility teachers could use more effective diagnostic teaching to improve their students' problem solving performance.

REFERENCES

APPENDIX A

THE PROBLEM SOLVING TASKS - Five Non-Routine Mathematical Problems

1 Loans
Sally loaned $7 to Betty. But Sally borrowed $15 from Estella and $32 from Joan. Moreover, Joan owes $3 to Estella and $7 to Betty. One day the girls got together at Betty’s house to straighten their accounts. Which girl left with $18 more than she came with?

2 Palindromes
A number like 12321 is called a palindrome because it reads the same backwards and forwards. A friend of mine claims that all palindromes with four digits are exactly divisible by eleven. Are they?

3 Chicken-Coop
The Smith family wants to build a chicken coop and they have bought enough wire for 19 metres of fence and one gate that is one metre wide. They have decided to make the coop rectangular or square. What width and length would they make the coop so that the chicken can have the largest possible area inside the coop?

4 Diagonals
A diagonal of this 5 x 7 rectangle passes through 11 squares (shaded). Can you find a way of forecasting the number of squares passed through if you know the dimensions of the rectangle? How many squares will the diagonal of a 1000 x 800 rectangle pass through?

5 Christmas Presents (this task was used only in the pilot study)
Mrs Tan is buying Christmas presents for her six children to give one another. Each child gives a present to each of the others. How many presents must she buy?
APPENDIX B

THE FINAL TAXONOMY WITH A SYSTEM OF CODES

P-Category: Problem-Orientation Heuristics

P1 - First reading of the whole problem
P2 - Re-reading the whole or part of the problem
P3 - Paraphrase problem statement
P4 - Examine dimensions
P5 - Construct Representation

H-Category: Problem-Solution Heuristics

H1 - Recall similar problem
H2 - Draw diagram(s)
H3 - State an answer
H4 - Look at particular or simple cases
H5 - Make a guess and check
H6 - Search for a pattern
H7 - Generalise to a rule
H8 - Make a logical deduction
H9 - Check computation or final answer

K-Category: Domain-Specific Knowledge

K1 - Apply arithmetic algorithm
K2 - State a fact, principle or theorem
K3 - Apply routine mathematical procedures
K4 - Mathematical misconceptions

M-Category: Metacognitive Behaviour

M1 - Suggest a plan
M2 - Assess task facility
M3 - Review progress
M4 - Recognise error
M5 - Recognise new development
Q1 - Task-relevant self-question
N1 - Task-irrelevant rhetoric

A-Category: Affective Behaviour

A1 - Negative self-evaluation
A2 - Giving up
A3 - Emotional expression
APPENDIX C

SOME OPERATIONAL DEFINITIONS OF PROBLEM SOLVING BEHAVIOURS

The operational definitions helped to specify the phenomena i.e., the problem solving behaviours under investigation in the present study, to provide a basis for distinguishing between various items of the taxonomy. For a better introduction to the general nature of the phenomenon to be observed, some examples from the subjects’ problem solving protocols were extracted to provide a frame of reference for the operational definitions.

Every behaviour was assigned a code, eg P3 for “Paraphrase problem statement” and if the behaviour was incorrect, inappropriate or irrelevant to the solution then a bar was drawn over the code, eg P3 for “Misinterpret problem statement”. The following are sample definitions of selected behaviours in the five categories of the taxonomy:

PROBLEM-ORIENTATION HEURISTICS

Eg

P3 - Paraphrase problem statement.
Attempt to give meaning to a certain word or to rephrase the problem in their own words.

P3 - Misinterpret problem statement.
A subject sometimes misinterpreted the meaning of a certain word or of a certain part of the text from the problem.

Some examples of code P3:
(a) “...loan and borrow...owes and borrow is the same.”
(b) “...Can you find a way of forecasting ....forecasting means what.?...ah..is it to predict?........Yes.”

PROBLEM-SOLUTION HEURISTICS

Eg

H5 - Make a guess and check
Subject may predict or expect a certain result while considering a particular case and then proceed to check it.

An example of code H5:
(a) “...so, I’ll expect 3 by 7 to be 9...let’s check that 1,2,3,1,2,3,4,5,6,7,...Ok, I expect 9 squares cut ...1,2,3,4,5,6,7,8,9..yes.”

Eg

H6 - Search for pattern
This behaviour follows from H4 or H5 where subjects, after trying some cases, attempted to identify relationships for a pattern or common property among them.

Some examples of code H6:
(a) “...Ok, let me see the relationship again..if it’s 5 x 7 is 11, 10 by 8 is 16, 6 by 4 is 8....”
(b) “...let’s try and get pattern...2 by 3 you have 4 squares, 2 by 4 you have 4 squares, 2 by 5 ....6.”
**DOMAIN-SPECIFIC KNOWLEDGE**

**Eg K3 - Apply routine mathematical procedures**  
Subjects may apply domain specific routine procedures in algebra, geometry, calculus or others in their solution paths.

**K3 - Inappropriate application of procedures**  
Structural errors may stem from a misunderstanding of the problem or some of the principles necessary for its solution. However, any computational error made while manipulating these procedures is coded as K1. Sometimes subjects may try to use certain mathematical procedures which are inappropriate to the solution of the problem.

Some examples of code K3:
(a) "...I can do it this way...ABBA is equal to 1000A + 100A + 10B + A, alright...and then I group them together and get 1001A + 110B, right."
(b) "...can use calculus to calculate...so that gives me ...differentiating with respect to small a ,we get 10 - 2a with zero, we get a=5 and b=5."

**METACOGNITION**

**Eg Q1 - Self-questioning**  
In a problem-solving situation it is quite natural for subjects to ask questions of themselves. These self-questions could be task-specific and were often junctures at which decisions had to be made, eg "...let's say if it's a rectangle in shape...can I have other possible numbers?"

**Eg M1 - Suggest a plan or subgoal operation**  
Subjects might consider the possibility of using a particular plan or subgoal operation. This might be a global plan or hypothesis where the subjects could see the solution and sometimes it could just be an exploration of ideas.

Some examples of M1:
(a) "So we need to generalise first and then find answer for the specific dimensions of 1000 x 800."
(b) "...Ok, let's draw a table and see if I can generalise anything from there."

**Eg M2 - Assess the task facility**  
When presented with the task some subjects would quite spontaneously perceive it as easy, difficult or confusing. Sometimes in the middle of the solution the subject may say a certain operation is easy or hard to do.

Some examples of code M2:
(a) "Let me see...it's simple...straightforward..."
(b) "If it's a square...then it'll be easier."

**Eg M3 - Review Progress**  
Subjects may review and evaluate progress at any stage of the solution path, in terms of adequacy and reasonableness of any results obtained; method used; or outcome of any action or decision made, for example: "What I'm doing is I'm adding it this way...OK... no, that's not very good."
AFFECTIVE BEHAVIOURS

Eg. A1 - Negative Self-evaluation
Subjects assess their own ability, confidence or personality. Express self-doubt, they say they
don't know how to do; can't remember; don't understand; not sure etc.

Some examples of code A1:
(a) "....Umm...not again!..I can't do!"
(b) "....Aiyah, my math is not working!"

Eg. A2 - Giving Up
Lack of persistence. Subjects in this study were not given the chance to ask for help if they were
stuck. This behaviour was coded when subjects indicated their intention to quit the task.

Some examples of code A2:
(a) "Hiyah, I think I give up, I just don't know how to do!"
(b) "I surrender already...Ok, I surrender here."

APPENDIX
A SAMPLE OF

<table>
<thead>
<tr>
<th>Segment</th>
<th>Co</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P1</td>
</tr>
<tr>
<td>2</td>
<td>P2</td>
</tr>
<tr>
<td>3</td>
<td>M2</td>
</tr>
<tr>
<td>4</td>
<td>M1</td>
</tr>
<tr>
<td>5</td>
<td>H2</td>
</tr>
<tr>
<td>6</td>
<td>P4</td>
</tr>
<tr>
<td>7</td>
<td>P4</td>
</tr>
<tr>
<td>8</td>
<td>P4</td>
</tr>
<tr>
<td>9</td>
<td>P4</td>
</tr>
<tr>
<td>10</td>
<td>P2</td>
</tr>
<tr>
<td>11</td>
<td>M1</td>
</tr>
<tr>
<td>12</td>
<td>H4</td>
</tr>
<tr>
<td>13</td>
<td>H4</td>
</tr>
<tr>
<td>14</td>
<td>H4</td>
</tr>
<tr>
<td>15</td>
<td>H4</td>
</tr>
<tr>
<td>16</td>
<td>H6</td>
</tr>
<tr>
<td>17</td>
<td>H3</td>
</tr>
</tbody>
</table>

Sequence of I
Conversely think

Subjects plan a plan of the plan of the plan of the plan of the plan of the plan...specifically: (H8) and...
APPENDIX D

A SAMPLE OF THE PROTOCOL ANALYSIS ON THE LOAN-PROBLEM

Verbatim Transcript

Segment Code
1 P1: "Sally loaned $7 to Betty. But Sally borrowed $15 from Estella and $32 from Joan. Moreover, ...Which girl left with $18 more than she came with?"
2 P2: "Ok, one more time...Sally owes $7 to Betty....(repeats reading the question)."
3 M2: "Lmm...this is quite confusing."
4 M1: "Ok, I'll consider each person one at a time."
5 H2: "S..B..E..and J.. (drawing a table) "
6 P4: "Sally loans $7 to Betty, so Sally will receive $7 in the end and Betty returns $7, so plus 7 here and minus 7 here."
7 P4: "But Sally borrow 15...borrow 15...so minus 15 from Estella...so from E put +15.."
8 P4: "$32 from Joan...so minus $32 for Sally and Joan plus $32."
9 P4: "Moreover, Joan owes $3, so she must minus 3 to E, E must plus 3 and owes $7 to Betty so another minus 7 to B, B must plus 7, alright"
10 P2: "One day the girls got together and ...which girl left with $18 ? "
11 M3: "Now is easy...to add up for each one."
12 H4: "Sally, cannot because she's still in debt."
13 H4: "Er...Betty..is zero...ya...
14 H4: "Estella got $18 right, plus..."
15 H4: "and Joan has 32 minus 10, $22 to take home."
16 H8: "So looks like Estella is the girl."
17 H3: "The answer is Estella left with $18 more than she came with."

Sequence of behaviours in string of codes in Protocol 1A:

P1 —> P2 —> M2 —> M1 —> H2 —> P4 —> P4 —> P4 —> P4 —> P2 —> M3 —> H4 —> H4 —> H4 —> H4 —> H8 —> H3

Conversely the sequence of codes can be interpreted as:

Subject first read the problem (P1), reread the problem (P2), found it confusing (M2), suggested a plan to organise the given information (M1), by tabulating (H2), then examined the dimensions of the problem statement by breaking down into the table drawn (P4, P4, P4, P4...), reread part of the question (P2) then reviewed to sum up the numerical relationships (M3) by looking at specific cases of the borrowers and lenders (H4, H4, H4, H4) and making a logical deductions (H8) and finally stated the correct answer (H3). [time taken : 2 min. 15 sec.]
Influence of Computer-Generated Visuals on Word-Problem Solving

Philip Wong

Abstract

This study investigated the effects of different visual strategies in the solving of word problems in computer-based lessons. There were three visual treatments, namely, visual-supplied, self-generated, and no-visual. Together with this, students were given two option-control treatments of programme-control where subjects solved ten problems, and learner-control where subjects solved six problems and were then allowed to choose up to ten additional problems.

Primary four students (n = 138) from two schools were randomly assigned to treatments in a (3 x 2) factorial design. To complete the experimental task, subjects worked individually at the computer for three sessions, each lasting approximately forty-five minutes. During the lessons, subjects were shown examples and prototype solutions, and solved word problems. A delayed post-test was administered a week later. Using on-task scores as the dependent variable, no visual treatment effect was detected. The self-generated group's post-test mean scores were significantly higher than those of the other two visual groups. Subjects in the learner-control group attempted five more problems than subjects in the programme-control group but the post-test scores were lower in the learner-control group. Results from this study suggest that requiring students to generate their own visuals is an effective strategy for solving word problems. Although students in the learner-control group attempted additional problems, their additional exposure did not result in higher post-test scores.

Descriptors: Word problems, computer-assisted instruction, visualisation, elementary school students, Mathematics instruction, problem solving.

The ability to solve word problems is one of the most important objectives in the study of mathematics. Yet teaching word problem solving is recognised to be difficult. One of the reasons for this difficulty is that finding appropriate solutions to word problems is a complex task and indications of this complexity can be gleaned by reviewing studies which report that students score lower on word-problem solving than on computation (NAEP, 1979). To counteract the difficulty of solving word problems, many mathematicians and teachers have proposed untested methods to help students. Of the many hints, strategies and methods, the use of visuals is one that is widely endorsed. However, very few empirical studies have shown positive results on the effectiveness of self-generated visuals for solving word problems in elementary schools.

With the availability of microcomputers in schools and in classrooms, the emphasis on the use of computers for word-problem solving has shifted from process research to classroom application. One distinct advantage of computers over other teaching materials is their ability to offer individualised adaptation to student needs. For instance, an instruction (C) program may be given to students in the selection of one or more of the above strategies to assist them in the solution of one or more word problems. Although the self-generated treatment did not result in significantly higher post-test scores for the learner-control group, it is possible that students may have adopted additional strategies which they did not use before. Further studies are needed to explore these possibilities.

Visual Representation

Understanding and using visual representations are important initial steps in the solution of word problems (Lesh and Landau, 1984). In this initial study on visual word problem solving, we have emphasised various techniques for representing situations such as rewording the context, deriving an external representation (pictures and diagrams) and an internal (mental) representation (Lesh and Landau, 1984). The former may help students to support their understanding of the problem while the latter may lead to a more refined solution. Kaufman (1981) suggests that students who are trained to support their thinking with visual representations show better performance than those who do not use such aids.

1993 Vol.13, No.1 76
over other traditional instructional modes is their ability to offer individualised instruction by adapting instructional materials to suit individual needs. For example, in computer-based instruction (CBI), some levels of control can be given to students to allow them to make choices in the selection of additional instruction. There is evidence that increased control promotes feelings of self-efficacy and self-determination and assists students to become independent and responsible learners (Gay, 1986). Many learner-control studies have investigated the influence of one or more individual characteristics such as cognitive styles, learning styles and ability levels on learner-control strategies in CBI. But extraneous factors such as social and cultural climate, may also determine how well these strategies work. No studies, however, have looked into other characteristics not under the control of the individual that could influence learner-control strategies.

Visual Representation in Word Problems

Understanding and creating internal representation of a problem situation forms an important initial step toward successfully solving word problems. Building students’ competence in this initial step can help them be better problem solvers and courses in problem-solving have emphasised this aspect and have taught various techniques of representing problem situations such as rephrasing problem situations, rewording the problem situation into another context, deriving equations, and drawing visuals (pictures and diagrams).

A number of theoretical propositions have been proposed to account for the importance of drawing visuals during word problem solving. Lesh and Akerstrom (1982) believe that forming an external representation by drawing visuals helps young problem solvers refine their mental representations. The iteration process of forming internal (mental) and external representations may lead to the generation and selection of more refined representations and, thus, to the solution. Kaufmann (1985) cites research studies to support his theory that problem solvers process a problem visually when the problem task is new and novel, but, if they are familiar with the task, will switch to verbal processing.

Although theoretical propositions indicate that drawing visuals is helpful, this procedure is not frequently practised by many people. However, there is evidence to show that expert problem solvers draw visuals to represent problem situations more often than novices do. Heller and Greeno (1979), after reviewing a number of studies that investigated word-problem solving in arithmetic, physics, and thermodynamics, note that experts construct diagrams whenever they are useful while novices rely mainly on verbal statements. On the other hand, children do not draw visuals or diagrams to help them solve word problems. Ekenstam and Greger (1983) conducted a large scale survey of sixth-grade children’s strategy in solving word problems. In one of the subtasks, students were asked to assess the correctness of a worked answer to a word problem. Most students concentrated on determining the accuracy of the calculation and very few bothered to check on the set-up of the problem. Hardly any students drew visuals to help in the solution.

Effects of Visuals on Word-problem Solving

It is well known that visuals lend to better performance than words in paired-associate learning because of the dual channels for encoding information (Paivio, 1971). Visuals have also been incorporated in textual materials to help in comprehension and to aid in recall. However, in word-problem solving, the functions of supplied visuals are not clear. It has been hypothesised that visuals help students: (a) to conceptualise the problem, (b) symbolise the problem situation through familiar images, and (c) to understand the problem better (Campbell, 1984).

The literature on the effectiveness of visuals in solving word problems has produced contradictory results. Threadgill-Sowder and Sowder (1982) in their review of literature
reported several studies which showed that young students performed better when word problems were accompanied by visuals. Most of the studies reviewed involved subjects solving word problems in a testing situation without any extra instruction or practice in using the visuals to solve the problems. While visuals are effective in group testing situation, they are not when students are tested individually or when time constraint is imposed. Further studies by Sowder and associates (Moyer, Sowder, Threadgill-Sowder and Moyer, 1984; Threadgill-Sowder & Sowder, 1982; Threadgill-Sowder, Sowder, Moyer, & Moyer, 1985) have consistently shown that students obtained higher scores when word problems are presented in a visual format than in the verbal format. Moyer et al. (1984) suggested that word problems presented visually help to reduce reading-related working memory overload, recall similar past experiences and link visuals to appropriate schema, provide another channel to enable students to understand the problem situation, and organise the problem information to facilitate the selection of appropriate schema and the recall of the data for computation.

While word problems accompanied with visuals help to enhance word-problem solving performance, the effects of pupils' self-generating visuals are not so clear. Three studies showed that some form of training in the use of visuals in word-problem solving heightened performance (Canner, 1979; Nelson, 1974; Wersan, 1981). One study showed no difference in performance when visuals were generated by students (Wersan, 1981), and another showed that it was difficult for students to generate pictorial representations of the problem statement without training (Walter, 1984).

These contradictory results from different studies can perhaps be explained by the variability of the design. Length of treatment, age of subjects, type of problems, and the type of covariates, are some factors that will affect the results of the studies. The ability to generate visuals and to use them depends not only on the age and ability of students, but also on the type of word problems and the type of visual tasks involved. Word problems involving complex concepts (e.g., proportion and commutativity) as used in Wersan's and Walter's studies, can influence the effectiveness of the method of self-generation of visuals.

Learner Control

One exciting feature of computer-based instruction is its ability to offer various types of options for learners. Options such as selecting the amount and type of materials, specifying the type of instructional strategies, and choosing the amount of instructional support are commonly incorporated into CBI materials; the instructional effectiveness of these options has been investigated (Kinzie, 1990).

It is generally believed that it is advantageous to allow students to exercise control in the learning process. By allowing students to make decisions during the learning process it will improve achievement by increasing motivation, reducing anxiety, and improving attitude toward learning. Most research studies in CBI have not provided support for this notion and have found negative correlation between control and achievement (Kinzie, 1990). But under certain conditions, for example, when provided with advice about their learning progress, students are able to judge their own learning ability (Tennyson & Buttrey, 1980).

Students can be given the choice of selecting the types or amount of materials they deem most suitable for their needs. But most students are not good judges; this is especially true for young students. While it is important for less able students to obtain more instructional support, they did not seek it when allowed to control lessons (e.g., Carrier, Davidson, & Williams, 1985). Fisher, Blackwell, Garcia, and Greene (1975) investigated the effects and the pattern of selection of problems by fourth and fifth graders during an arithmetic CBI lesson. Some students consistently chose the easiest problems; others chose the more difficult problems. Tennyson & Buttry (1980) found that younger students, when given complete control of lessons, tended to choose an amount of material insufficient to master concepts and to terminate lessons too early. This resulted in poor achievement scores. Students who received insufficient CBI instruction did not perform as well as students who received additional instructional support.

Besides age and ability characteristics of pupils, such as cultural background, environment can also affect their additional instructional needs. Numerous cross-cultural studies have shown that Asian students are more positive toward self-generated visuals and are more likely to spend more homework time with the same number of school instruction (Walter, 1984). Students from high and cultural environments are more cited as the most suitable for CBI instruction, and these factors are related to the amount of additional instruction they receive, which can influence their performance and learning process. Furthermore, under some conditions, computer-based instruction may be more effective when provided by a teacher who is familiar with the learner's needs and abilities.

The Current Study

One purpose of this study was to investigate the effectiveness of using visuals to help students solve word problems. The main focus was to determine the methods of using visual aids to improve the self-generated self-help strategies and the computation abilities of students.
achievement scores. On the other hand, students who received feedback about their progress toward the mastery of concepts, stayed on-task long enough to master the concepts. College students are better judges of the amount of practice and were able to judge and select an appropriate quantity of practice items required to master mathematical concepts (Judd, Bunderson & Bessent, 1970). But they overpractised and, consequently, took more time to complete a module. The ability to be good judges, thus, will depend on the age and ability level of the students. Hannafin (1984) suggests that older and more academically capable students may have more refined cognitive strategies and are able to apply them to instruction. But younger children and less able students do not have the refined cognitive strategies or the self-evaluation skills needed to monitor their own progress during instruction.

Besides age, ability, and personal characteristics (eg level of motivation), factors such as cultural, social, and educational environment can influence students opting for additional instructions during the lesson. Numerous cross-cultural studies have indicated that Asian students (eg Japanese and Taiwanese) are more positive toward school learning and spend more hours on homework and on after-school instruction (Walberg, Harnisch, & Tsai, 1986). Social factors such as parental support and cultural emphasis on education have been cited as the main contributors to the higher mathematics achievement of Asians. Perhaps, these factors and the extra instruction students receive can influence students' task perseverance and on their selection of the amount of instruction.

The Current Study

One purpose of the current study was to investigate the effectiveness of the strategy of using visuals to represent problem situation for solving word problems. It compares two different methods of using visuals: supplied visuals versus self-generated visuals. Reading ability, computation ability and spatial visualisation are some individual difference factors that have been found to be correlated with students' word-problem solving performance and these factors will be used as covariates.

Another purpose was to extend the range of variables of how learners operate when given options to select additional instruction. Do Singaporean students select additional instruction when given the choice? If they do, do the students benefit from extra instruction?

Method

Subjects

The sample was 138 Primary Four students from two primary schools in Singapore. These students follow a six-year primary curriculum consisting of studies in the first language (English), a second language (Chinese, Tamil or Malay), science, mathematics, social studies, and aesthetics. At the end of their sixth year, students sit for a national examination, the Primary School Leaving Examination, and if successful, proceed to secondary education. Few students have had any regular exposure to computer work.

Using a random set of numbers, students names were drawn from the school register and were assigned to six cells of a 3 x 2 design (three levels of visuals by two levels of learner-control). Due to various reasons, 12 subjects were eliminated leaving 21 subjects per cell.

Instrumentation

The Comprehensive Tests for Basic Skills, Form U, Level F (CTB/McGraw-Hill, 1981) was used for testing students' reading comprehension skills and the Form U, Level G for the computation skills. The Punched Holes Test (Wilson, Cahen, & Begle, 1968) for elementary school students was used for assessing students' spatial visualisation ability. This test, an adaptation from the adult ETS Paper Folding Test (French, Ekstrom & Price, 1963 ), was modified for children by the National Longitudinal Studies of Mathematical Abilities (NLSMA) Group in 1967.
The word problems were drawn from previous research projects (Sowder, Threadgill-Sowder, Moyer & Moyer, 1984; Yancey, 1981). These word problems were similar to those found in school textbooks. Basically, each problem consisted of two parts. In the first part, information with numerical values describing the problem situation was presented and in the second part, a question was posed. Two criteria were used for the selection of word problems: (a) the problems must be appropriate for third to fifth graders and (b) the problems must involve double-step arithmetic operations (e.g., multiplication and then addition). The reliability coefficient (KR 20) of word problems used during lessons was 0.79, and for the ten word-problem post-test was 0.78.

Treatment

There were three visual treatments: visual-supplied, self-generated, and no-visual; and two option-control treatments: learner-control and programme-control.

Visual-supplied treatment

The visuals appeared for all the problems and examples in the three lessons. Simple iconic visuals with labels representing each problem were drawn. For example, to show five thousand three hundred trees, a group of trees were drawn with the label "5300 Trees." There were no attempts to represent the actual number of items in the visuals. Some problems required only two visuals to represent the problem while some required three. Figure 1 illustrates a sample visual screen for the word problem.

Self-generated treatment

For this treatment, subjects had to draw their own visuals on sheets of paper provided. To help subjects in drawing visuals, examples in the lessons were accompanied with visuals (same visuals as those in the visual-supplied treatment). Specific instructions on how to draw visuals were provided. Subjects were told to draw a group of articles and attached a label to them. Instead of visuals appearing, a prompt "Draw pictures to help you" was presented. Also to encourage them to draw visuals, subjects were told that they would be awarded an extra point for visuals drawn in each problem.

No-visual treatment

This was the control treatment. The treatment procedure was the same as the visual-supplied group except that visuals were not presented. To maintain consistency, the keyboard was locked for fifteen seconds before subjects were allowed to key-in their answers to the word problem.

Learner-control treatment

For Lessons 2 and 3, subjects were allowed to choose the number of problems they would see. In order to prevent students from not solving any problems, they had to work out three word problems before being given the option to select more problems. On completion of the three problems, subjects were asked whether they would like to try another problem. If positive, another problem was presented. If negative, the lesson terminated and the total scores were shown. This choice was presented until the total number of problems reached eight.

Problem Two

A Christmas tree farmer has 5460 trees. He cuts down 1230 trees. Next day, he cuts down 500 more. How many trees are standing now?

5460 trees  cuts 1230  cuts 500 more =

Type in your answer and press <RETURN>.

Figure 1: Screen Display - Visual-supplied Treatment
At that point, the lessons terminated and subjects were informed of their scores. The maximum number of problems chosen for Lessons 2 and 3 was eight (three mandatory and five optional) and the minimum number was three (three mandatory and no optional problems).

Programme-control treatment
Under this mode, subjects did not have a choice in selecting the number of problems to solve. Instead, each subject would have to solve four problems in Lesson 1 and five problems each in Lessons 2 and 3.

Procedure
Before the collection of data, subjects were given the reading, computation, and spatial visualisation tests. Subjects attended three CBI lessons, each lasting from thirty to forty-five minutes. The first lesson was an orientation one and scores were not collected. The lessons were delivered on Apple Ile computers in the schools' computer laboratories. Each computer station was numbered so that subjects could be randomly assigned to the stations and thus, to different treatments. Within each session, all of the six treatments were present.

A typical task during the lesson consisted of the following sequence. A word-problem situation was presented on the screen for students to read. They were advised to read the problem carefully. After a lapse of five seconds, instructions were provided for subjects to press the return key for the presentation of visuals. The visuals appeared followed by a time delay of ten seconds during which the keyboard was locked. After the pause, the problem question appeared. This procedure was adopted to prevent impulsive actions and also to encourage subjects view the visuals. Subjects then typed in their answers which were judged and if wrong, were shown the correct answer. For each correct answer one point was added to the On-Task Answer score. On pressing the return key, five options for the selection of process operations were presented (Figure 2).

Subjects selected one process operation that was perceived would yield the correct numerical answer. If successful, one point was added to the On-Task Process score, and they would proceed to the next problem. If unsuccessful, the prototype solution, which explained how the answer was obtained, was presented. They would then proceed to the next problem. The first lesson lasted for approximately thirty to forty-five minutes. For Lessons 2 and 3, subjects were required to sit at their assigned microcomputer stations which were loaded with the programmes they were working on previously. Subjects who completed their lessons ahead of others were sent back to their class. A delayed post-test consisting of 10 word problems were administered a week after the treatment.

Problem Two
A Christmas tree farmer has 5460 trees. He cuts down 1230 trees. Next day, he cuts down 500 more. How many trees are standing now? Choose only 1 method.

Type a, b, c, d or e and <RETURN>

a) 5460 - 1230 then subtract 500  
b) 5460 - 1230 then add 500  
c) 5460 + 1230 then subtract 500  
d) 5460 - 1230 then multiply 500  
e) 5460 + 1230 then add 500

Figure 2: Screen Display - Selection of Process Operations
In summary, the following sets of data were recorded:
(a) Individual-difference variable scores of computation, reading, and spatial visualisation
(b) On-Task Answer score for correct answer during the computer lessons
(c) On-Task Process score for correct selection of process operation during the computer lessons
(d) Post-test Answer score for correct answer to problems in the post-test
(e) Post-test Process score for correct process operation in the post-test
(f) Number of problems attempted by subjects in the learner control group.

Results

There were two different option-control groups. Subjects in the learner-control group solved six word problems before they were allowed to choose up to ten additional problems (total maximum number of problems was 16). Subjects in the programme-control group solved a fixed number of problems (10). The number of problems attempted by the subjects in the learner-control group was 14.1 for the no-visual group, 15.3 for visual-supplied group, and 14.9 for self-generated group. Tables 1 shows the means and standard deviations of the On-Task Answer and On-Task Process scores. To compensate for unequal number of problems attempted, the number of problems attempted was used as a covariate to adjust the means of the on-task scores.

Table 2 shows the post-test scores. Subjects in the learner-control group while attempted an average of five more problems than subjects in the program-control group their On-task performance for the two groups was not significant. Univariate analysis of the Post-test Answer showed that the program-control group means were significantly higher than those of the learner-control group (F = 4.47, p < .05). Post-test Process scores were also higher for the program-control group than the learner-control group (F = 5.26, p < .05).

Table 1: Means and Standard Deviations for On-Task Answer and On-Task Process

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>No visual</th>
<th>Visual-supplied</th>
<th>Self-generated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Answer</td>
<td>Process</td>
<td>Answer</td>
<td>Process</td>
</tr>
<tr>
<td>Learner Control</td>
<td>M*</td>
<td>4.81</td>
<td>7.33</td>
<td>7.19</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.01</td>
<td>2.58</td>
<td>3.19</td>
</tr>
<tr>
<td>Programme Control</td>
<td>M**</td>
<td>3.95</td>
<td>5.38</td>
<td>4.62</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.01</td>
<td>2.33</td>
<td>3.31</td>
</tr>
</tbody>
</table>

* maximum possible = 16
** maximum possible = 10

Table 2: Means and Standard Deviations for Posttest Answer and Posttest Process

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>No visual</th>
<th>Visual-supplied</th>
<th>Self-generated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Answer</td>
<td>Process</td>
<td>Answer</td>
<td>Process</td>
</tr>
<tr>
<td>Learner Control</td>
<td>M*</td>
<td>3.10</td>
<td>5.24</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.95</td>
<td>2.58</td>
<td>2.20</td>
</tr>
<tr>
<td>Programme Control</td>
<td>M*</td>
<td>4.57</td>
<td>6.48</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>3.44</td>
<td>3.59</td>
<td>3.21</td>
</tr>
</tbody>
</table>

* maximum possible = 10

Reading, computation, and spatial scores were used in the four multiple regression models: On-Task Answer, On-Task Process, Post-test Answer, and Post-test Process. The purpose of this was to reduce the amount of correlation among the variables controlling the On-Task and Post-test scores. Using this procedure, each variable was entered into the appropriate regression equation without violating the assumptions of the analysis. When all independent variables were entered, the number of independent varables for each equation was used. Spatial, number, and computation variables were two variables in the equations.
Reading, computation and visualisation scores were used as independent variables in four multiple regression equations with On-Task Answer, On-Task Process, Post-test Answer, and Post-test Process as dependent variables. The purpose of this set of analyses was to determine the amount of variance these independent variables contributed to the problem-solving scores. Using the forward regression procedure, each variable was added to the regression equation without removing the other variables already present. Table 3 shows the values of $R^2$ when all independent variables entered into the equations. Nearly one-third of the variance of the post-test scores were accounted by the independent variables. To select the independent variables for covariates, a stepwise regression was used. Spatial visualisation and computation were two variables (Table 3) that accounted for most of the variance in the on-task and post-test scores. When these variables were in the regression, reading scores did not contribute to any significant change in the variance and thus was not used as a covariate.

Multivariate analysis of covariance (MANCOVA) was used to analyse the data with On-Task Answer and On-Task Process as combined dependent variables. Computation was used as a covariate variable while spatial visualisation was included as a factor to assess aptitude-treatment interaction (ATI) effects. The main effect for visual treatment was found to be not significant $F(4, 230) = 1.35, p = .25$. After adjusting the means for total number of problems attempted, there was no significant difference between learner-control and programme-control. There was no ATI effect between spatial visualisation and visual treatments.

Table 3: Regression on Dependent measures

<table>
<thead>
<tr>
<th>Dependent Measures</th>
<th>Spatial</th>
<th>Computation</th>
<th>Reading</th>
<th>Visualisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-Task Answer</td>
<td>0.22</td>
<td>0.24*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>On-Task Process</td>
<td>0.38</td>
<td>0.61*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Post-test Answer</td>
<td>0.32</td>
<td>0.47**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Post-test Process</td>
<td>0.26</td>
<td>0.29**</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* $p < .05$ ** $p < .005$

Table 4: MANCOVA for Post-test Answer and Post-test Process

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pillais-Bartlett</th>
<th>F</th>
<th>$(df_1, df_2)$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>0.13</td>
<td>4.26</td>
<td>(4, 232)</td>
<td>0.002</td>
</tr>
<tr>
<td>Option-control</td>
<td>0.043</td>
<td>2.60</td>
<td>(2, 117)</td>
<td>0.078</td>
</tr>
<tr>
<td>Visual x Option</td>
<td>0.16</td>
<td>1.65</td>
<td>(4, 232)</td>
<td>0.160</td>
</tr>
<tr>
<td>Covariates</td>
<td>0.40</td>
<td>14.96</td>
<td>(4, 236)</td>
<td>0.001</td>
</tr>
</tbody>
</table>
In post-test scores, computation and spatial visualisation scores were used as covariates with the Post-test Answer and Post-test Process as combined dependent variables. The main effect for visual treatments was significant (Table 4). Both Post-test Answer and Post-test Process scores showed highly significant visual treatment effects. Simple contrast showed that the means of the self-generated group were significantly higher than both the visual-supplied (p < 0.05) and no-visual groups (p < 0.05) but not between the supplied-visual group and the no-visual group.

Discussion

It was interesting to note that subjects in the visual-supplied group, when provided with visuals during the lesson, had the highest proportion of correct on-task answers and process operations. In the written post-test, where no visuals were provided, they had the lowest scores. Perhaps the presence of visuals during the lesson and the absence of visuals in the post-test could have caused this pattern of results. A number of explanations could account for this.

First, there was no direct instruction during the lessons for subjects in the supplied-visual group to "learn" the strategy of using visuals. Subjects were unable to internalise the strategy of representing the word problems visually, and thus not able to transfer the strategy when visuals were not provided. One kind of evidence that led to support this proposition was the small number of subjects (only two subjects) in the visual-supplied group who attempted to draw visuals during the post-test. To provide better evidence for this proposition, other research methods, such as interviews or subject protocols, should be employed. Unfortunately, the design of this study did not include these methods and one can only postulate about this proposition.

Second, visuals supplied during the lessons might provide another source for processing the information as hypothesised by Paivio's (1971) dual coding model. For example, subjects who had difficulty in processing the textual statements verbally could do so visually. The presence of visuals "might call into play additional internal resources that the solver does not use spontaneously" when only verbal problems are presented (Threadgill-Sowder et al., 1985, p. 56). The visuals, for example, could have helped subjects to link the verbal statements with the visuals and make the problem more meaningful. The absence of visuals during the post-test could deprive subjects of this extra source.

Third, the presence of visuals could help subjects in understanding the problem by activating relevant past experiences and schemata. The activation of various linked knowledge structures could help in the construction of an appropriate semantic internal representation of the problem (Moyer et al., 1984).

Fourth, the visuals might have inadvertently provided additional help to subjects by organising and interpreting the information of the problem and indirectly provided hints to the selection of arithmetic operations (Threadgill-Sowder et al., 1985). For example, when a second visual contained less number of items than the first visual, it could imply subtraction or division. When visuals were no longer provided in the post-test, these extra functions of visuals were removed and could have affected subjects' performance in solving word problems.

Post-test scores showed significant visual effect with the self-generated visual group having performed better than the other two groups. Consistent with other observations (e.g., Lester, 1985), young children can be taught to use heuristic strategies in solving word problems. In the current study, subjects in the self-generated group were able to draw and use visuals for solving the word problems and in so doing, improved their performance. The process of generating their own visuals was a more effective strategy than being supplied with visuals as shown by the higher post-test scores of the self-generated group. The active process of generating visuals took a longer time, but once the strategy was learned, it was applicable even after a delay of one week.

While past studies on learner control in CBI indicated that young students tend to terminate instruction too early (Buttrey, 1980), the solving of six word problems stopped but instead did so for additional work on the possibility of further result. First, more experience with the computer might have helped the subjects to ask for additional help. Second, it was similar to real practice. Third, these factors might have inadvertently influenced the subject's performance. Numerous cross-cultural studies have shown that Asian students are more positional with much less emphasis on individual achievement. These factors could have influenced the number of additional problems solved.

The number of subjects who failed to complete the lessons did not correlate with the post-test scores. In fact, the Post-test scores of the control group were higher than in the experimental group. The self-generated group also showed a significant improvement in the post-test, the additional help supplied by the program-control group also provided an additional source for solving the word problems during the post-test. The additional help might have caused this pattern of results. First, the presence of visuals could help subjects in understanding the problem by activating relevant past experiences and schemata. The activation of various linked knowledge structures could help in the construction of an appropriate semantic internal representation of the problem (Moyer et al., 1984).

In conclusion, the results of this study suggest that the provision of visuals during lesson-time can improve learning outcomes. Visuals can provide additional sources for processing the information as hypothesised by Paivio's (1971) dual coding model. The presence of visuals "might call into play additional internal resources that the solver does not use spontaneously" when only verbal problems are presented (Threadgill-Sowder et al., 1985, p. 56). The visuals, for example, could have helped subjects to link the verbal statements with the visuals and make the problem more meaningful. The absence of visuals during the post-test could deprive subjects of this extra source.

Numerous cross-cultural studies have shown that Asian students are more positional with much less emphasis on individual achievement. These factors could have influenced the number of additional problems solved. The number of subjects who failed to complete the lessons did not correlate with the post-test scores. In fact, the Post-test scores of the control group were higher than in the experimental group. The self-generated group also showed a significant improvement in the post-test, the additional help supplied by the program-control group also provided an additional source for solving the word problems during the post-test. The additional help might have caused this pattern of results.
instruction too early if allowed (Tennyson & Buttrey, 1980), that was not the case here. After solving six word problems, subjects could have stopped but instead, they continued and opted for additional word problems. There are a number of possible factors that could account for this result. First, most of the subjects had limited experience with CBI and the novelty of working with the computer might have prompted them to ask for additional problems. Second, subjects might have believed that problem solving exercise was similar to rote learning and opted for more practice. Third, cultural and social factors could influence the subjects to attempt more problems. Numerous cross-cultural studies have indicated that Asian students (eg Japanese and Taiwanese) are more positive towards school learning and spend more hours on homework and on after-school instruction (Walberg et al., 1986). Social factors such as parental support and cultural emphasis on education have been cited as the main contributors to the higher mathematics achievement of Asians (Miura, 1987). Perhaps, these factors could influence subjects to attempt more problems believing that attempting more would improve their performance.

The number of problems attempted by students did not correlate with post-test scores ($r = -14$). In fact, the Post-test Answer for the program-control group were significantly higher than the learner-control group (means = 4.60 and 3.75) and in the Post-test Process, a similar trend was also observed (means = 6.63 and 5.56). This was observed even though subjects in the learner-control group attempted more problems than the program-control group. Exposure to more word problems during the lessons did not help students in the post-test. There was no positive transfer of the additional experience from the lessons to post-test performance. Instructional strategies which use worked examples for analogies and for comparison with other problems may facilitate schema learning. A worked example is a sample problem with its solution worked out to illustrate how the solution process is carried out. In this study, worked examples were provided at the beginning of each lesson. Prototype solutions provided at the end of each unsuccessful attempt at answering the word problem were also worked examples. For students in the learner-control group, they were shown more prototype solutions. However, they were unable to use their added exposure of prototype solutions and transfer the experience to the post-test. This may be due to the lack of specific instructions to “learn” from the prototype solutions. The fourth grade students in this study were not able to learn by themselves without this extra instructional procedure. While it is important for students to develop metacognitive skills in problem solving (eg Lester, 1985), subjects in this study were unable to do so on their own, as reflected by the lower mean scores of the learner-control group. Subjects in the learner-control group attempted more problems and had more exposure to instructions. However, during the lessons, subjects were not told to “learn” from the prototype solutions. It was assumed that subjects would learn and benefit from the extra instruction when they saw the prototype solutions. The assumption was wrong. Other findings have also shown that young students when left on their own, were unable to generate their own learning strategies (eg Biehler and Snowman, 1986).

Conclusion

Although Asian students may show different learner control patterns in computer-based instruction, students need more guidance to benefit from any extra exposure of instructional materials. The results of this study suggest that requiring students to generate their own visuals could be an effective strategy for solving word problems.
REFERENCES


Measuring Motivation to Learn Chinese and English Through Self-Reported Feelings and Behaviours

Soh Kay Cheng

Abstract

The socio-educational approach to measuring language motivation has its focus on social attitude rather than language-relevant feelings and behaviours in the classroom context. The development of two scales for measuring elementary students’ motivation to learn Chinese and English reported here is an exploration of a psychology-based approach which may be more meaningful to language teachers. The scales share common items and cover both affective and behavioural aspects of language learning. Very high internal consistency was found for both scales. Discriminant validity was established with reference to student backgrounds, including overall academic performance, gender, language examination grade, home language, and self-evaluation of language ability.

Descriptors: Language motivation, bilingualism, measurement.

INTRODUCTION

Teachers know well that students who are motivated to learn pay more attention and put in more effort and these make teaching more rewarding with better pupil behaviour in class and better achievement subsequently. This is true of any subject in the school curriculum and is, of course, true of language instruction. The causal effect of motivation on language learning is most emphatically pronounced by Gardner (1985, p 85), “The prime determining factor in language learning success is motivation.” Even though Carroll did not include a motivation measure in the MLAT (Carroll & Sapon, 1959), as did Pimsleur (1966) in his foreign language aptitude inventory, Carroll (1981, p 84), however, takes motivation for granted when defining language aptitude, thus, the individual may be thought of as possessing some current state of capability of learning that task - if the individual is motivated, and has the opportunity of doing so. (Emphasis added)

Taking the earlier research on language motivation (Gardner & Lambert, 1972) further, Gardner (1980) expanded the concept of language attitude/motivation index (AMI) to include three components: integrativeness, attitudes toward the learning situation, and motivation. AMI, while relatively independent of aptitude as measured by MLAT which correlates as expected with language grade, was found to be persistently correlated with French achievement for grades 7-11 students across several school areas. It is of note that, the first component of AMI, integrativeness, is akin to the integrative motivation of the earlier conception (Gardner & Lambert, 1972), whereas the other two components of AMI, attitude and motivation, have greater classroom relevance since they, for instance, pertain to affective reactions toward the course and the teacher and effort expended toward learning the language. Notwithstanding the debate on the validity of the socio-educational theory (Au, 1988; Gardner, 1988), there is no denying the importance of affect in language learning.

The need for a more psychology based motivation, which takes motivation for granted, is becoming more apparent with the students do as much as the teacher does about the language, as true of any subject in the school curriculum, has been voiced by Gardner (1980) when they describe the lack of this approach direct classroom pedagogical implications.

It is probably fair to describe a student who learns language with more mature and hence better understanding of their language-relevant feelings and behaviours, Hart & Soh (1988) further, requires one to start the language motivation research on language aptitude inventory scales that are more useful for research and practical purposes. Moreover, the affective or attitude aspect deserves as much attention as the motivation is well established and hence better useful for research on language aptitude. In view of the need for the need for a more psychology based motivation is well established and hence better useful for research on language aptitude inventory scales that are more useful for research and practical purposes.
The Lambert-Gardner approach to language motivation is social-psychological in orientation and concerns itself almost exclusively with social attitudes toward second-language acquisition and second-language community. While the educational and social significance of such attitudes are important in their own rights, the lack of this approach to language motivation in direct classroom relevance and immediate pedagogical implication is not to be denied. The need for a more practitioner oriented and psychology based approach to language motivation, which has a closer link with what the teacher does in terms of pedagogy and what the students do as classroom activities for and about the language to be learned, has recently been voiced by Crookes and Schmidt (1991) when they describe succinctly the concern of motivation among practising teachers:

It is probably fair to say that teachers would describe a student as motivated if he or she becomes productively engaged in learning tasks, and sustains that engagement, without the need for continual encouragement or direction. 

Moreover, for obvious reasons, much of the research on language motivation has been done with more mature students of high schools and colleges than with younger children in elementary schools, although it is a widely accepted notion that to be good in a language requires one to start young. Thus, for measuring language motivation of younger school children, scales that are more relevant to their experience and hence better understood by them would be useful for research and, ultimately, instructional purposes. Moreover, while the importance of the affective or attitudinal aspect of language motivation is well-recognised, the behavioural aspect deserves as much attention if the measures are to be of practical value to classroom teachers who may then guide the students in changing their language-relevant behaviours, which understandably is more amenable to change than language-relevant attitude.

In view of the need for a motivation measure to be anchored in the classroom context and to take into account both students' affects and behaviours, Hart's (1981) conception of a motivational and an informational components is relevant. She attempted at developing a factor scale of motivation for classroom learning, though not specific to language learning. The five dimensions defined are: preference for challenge versus preference for easy work, curiosity and interest versus teacher approval, independent mastery attempts versus dependence on the teacher, independent judgement versus reliance on the teacher’s judgement, and internal versus external criteria for success/failure. Hart’s study shows that five dimensions can be established, where the first three are motivational in nature while the last two are cognitive-information. When considering the possible structure of the language motivation scales to be developed, it was felt that Hart’s dimensions would provide a conceptual scaffolding for the scales, with slight modification by expanding the sixth dimension, criteria, into two for social motivation and self-motivation.

In Singapore, the bilingual education policy requires Chinese-Singaporean children to learn Chinese (Mandarin) and English right from first grade. As would be expected, related to variation in language environment at home and through the family’s network, some students do well in one but not the other language, while only a small proportion is expected to be able to, and actually show, mastery of both languages; it has been estimated that only about 10% of elementary school students are able to attain balanced bilingualism, that is, in Singapore terminology, “to offer English and the mother tongue at first language level” (Yip et al., 1991). In Singapore, as in elsewhere, attitude/motivation is frequently cited as an important contributing factor to explain success or the lack of it in language learning. Thus, the students’ language motivation deserve research effort and instruction attention. This implied of course, the need for valid, reliable and user- (teacher-) friendly measuring devices.

Incidentally, maybe of side interest, most language motivation research deals with two related languages (eg English and French, Spanish and English, etc.). The languages of concern here are from two distinctly different
linguistic families, and hence it would be interesting to see if the structures of language motivation in this situation are similar.

English was considered a more suitable language for presenting the scales to students for their response, instead of using both languages for their respective versions. This decision was taken because the students in general may not have adequate reading ability in Chinese to understand the items if they were in Chinese. Moreover, students with low motivation to learn Chinese language are most likely those who are weak in this language but strong in English. Furthermore, when the scale is ultimately used for evaluation, classroom-based research and counselling purposes, information gathering may be undertaken not by the Chinese language teacher but the form teacher, the departmental head or even the principal who may find an English version more meaningful and convenient to use.

With the above considerations, a scale was designed which has common items that can be used for measuring motivation to learn Chinese (Mandarin) and English after the necessary adjustment for the referenced object (the language). This article reports the attempt to develop such a scale.

Chinese Language Motivation Scale (CLMOTS)

The Scale

The draft version of the CLMOTS consists of 30 items. The items depict feelings the students may experience and behaviours they may display in connection with the learning of Chinese, especially in the context of the classroom. Five items were intended for each of the six aspects of language motivation mentioned earlier, following largely Hart's (1981) conceptualisation: challenge, curiosity/interest, mastery, judgement, social motivation and self-motivation. When considering motivation at the classroom level, Crookes and Schmidt (1991) highlight preliminaries, activities, feedback, self-evaluations, and material as specific components of language motivation. These were taken into account in developing the scale for the present study. The items were presented with a four-point scale indicating the degree of applicability of the statements to the pupil, ranging from Always true, True, Not true and Always not true. When the draft CLMOTS was administered, the teacher read the test instructions and the items aloud while the students followed silently and responded accordingly. Data collection was carried out in the early part of the school year.

The Sample

The trial of the CLMOTS involved four classes of Primary Five students from a school with a strong Chinese tradition located in a densely populated public housing estate. After exclusion of scripts with incomplete data, there were 123 students whose responses were analysed for CLMOTS. These students came from classes of a wide range of overall academic standing with slight preponderance from the higher end (55.1%). However, this distribution does not deviate significantly from equal distribution of high (classes A and C) and low (classes E and F) achievers (chi-square 1·829, df 1, p>0·05). There is also a slight preponderance of females (54·5%) but this does not deviate significantly from an equal distribution of sexes (chi-square 0·984, df 1, p>0·05). These students' results for the school-based Chinese language examination for the year prior to data collection was somewhat on the high side, with 44·7% in the high band, 43·1% in the middle band, and only 12·2% in the lowest band. Considering the location of the school and the Chinese language grades of the students involved, it was not unexpected that Mandarin was the home language of the great majority (78·9%). This was further reflected by the students' self-evaluation of Chinese language ability, with more than half (53·7%) rating themselves as being good or very good in the language, another one-third (33·3%) rating themselves as being average, and only a small proportion (13·0%) giving low self-ratings (Not so good and Very poor).

Factor Analyses

The 123 students' raw version of the CLMOTS principal component yielded eight unrotated factors explaining 67.07% of variance. The first factor explained 34.85% of variance, which explained the remaining variance factors explaining 14.61%, 7.53%, 6.49%, 3.27%, 2.45%, 1.45% and 0.85%, respectively. The first factor, 12 items, varied from 0·61 to 1·00 and thus, suggest its retention for further analysis.

The second retained factor, 11 items, ranged from 0·60 to 1·00 and thus, was not varied, thus: Challenge (1), Mastery (4), Judgement (1), Curiosity/interest (2), Social motivation (2), Self-motivation (2). The reliability of these 12 items was calculated using Cronbach's Alpha, yielding an internal consistency for the scale of 0·91.

Reliability

The reliability of the CLMOTS was assessed using Cronbach's Alpha. These were estimated for sex, examination results and the students as a whole. The CLMOTS has high consistency for overall sex, ranging from 0·86 to 0·90. The consistency for the students as a whole, the CLMOTS average of 0·67.
**Factor Analyses**

The 123 students’ responses to the 30-item trial version of the CLMOTS were submitted for a principal components analysis. The analysis yielded eight unrotated orthogonal factors explaining 67.07% total variance. The first factor explained 34.85% and the remaining seven factors explaining from 7.04% to 3.47%. A scree plot shows the drastic drop of proportion of explained variance after the first factor and, thus, suggest its retention and the exclusion of the remaining factors from subsequent analysis. The first factor, 12 items with the highest loadings varying from 0.61 to 0.75, were selected for further analysis.

The second run yielded only one factor formed by all 12 retained items, with loading ranging from 0.67 to 0.75 and 52.03% total variance was accounted for. The items and loadings are shown in Table 1. It can be seen that the items cover all six aspects of the original conceptualisation, though numbers of items varied, thus: Challenge 1, Curiosity/interest 1, Mastery 4, Judgement 3, Social Motivation 1, and Self-Motivation 2. Although the scale formed by these 12 items measure the students’ motivation to learn Chinese language in a broad and general manner, there is a strong element of making a conscious effort to master the language and keen concern for improving the ability in the Chinese language.

**Reliability**

The reliability of the final version of the 12-item CLMOTS was evaluated for its internal consistency using Cronbach’s alpha coefficients. These were estimated for different groups by sex, examination grade, self-evaluation and for the students as a whole. As Table 2 shows, CLMOTS has high reliability coefficients varying from 0.86 to 0.92 which indicate high internal consistence for various groupings. For the sample as a whole, the CLMOTS has a Cronbach alpha of 0.91.

**Validity**

The validity of the CLMOTS was evaluated by comparing students on their motivation score with reference to the background variables. First, the motivation means decreased with the general academic levels of the classes, with Class A scoring 40.30, Class C 38.92, Class E 38.09, and Class F 30.78. One way ANOVA show an overall difference among the groups (F=13.94, df 3, p<0.01). However, Duncan’s test shows classes C and E to be not significantly different from Class A, indicating that the CLMOTS was more effective in discriminating between low achievers and those average and higher achievers. Secondly, the mean of 34.84 for males is significantly lower than that of 38.85 for females (F=9.49, df 1, p<0.01); this indicates that girls were more highly motivated than boys in learning the Chinese language. The third comparison was made between students with high and low Chinese examination grades. As would be expected, students whose last Chinese examination results placed them in Bands 1 and 2 were motivated when compared with their peers in Bands 3 and 4. The motivation means of 39.23 and 34.68 differ significantly (F=12.74, df 1, p<0.01). The fourth significant difference was found between those who always speak Mandarin at home (37.81) and those who do not. The motivation means of 37.81 and 34.08, for the two groups respectively, are significantly different (F=5.36, df 1, p<0.05). This suggests the importance of home support in motivating students to learn the Chinese language. Fifthly, students who evaluated their own Chinese language ability most favourably (Very good and Good) obtained a motivation mean of 38.70. This differed significantly from the mean of 36.09 of those whose self-evaluation was less favourable (Average, Not so good and Very poor). Again, there is a significant difference between these means (F=7.59, df 1, p<0.01). Thus, students who saw themselves in better light were more strongly motivated to learn Chinese, as would be expected.
English Language Motivation Scale (ELMOTS)

The Scale

The draft version of the ELMOTS consists of 30 items which are exactly the same as those of the CLMOTS, except that the word “Chinese” was replaced by “English”. It will be recalled that the items depict students’ feelings and behaviour relevant to the language, with five items for each of the six aspects of motivation. The ELMOTS was administered together with the CLMOTS in the early part of the school year (February). The teacher read the test instructions and the items aloud for the students to follow.

The Sample

The same groups of students who completed the CLMOTS were administered the ELMOTS. After eliminating scripts with incomplete data, 138 were retained. Their school-based English language examination results for the past year were, not unexpectedly, somewhat lower than those for their Chinese language examination, with 40.6% in the high band, 35.5% in the middle band, and 23.9% in the lowest band. Considering these students’ home background, it was not unexpected that English was the home language of only 39.1%. Hence, less than half (44.2%) gave favourable self-evaluation rating themselves good or very good in the English language, two-fifths (39.9%) rated themselves as being average, and one-sixth (15.9%) rated themselves as Not so good and Very poor. The slight preponderance of students from the lower-end (54.4%) classes is not significantly different from an equal distribution (chi-square 1.043, df a, p>0.05) and so was the slight preponderance of female students (53.6%, chi-square 0.725, df 1, p>0.05).

Factor Analyses

As was done for the CLMOTS, the 138 students’ responses to the ELMOTS were submitted for a principal components analysis which yielded eight unrotated orthogonal factors explaining 63.11% total variance, with the first factor explaining 31.6% and the remaining seven factors explaining from 5.92% to 3.48%. When the factor pattern of the ELMOTS was compared with that of the CLMOTS, there are 11 items common to both scales among the 12 highest-loading items. The exception in the ELMOTS is item number 10 (“When I see English words on signboards, I like to know how to read them and what they mean.”) with a loading of 0.62 when compared with item number 30 (“I study for English tests even before the teacher tells us to do so.”) with a loading of 0.61. Given such a minute difference, it was decided to include item number 30 in the ELMOTS in its final version so that the scales for the two languages can have identical wordings except, of course, the language each refers to. A second run of the 12 selected items yielded only one factor with loading ranging from 0.62 to 0.85 and 51.06% total variance was accounted for (Table 1).

Reliability

As Table 2 shows, the 11 item final version of ELMOTS has high reliability coefficients varying from 0.89 to 0.92 which indicate high internal consistency for various groupings. A Cronbach alpha of 0.91 was obtained for the ELMOTS for the sample as a whole.

Validity

The ELMOTS was evaluated for its validity through comparing students’ motivation scores derived from their responses to the 11-item scale with reference to their background variables. First of all, the motivation means varied with the general academic levels of the classes: Class A 41.31, Class C 39.64, Class E 38.90 and Class F 28.78. One way ANOVA shows the differences to be significant, overall (F=29.80, df 3, p<0.01). However, Duncan’s test shows Classes C and E to not be significantly different from Class A. Thus, ELMOTS was more effective in discriminating between low achievers and the average and higher achievers. Next, the motivation means of males (34.19) and females (39.02) are significantly different (F=9.49, df 1, p<0.01); this indicates that girls were more highly motivated than boys in learning English.

<table>
<thead>
<tr>
<th>ITEMS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Factor Loading for 12 Items

<table>
<thead>
<tr>
<th>Groups</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thirdly, students’ motivation grades (Ban 1988) had a mean of 39.19 for high grades (Ban 1988), 33.35 for medium grades and 36.35. The results were significant (F=29.80, df 1, p<0.01).

A fourth analysis showed that students who were not interested in Chinese language, those who achieved lower grades and those who were bored in Chinese contrasted with those who were interested in Chinese, those who achieved higher grades and those who were interested in Chinese.

Table 2: Reliability Coefficient
Thirdly, students with high English examination grades (Bands 1 and 2) obtained a motivation mean of 39.40 whereas those with medium and low grades (in bands 3 and 4) obtained a mean of 36.35. The means differ significantly (F=9.67, df 1, p<0.01).

A fourth significant difference (F=5.36, df 1, p<0.05) in motivation means was found between those who always speak English at home (39.74) and those who do not (35.04). This again suggest the importance of home support in motivating students to learn English. Fifthly, students who evaluated their own English language ability most favourably (Very good) had a motivation mean of 39.46 whereas those whose self-evaluation was not favourable (Not so good and Very poor) had a mean of 34.83. These means differed significantly (F=12.41, df 1, p<0.01), indicating that students with stronger motivation to learn English were those who evaluated themselves more favourably.

Table 1: Factor Loadings

<table>
<thead>
<tr>
<th>ITEMS</th>
<th>CLMOTS</th>
<th>ELMOTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Even if my Chinese/English homework is difficult, I still enjoy doing it.</td>
<td>0.75</td>
<td>0.62</td>
</tr>
<tr>
<td>2 When the teacher tells us something about Chinese/English words, I listen carefully</td>
<td>0.71</td>
<td>0.73</td>
</tr>
<tr>
<td>3 I listen carefully in Chinese/English class, because I do not want to miss anything the teacher is teaching us.</td>
<td>0.70</td>
<td>0.85</td>
</tr>
<tr>
<td>4 Some Chinese/English words are difficult to write but I keep trying until I can write them correctly.</td>
<td>0.72</td>
<td>0.76</td>
</tr>
<tr>
<td>5 If a Chinese/English word is difficult to read, I'll practise so that I can 'say' it correctly.</td>
<td>0.77</td>
<td>0.71</td>
</tr>
<tr>
<td>6 If I have made mistakes in my Chinese/English homework, I like to do corrections.</td>
<td>0.71</td>
<td>0.64</td>
</tr>
<tr>
<td>7 When my Chinese/English teacher tells me about my mistakes, I listen carefully.</td>
<td>0.75</td>
<td>0.83</td>
</tr>
<tr>
<td>8 I like my Chinese/English teacher talking to me about my homework.</td>
<td>0.67</td>
<td>0.77</td>
</tr>
<tr>
<td>9 I study hard to get good marks for my Chinese/English homework.</td>
<td>0.73</td>
<td>0.66</td>
</tr>
<tr>
<td>10 I like to help my classmates learn Chinese/English.</td>
<td>0.71</td>
<td>0.66</td>
</tr>
<tr>
<td>11 I study for Chinese/English tests even before the teacher tells us to do so.</td>
<td>0.70</td>
<td>0.64</td>
</tr>
<tr>
<td>% total variance explained</td>
<td>52.03</td>
<td>51.06</td>
</tr>
</tbody>
</table>

Table 2: Reliability Coefficients

<table>
<thead>
<tr>
<th>Groups</th>
<th>CLMOTS</th>
<th>ELMOTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>Females</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>High grades</td>
<td>0.86</td>
<td>0.92</td>
</tr>
<tr>
<td>Medium &amp; low grades</td>
<td>0.92</td>
<td>0.90</td>
</tr>
<tr>
<td>High self-evaluation</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>Low self-evaluation</td>
<td>0.91</td>
<td>0.89</td>
</tr>
<tr>
<td>All Students</td>
<td>0.91</td>
<td>0.91</td>
</tr>
</tbody>
</table>
The scales for measuring elementary students' motivation to learn Chinese and English, as reported here, share common items, and cover both the affective and behavioural aspects of language motivation in a bilingual schooling context. The two scales show a high degree of internal consistency and discriminant validity with reference to students' background variables including overall academic performance, gender, language examination grade, home language and self-evaluation of language ability. The scales are intentionally kept short so that they will be convenient to use for evaluation, classroom-based research and counselling.

The results seem to indicate that a psychology-based and pedagogy-oriented approach to the measurement of language motivation in the classroom is feasible and promises fruitful development. As for research into language motivation, this approach perhaps serves to complement the attitude-based, integrative-instrumental approach of earlier research carried out in a broader, social context outside the perimeters of the language classroom. Crookes and Schmidt (1991) propose as an item of research agenda "How is SL motivation to be measured or operationalised? Apply validity and reliability criteria to instruments and operationalisations" (p 498). The attempt made in the present study, by incorporating Hart's (1981) conception of classroom motivation and the specifics suggested by Crookes and Schmidt in a language learning situation, is just one possible answer to this call and, obviously, further research is warranted. Further work that could be considered is to ascertain the validity of the scales when used with students at higher as well as lower academic levels, since the development of the scales involved only grade five students. It would also be useful for both theoretical and practical purposes to check empirically on the predictive (as contrasted with the correlational) efficacy of the approach adopted here.

REFERENCES


Optimal Test Designs with Content Balancing and Variable Target Information Function as Constraints

Peter Lam

Optimal test design involves the application of an item selection heuristic to construct a test to fit the target information function in order that the standard error of the test can be controlled at different regions of the ability continuum. The real data simulation study assessed the efficiency of binary programming in optimal item selection by comparing the degree in which the obtained test information was approximated to different target information functions with a manual heuristic. The effects of imposing a content balancing constraint was studied in conventional, two-stage and adaptive tests designed using the automated procedure.

Results showed that the automated procedure improved upon the manual procedure significantly when a uniform target information function was used. However, when a peaked target information function was used, the improvement over the manual procedure was marginal. Both procedures were affected by the distribution of the item parameters in the item pool.

The degree in which the examinee empirical scores were recovered was lower when a content balancing constraint was imposed in the conventional test designs. The effect of uneven item parameter distribution in the item pool was shown by the poorer recovery of the empirical scores at the higher regions of the ability continuum. Two-stage tests were shown to limit the effects of content balancing. Content balanced adaptive tests using optimal item selection was shown to be efficient in empirical score recovery, especially in maintaining equiprecision in measurement over a wide ability range despite the imposition of content balancing constraint in the test design.

The study has implications for implementing automated test designs in the school systems supported by hardware and expertise in measurement theory and addresses the issue of content balancing using optimal test designs within an adaptive testing framework.
Modelling Children's Thinking in Solving Ratio and Proportion Problems: An Information Processing Perspective

Fong Ho Kheong

This thesis is concerned with the development of the Information Processing Taxonomy (IPT) Model which attempts to model children's thinking in solving ratio and proportion problems at different hierarchical levels. The main features of the Model consist of retrieving information from external source and primary and secondary productions of type A and B information. These features constitute the components of different levels of thinking for describing the problem solver's cognitive processes.

The first phase of the study involved validation of the IPT Model. Test items which were constructed in accordance with the Model were validated to check whether they were scalable. Qualitative data obtained from protocol were examined to verify the existence of the hypothetical features depicted at each level of the Model.

In the second phase, the Guttman scale was applied to test the hierarchical nature of the Model. Qualitative analysis involved:

1. analysing pupils' solutions of each ratio and proportion problem,
2. identifying the strategies used by them to solve the problems,
3. generalising strategic models which represent pupils' cognitive processes in solving each ratio and proportion question and the ratio and proportion problems as a whole, and
4. determining the extent to which the IPT Model could be used to explain pupils' thinking processes.

The findings of the research showed that the items from each topic of fractions, composite figures and ratio proportion were scalable. The Model's features could be identified using pupils' solutions and interviewing data. Besides the hypothetical features, additional phenomena were also identified. The study results provided information for:

1. updating the hypothetical IPT Model,
2. identifying a strategic model for solving each of the ratio and proportion problems, and
3. establishing a generic schematic model which represents pupils' thinking processes in solving the ratio and proportion problems designed for the study.